Formas latentes: protocolos de visión artificial para la detección de analogías aplicados a la catalogación y creación artísticas

Pilar Rosado Rodrigo

TESIS DOCTORAL

FORMAS LATENTES:
PROTOCOLOS DE VISIÓN ARTIFICIAL PARA LA DETECCIÓN DE ANALOGÍAS APLICADOS A LA CATALOGACIÓN Y CREACIÓN ARTÍSTICAS

de

Pilar Rosado Rodrigo

2015
Universidad de Barcelona

Programa de Doctorado HD203: La Realidad Asediada: Posicionamientos Creativos
Directores : Dr. Ferran Reverter Comes y Dr. Miquel Àngel Planas Rosselló
Tutora: Dra. Eva Figueras Ferrer

Esta investigación ha sido financiada por el AGAUR (Agència de Gestió d'Ajuts Universitaris i de Recerca de la Generalitat de Catalunya) mediante la contratación de la autora de la tesis como personal investigador novel FI-DGR 2012 en el Departamento de Escultura de la Facultad de Bellas Artes de la Universidad de Barcelona durante el periodo 2012-2015.

TESIS DOCTORAL

Formas Latentes:
Protocolos de Visión Artificial para la Detección de Analogías Aplicados a la Catalogación y Creación Artísticas

Latent Patterns:
Using Computer Vision Models to Detect Analogies Applied to the Classification and Creation of Art

de

Pilar Rosado Rodrigo

2015
Universidad de Barcelona

A Dídac y Mar

AGRADECIMIENTOS

Agradezco su ayuda y aportaciones especialmente a Ferran por creer en mis sueños, a Eva Figueras por estar siempre disponible, a Miquel Planas por compartir conmigo esta aventura, al archivo de la Fundación Antoni Tàpies por facilitarme el acceso a la colección de imágenes digitales del artista, en especial a Laurence Rassell y Núria Solé por su amabilidad y confianza, a Josep María Jori por brindarme generosamente su sabiduría, a Dariusz Frejlichowski y Piotr Czapiewski por la ayuda que me han proporcionado, a Ignasi Labastida por sus eficientes recomendaciones y a Àlex Nogué y Joan Descarga por demostrarme su interés y su apoyo.

Este doctorado ha supuesto un recorrido a lo largo de un camino desafiante que no habría sido capaz de recorrer sin la ayuda de mi familia, en especial de mis padres.

> Creo que en el origen de la creatividad en todos los campos existe lo que yo llamo la capacidad o la disponibilidad para soñar; para imaginar mundos diferentes, cosas diferentes, intentando combinarlos en la propia imaginación de otro modo. A esta capacidad, tal vez muy semejante en todas las disciplinas, de la matemática a la filosofía, de la teología al arte, de la pintura a la escultura, a la física, a la biología, se le une la capacidad de comunicar los propios sueños; una comunicación no ambigua requiere el conocimiento del lenguaje, de las reglas internas propias de las diversas disciplinas. Creo que existe una capacidad para soñar generalmente indistinta, como era generalmente indistinto el sentimiento que los antiguos llamaban filosofía, amor por la sabiduría, y los diversos modos de comunicar de forma no ambigua esos sueños, ese amor por la sabiduría, utilizando lenguajes diferentes, esquemas diferentes que son propios de las diversas disciplinas, y de las diversas artes, de las diversas formas del saber humano (Emmer, 2005, p. 2-3).

ÍNDICE DE CONTENIDOS

RESUMEN

Del mismo modo que Maria Zambrano (1989), esta tesis considera que la pintura "es un lugar privilegiado donde detener la mirada" (p. 11) . La pintura relaciona al hombre con lo que le rodea. La autora no se posiciona ante ella como teórica del arte, ni como crítica, sino como creadora. Zambrano nos explica que sólo es posible la creación para el que sabe mirar, poniendo especial atención en las sombras "para desvelar el enigma que encierra la pintura" (p. 12). Nos habla de ver desde dentro tras haber mirado el cuadro desde fuera. El presente trabajo de investigación se aproxima a las imágenes digitales de obras de arte desde el interior, valiéndose de protocolos de visión artificial.

Frecuentemente la creatividad es acumulativa; suma, enriquece un ámbito de trabajo. A menudo el creador se siente extraño en su dominio, se cuestiona las tradiciones y se sumerge en las nuevas posibilidades que le proporcionan las técnicas, la mezcla de disciplinas. El artista y la necesidad de innovar a lo largo de la historia son una constante y así las revoluciones tecnológicas han comportado cambios en la representación de la realidad. Muchos artistas han sido capaces de utilizar en su favor los nuevos avances de su época; la perspectiva, los estudios de las propiedades de la luz y del color, la fotografía, el cine, el vídeo, la web, etc.

Si en el año 1990 fue el proyecto Genoma, en el 2013 se han iniciado investigaciones multimillonarias transcendentales para el estudio del cerebro humano. Por un lado, desde Estados Unidos, el proyecto BRAIN (Brain Research through Advancing Innovative Neurotechnologies) pretende hacer un mapa de cada neurona del cerebro humano y por otro lado, desde la Unión europea, arranca el proyecto HBP (Human Brain Project) que tienen como objetivo simular el cerebro a través de supercomputadores. Es seguro que en las próximas décadas la inteligencia artificial será fundamental y a su vez una fuente inestimable de nuevas herramientas destinadas a la extracción y producción automática de conocimiento, de las cuales los artistas se podrán beneficiar.

La visión por computador o visión artificial es un subcampo de la Inteligencia Artificial cuyo objetivo es programar a un ordenador para que "entienda" o "interprete" una escena o las características de una imagen. En este ámbito concreto, los investigadores se enfrentan a dos grandes problemas: en primer lugar a las limitaciones que supone registrar las características de las imágenes en un código abstracto, en segundo lugar a la dificultad de elaborar interpretaciones a partir de este código generado. Para superar estos inconvenientes se han creado multitud de metodologías y se evalúan sus rendimientos.

El objetivo de esta tesis es desarrollar un programa informático que implemente algoritmos de visión por computador que permitan, de manera automática, buscar analogías formales en grandes colecciones de imágenes de obras de artista abstractas, basadas únicamente en su contenido visual y sin poyo de anotación textual alguna. De esta manera se espera obtener una herramienta de utilidad tanto en la producción artística como en el análisis de obras de arte.

En el capítulo 1, tras presentar las motivaciones personales que mueven este proyecto, se ponen de manifiesto las enormes diferencias que existen entre el lenguaje visual y el lenguaje verbal o textual; tanto a nivel de lectura como de interpretación, y la importancia que tendría la posibilidad de "dar voz a las imágenes" accediendo directamente a su contenido visual, sin el auxilio de textos y contextos.

Se presentan como antecedentes del análisis de las formas, por un lado a D'Arcy Wentworth desde la biología, como estudioso de la descripción de la forma en términos físico-matemáticos, y por otro lado, desde la psicología, al estructuralismo y la Gestalt como precedentes de estudio de la sintaxis visual y el problema del significado contenido en las artes visuales; cómo y qué comunican las artes.

El objeto de estudio de esta tesis son colecciones de obras de arte abstractas y se apela a la mirada del artista como recolector y productor de formas y analogías de sentido a partir de su entorno, utilizando principios estadísticos desde el momento en que observa la diversidad, la procesa y abstrae el modelo que considera significativo.

En cuanto a la interpretación del arte, al discurso que puede desprenderse del análisis de sus colecciones, se recuerda el intento visionario de Aby Warburg que, con su Atlas Mnemosyne, ya intentó construir una memoria de la civilización europea en función únicamente del contenido de sus imágenes, sin apenas relato de apoyo.

En este mismo capítulo se realiza un recorrido para situar la utilización en la actualidad de las metodologías de visión artificial en el análisis de obras de arte, precisando su profuso empleo en tareas de autentificación o para descubrir los métodos y herramientas utilizadas en la historia de la pintura. En un apartado concreto se especifican los antecedentes de aplicación de estas técnicas en la clasificación de imágenes de artistas, algunos con la intención de categorizar estilos pictóricos, pero todos ellos aplicando métodos de aprendizaje automático que requieren una clasificación previa realizada por expertos.

Las novedades que aporta nuestro planteamiento en este contexto serían; por un lado la búsqueda de formas latentes en colecciones de arte abstracto, y por otro, la aplicación de un método totalmente automático que no requiere intervención previa de nadie para establecer la taxonomía visual. Se anticipa el hecho de que la aplicación de la metodología objeto de estudio en la presente tesis para el análisis de arte abstracto es novedosa ya que no se encuentran antecedentes y únicamente se ha puesto a prueba en la clasificación de escenas naturales (fotografías de paisajes, escenas de interior, paisajes urbanos, detección de objetos). En estos contextos se han obtenido excelentes resultados que animan a la extensión de su uso. En nuestra hipótesis se presupone que en una colección de obras de artista abstractas existen constantes visuales, correlaciones formales que son susceptibles de ser calculadas mediante estas técnicas de visión por computador. La imagen como superficie de significado es explorada por la mirada artificial y el sentido viene dado por criterios matemáticos de similitud.

En el capítulo 2 se explica exhaustivamente la metodología con el apoyo de los Anexos A y B, en los que se incluyen la formulación matemática y la terminología más empleada, respectivamente.

Se explora un modelo concreto de descripción de imágenes utilizado en visión artificial

cuyo enfoque consiste en colocar una malla regular de puntos de interés en la imagen y seleccionar alrededor de cada uno de sus nodos una región de píxeles para la que se calcula un descriptor invariante a la transformación de la imagen, que tiene en cuenta los gradientes de grises encontrados. Analizando las distancias entre el conjunto de descriptores de toda la colección de imágenes, se pueden agrupar en función de su similitud y estos grupos resultantes pasarán a determinar lo que llamamos palabras visuales. El total de palabras visuales de una colección de imágenes genera un vocabulario visual concreto del conjunto. El método se denomina *Bag-of-Words* (*BoW*, bolsa de palabras) porque representa una imagen como una colección desordenada de características visuales locales.

Se detalla la implementación de una nueva descripción de las características de la imagen que sí tiene en cuenta la distribución espacial, y posteriormente se explica cómo, una vez construido el vocabulario visual de la colección de imágenes, es posible obtener un nivel más de información utilizando modelos estadísticos que son capaces de discriminar patrones de distribución entre estas palabras.

En este mismo capítulo se explican también en detalle otro tipo de descriptores que se han utilizado en la tesis para obtener unos resultados comparativos; los descriptores de textura de Haralick.

En el capítulo 3, en primer lugar se pormenorizan los cuatro algoritmos desarrollados en la presente tesis: el de categorización supervisada, el de categorización no supervisada, el de agrupación basado en descriptores de textura de Haralick y el de cálculo de la distancia de Bhattacharyya. El uso de estas herramientas puede hacerse extensivo en el futuro al estudio de otras colecciones de obras de arte: proporcionando un punto de vista auxiliar, ampliando y facilitando las relaciones que se establecen entre obras de un mismo artista y diferentes periodos, y entre artistas de diferentes épocas.

En segundo lugar, en el capítulo 3 del presente estudio se comentan las particularidades de los resultados obtenidos al aplicar los algoritmos informáticos en las colecciones de obras de arte a las que se ha tenido acceso en la tesis. Los tres experimentos que se han realizado en el presente estudio han sido: primero, un análisis sobre la colección de 2846

imágenes fotográficas que el artista Miquel Planas utiliza como fondo de ideación artística en el que, en primera instancia se etiquetó manualmente el conjunto de datos para entrenar al sistema y así poder predecir la clasificación de imágenes problema; después, sobre la misma colección de imágenes, un estudio de clasificación totalmente automática en la que el sistema es capaz por si solo de detectar las categorías formales existentes; y por último se detallan los resultados de aplicar esta última metodología sobre la colección de 434 imágenes digitalizadas de pintura y obra gráfica (gran parte perteneciente a libros de artista) de Antoni Tàpies que posee su Fundación en Barcelona (Tàpies, 2001). El paso de imagen fotográfica a imagen de obra pictórica supone un nuevo grado de complejidad para el sistema dado que ya no se trata de imágenes extraídas directamente de la realidad en la que las palabras visuales se corresponden con elementos naturales como agua, piedras o cielo, sino que son construcciones del artista, lo que supone un reto mayor de categorización.

En este capítulo también se especifican los resultados de aplicar métodos basados en distancias matemáticas entre imágenes en la colección de Tàpies y con ellos se dibuja un dendograma de toda la colección que resulta muy informativo acerca de las relaciones formales que se establecen entre grupos de imágenes y sobre su grado de similitud.

Para finalizar se muestran y se comentan las agrupaciones obtenidas en base a los descriptores de textura de Haralick y se comparan con los resultados previos hallados con los descriptores invariantes a la transformación de la imagen.

Finalmente en el capítulo 4 se describen y discuten las aportaciones y conclusiones de la tesis y se realizan propuestas para futuras aplicaciones.

SUMMARY

This thesis supports María Zambrano's notion that the world within a painting is "a special place to stop and stare"[1]. Painting relates people to the world around them and Zambrano understood this from the point of view of the creator rather than the scholar or critic. To create, she argued, you need to be able to look; and to look, you need to pay special attention to the shadows, which is where we "unveil the enigma that is closed inside painting"[2] . Zambrano talked about seeing paintings "from the inside" after looking at them from outside. By using computer vision techniques to study the digitised images of large painting collections, the present study could also be said to examine paintings from the inside.

Creation is often informed by accretion. Things come together and a line of activity is gradually embellished. But when the line becomes too narrow for comfort, the creator questions traditional practices and finds new techniques and hybrid disciplines. Throughout history, creating art and being innovative have been inseparable and this is why revolutions in technology are closely tied to our changing representation of reality. Many artists have found their own uses for technological innovation, whether borrowing from the advances in the early study of perspective, from the periods in history when light and colour were researched or from the advent of photography, film, video and the Internet.

If 1990 was the year of the Human Genome Project, 2013 will be remembered for the US launch of the billion-dollar BRAIN Initiative (Brain Research through Advancing Innovative Neurotechnologies), which eventually hopes to map every neuron in the human brain, and the beginning of the EU's equally costly Human Brain Project, which is creating new IT platforms in the field of brain simulation. In the decades to come, R&D in artificial intelligence is likely to generate a wide array of applications to extract and produce knowledge, which artists will be able to turn to their favour.

1 In Zambrano's own words, "un lugar privilegiado donde detener la mirada" (Zambrano, 1989, p. 11).
2 "…para desvelar el enigma que encierra la pintura." (Ibid., p. 12)

A particularly important area of development will be computer vision, a subfield of artificial intelligence which programs computers to "understand" or "interpret" the content of a given scene or feature-rich image. Computer vision research currently faces two key issues: the limitations involved in recording the features of a given image in an abstract code and the difficulty of then interpreting the codes. The various models that have been developed and tested to negotiate these problems are regularly discussed and assessed in the literature.

The objective of this thesis is to develop a series of computer vision programs to search for analogies in large datasets—in this case, collections of images of abstract paintings—based solely on their visual content without textual annotation. In this way, the researchers hope to develop a tool both for producing and analysing works of art.

Chapter 1 begins by outlining the personal reasons why this research was undertaken and describes the major differences between visual language and verbal or textual language, evidenced by how we read and interpret each. It discusses the value to be derived from "letting images speak for themselves" and having direct access to the visual content of abstract paintings without textual annotation or contexts.

It discusses antecedents in the history of the study of visual patterns, citing biologist D'Arcy Wentworth Thompson (who used physics and mathematics to study pattern-formation in the natural world), the visual syntax practised by structuralism and gestalt psychologies (which organise the elements in images into various groups), and the subject of how meaning is contained and expressed in the visual arts today.

The researchers then describe the basic material this thesis uses—large collections or datasets of images of abstract paintings—and proposes that in abstract art the painter's eye becomes the eye of a gatherer and producer of patterns and analogies culled from that person's immediate environment. It argues that artists use essentially statistical principles from the moment they observe diversity to the time they process and finally abstract this into models they consider meaningful.

About how art is interpreted and the discourse that emerges from the analysis of art collec-

tions, the researchers recall the visionary attempt by Aby Warburg's *Bilderatlas Mnemosyne* to reconstruct an account of European civilisation almost solely on the basis of pictures and photographs, with hardly any recourse to textual annotation.

Chapter 1 also observes that computer vision has already been employed to examine works of art, whether in the extensive use of fractal analysis in authentication studies or in the computer vision algorithms that help researchers study painters' methods and tools in different periods of art history. One section describes the use of these techniques to classify artists' paintings, for example to group paintings by pictorial style, in all cases applying machine learning techniques to a prior classification performed by art experts.

As the researchers then explain, this thesis finds two new uses for computer vision techniques in art. First, it proposes that computer vision can help detect latent patterns in collections of abstract paintings; second, the method it develops to establish a visual taxonomy is totally automated and requires no previous intervention. The researchers argue that this application is novel and that, to date, any similar research has been limited to natural scene classification (with photographs of landscapes, interiors, cityscapes) and object detection. But the excellent results in these areas have encouraged the present study, whose premise is the following: any collection of abstract art will contain visual constants and formal correlations that can be computed with computer vision techniques, and these can incorporate mathematical similarity to explore an abstract painting as a surface of meaning.

Chapter 2 provides a thorough account of the research methodology and is supported by appendices A and B, which describe the most important mathematical formulae and terminology, respectively.

This chapter studies a specific model for describing pictures with computer vision. This consists in positioning a regular mesh of interest points in the image and selecting, around each mesh node, a region of pixels to be assigned a descriptor that remains invariant under different transformations and anticipates grayscale. By analysing the distances between the set of descriptors across the entire image collection, images can be grouped by similarity and groups can determine what we call 'visual words', meaning the arrays of pixels within

an image that would correspond to the words within a text. The total number of visual words in a collection of images generates a visual vocabulary specific to that collection. In the literature, this is referred to as the *Bag-of-Words* model (hereafter, *BoW*) because it ignores spatial relationships and simply represents the image as a disordered bag of local visual features.

Next, the chapter describes the implementation of a new description of the features of the image that captures spatial information. It explains how, once the visual vocabulary of the collection of images has been constructed, another level of information can be obtained using statistical models which discriminate distribution patterns between the visual words.

Finally, this chapter also reports on the use of Haralick's texture descriptor to obtain comparative results.

Chapter 3 starts by presenting the four algorithms developed in this thesis: the algorithm for supervised classification, the algorithm for unsupervised classification, the algorithm based on Haralick's texture descriptor and the algorithm for calculating Bhattacharyya's distance. In future studies of other art collections, the use of these instruments may become more widespread, providing a helpful point of view, broadening and facilitating the associations established between the works of the same artist in different periods or between different artists and periods.

Chapter 3 then considers the results obtained by applying the algorithms to specific art collections. Three experiments were performed. First, the researchers analysed a set of 2846 photographs used by the artist Miquel Planas as a basis for artistic ideation, manually labelling the dataset to train the system to predict the classification of problematic images. Second, the same collection of images was subjected to a totally automated classification study in which the system autonomously detected the existing formal categories. Third, this same procedure was applied to a collection of 434 digitised images, mainly art book reproductions, of paintings and graphic works by Antoni Tàpies that belonged to the Tàpies Foundation in Barcelona (Tàpies, 2001). In this third experiment, the progression from photographs (Planas) to abstract paintings (Tàpies) Involved a new and complex challenge, given that

the system had to classify images whose visual words (pixel arrays) did not identify natural features of the real world ('water', 'stones', 'sky') but rather the artist's abstract constructions.

This chapter also reports on the results of applying methods based on mathematical distances between images in the Tàpies collection and draws a dendogram of all the collection. This provides valuable insight on the formal relationships between groups of images and their degree of similarity.

Chapter 3 concludes by analysing the groupings obtained with Haralick's texture descriptor compared with the prior findings obtained with descriptors that remained invariant under different transformations.

Finally, Chapter 4 discusses the contribution made by this study, draws conclusions and proposes future applications.

1 - INTRODUCCIÓN

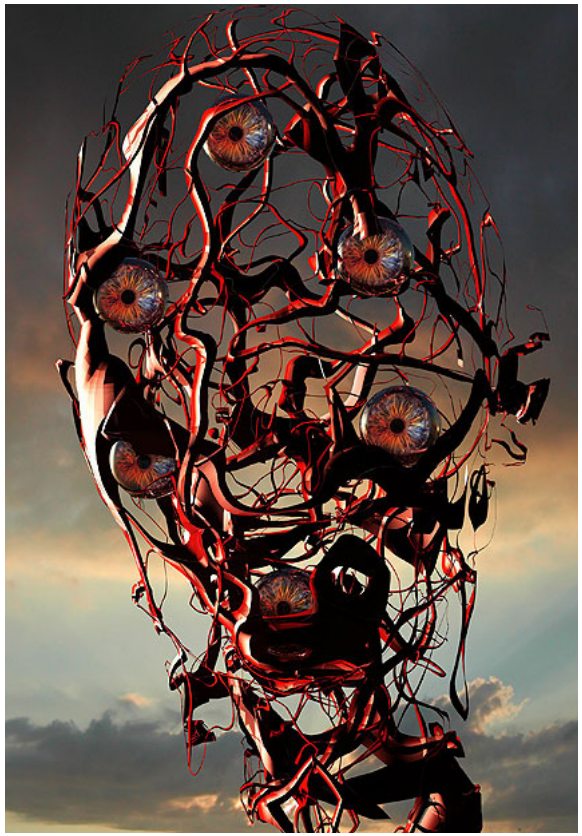


Figura 1.1. Marina Núñez, 2007. "Ocaso", Video monocanal © (Núñez, 2008)

1.1 MOTIVACIÓN PERSONAL

Una tesis doctoral es un trabajo académico eminentemente científico que lleva implícita una importante implicación personal, más si se presenta, como en este caso, en el ámbito de las Bellas Artes. Por este motivo, aunque el presente proyecto no consista en un trabajo enmarcado en el contexto de mi trayectoria artística, sí que es producto de un previo recorrido personal concreto que me gustaría compartir con el lector dado que considero estos aspectos fundamentales para que la presente tesis sea lo que es, tanto en la forma como en el fondo.

Me considero una persona constitutivamente multidisciplinar y este hecho ha propiciado que, en todas las fases de mi vida en las que me he visto obligada a escoger entre disciplinas como ciencias o letras, haya experimentado un sentimiento de pérdida, pues considero que las personas están por encima de las disciplinas. En bachillerato escogí ciencias para más tarde estudiar Ciencias Biológicas, que en los años 80 era una especie de cajón de sastre que contenía disciplinas tales como zoología, botánica, microbiología, ecología, bioquímica, genética, muchas de las cuales más tarde han dado lugar a grados completos. En aquel tiempo todas estas materias se estudiaban juntas a lo largo de 5 años con el objetivo de "estudiar la vida". El último año se daba la oportunidad de escoger 3 asignaturas y opté por la genética, un ámbito que resultó apasionante porque ayudaba a comprender los mecanismos con los que el ADN contiene la información que genera las formas de los seres vivos. A pesar de ser una disciplina con un futuro impresionante, los años 80 no eran para España precisamente fáciles, especialmente para trabajar en investigación.

De forma paralela, a principios de los años 90, la compañía informática IBM (International Bussiness Machines Corp.) había vendido a muchas grandes empresas del mundo su mainframe IBM 3090[1]. La licenciatura de informática en España apenas hacía unos años que

1 Gran computada que comercializó IBM en 1985 con Sistema Operativo MVS, Bases de Datos IMS (DL/1), gestor de teleproceso IMS/DC y software de Aplicación programado en Cobol.

se había iniciado y existía en nuestro país demanda de personal con formación específica para este sistema[2]. Para suplir esta carencia surgieron empresas de servicios informáticos que proponían a licenciados en disciplinas tan dispares como física, matemáticas o biología, un plan de formación remunerado para el aprendizaje en análisis y programación enfocada a estos grandes sistemas informáticos. La propuesta era perfecta para una estudiante en paro; nada que perder y cobrar por aprender constituía un reto tan insólito como interesante. Y el desafío resultó fascinante; el proceso de diseñar, codificar y depurar un código fuente en un lenguaje de programación determinado con el objetivo de que el ordenador produzca un comportamiento concreto me sedujo durante los siguientes 10 años.

Aunque no lo he mencionado con anterioridad, desde mi infancia la pintura ha formado parte siempre inseparable de mi cotidianeidad; veía a mi madre pintar copias de Goya y Velázquez en el salón de casa en sus ratos de ocio, a tamaño original, y en mi edad adulta la academia de pintura era el socorrido refugio de mi creatividad tras horas interminables de trabajo. La pintura siempre ha estado latente en mi vida, hasta que un buen día tomó fuerza y llegó la imperiosa necesidad vital de dedicar más tiempo al arte. Dando por concluido un ciclo, reconfiguré mi vida de nuevo para poder estudiar escultura, pintura, grabado e inicié mi formación en la facultad de Bellas Artes de Barcelona en el año 2000, con el cambio de siglo. Después cursé el master, el doctorado y finalmente me encuentro en proceso de defender la tesis, que principalmente ha servido para reconciliar mis pasiones, para que todo tenga sentido, las piezas del puzzle encajen y tomen su lugar. Ahora no he necesitado renunciar a nada porque precisamente mi trabajo de investigación ha consistido en integrar algunas de las habilidades que he podido ir recopilando en mi trayectoria académica y profesional: la posibilidad de programar un ordenador, la formación científica y la artística se convierten en fundamentales a la hora de desarrollar este proyecto. Con cierta perspectiva veo más claro que mis intereses pluridisciplinares en el caso concreto del planteamiento y desarrollo de esta investigación se han entrelazado produciendo un efecto de simbiosis y sinergia.

2 En 1976 los Decreto 327/76 y 593/76 crean en el estado español las primeras Facultades de Informática (licenciaturas), entre ellas la Facultad de Informática de Barcelona.

Figura 1.2. Marina Núñez, 2012. Sin título (ciencia ficción). Infografía sobre papel. 6 piezas de 45 x 70 cm. © (Núñez 2008)

El hecho de escoger como directores de este trabajo de investigación al Dr. Reverter, estadístico y al Dr. Planas, escultor, ha constituido también un factor determinante por la mirada poliédrica que han aportado sobre el tema en cuestión. El apoyo recibido por los doctores Dariusz Frejlichowski y Piotr Czapiewski, especialistas en visión por computador (concretamente en las áreas de inteligencia artificial y reconocimiento de patrones), durante mi estancia en la Faculty of Computer Science and Information Technology de Szczecin, ha resultado decisivo para la consecución de los objetivos planteados en la tesis.

Reconozco que ha supuesto una dificultad encontrar el tono descriptivo de la redacción del trabajo dado que cada disciplina tiene una terminología propia, pero considero que el diálogo ha resultado fructífero y que ha favorecido planteamientos que de otro modo hubiesen sido difíciles de imaginar.

1.2 PREÁMBULO

En el cerebro existen más de treinta áreas dedicadas al procesamiento de la información recogida por los ojos. No hay otra actividad que requiera una cantidad de recursos equivalentes. Cuando se explica alguna idea compleja a una persona inglesa o a una española, cuando la entienda es muy probable que exclame: I see! (inglés) o Ya veo! (español), porque ver y entender, en nuestro interior son procesos muy cercanos (Cairo, 2011).

El proyecto que aquí se presenta, tal y como nos sugiere su título, trata fundamentalmente de descubrir formas latentes en colecciones de imágenes de artista para que las analogías encontradas nos aproximen y hagan mas comprensibles los valores originales del creador. En el ámbito de la creación artística y de su enseñanza, es una práctica muy habitual colocar grupos de imágenes fotográficas, esculturas o pinturas, unas junto a las otras. Es sorprendente observar cómo los objetos visuales dialogan entre sí, unas veces ratificándose y otras contradiciéndose. Es un tipo de conocimiento que resulta difícil de explicar de otro modo, especialmente complicado con palabras, pero que al presentarlo a la vista en su totalidad,

proporciona una comprensión inmediata de sentido. De esta forma el artista, el profesor o el estudiante de arte tienen a su alcance una información que, tratando los mismos objetos de forma aislada, sería inaccesible. Este aspecto es especialmente valioso cuando se trata de estudiar imágenes de contenido abstracto, dado que en ellas el tema, el significado o el sentido no es producto de un acuerdo social, sino que se trata de resonancias visuales y sincronías que el artista creador relaciona y vincula.

Por otro lado, las "buenas" obras, las "afortunadas" que pasan a formar parte de colecciones museísticas, de galerías o de particulares, vuelven de nuevo a su estado de soledad. Se guardan en cajones o se almacenan, bien protegidas para salvaguardar sus cualidades y, en alguna ocasión se exponen junto a acompañantes accidentales con los que no necesariamente comparten una relación de contenido visual. Así que, esta potente herramienta de conocimiento por yuxtaposición, se infrautiliza por razones obvias; simples restricciones organizativas.

Las nuevas tecnologías permiten la digitalización de estos contenidos y con ello favorecen el rápido acceso a su visualización. Pero para buscar relaciones genuinas entre imágenes, de propio contenido visual, son necesarias herramientas de búsqueda más sofisticadas, que no requieran anotación textual alguna, ya que este mismo escrito condicionará las relaciones que la imagen podrá establecer a posteriori. A este fin se dedica un subcampo de la inteligencia artificial, la visión por computador, una disciplina multidisciplinar apasionante en la que se han conseguido grandes retos. El que más nos interesa en el presente proyecto es el de la recuperación de imágenes en base a su contenido visual. La resolución de la histórica dialéctica entre textos e imágenes podría provenir de estos métodos que permitirían a las imágenes "hablar con su propia voz".

1.3 HIPÓTESIS Y OBJETIVOS

El propósito de esta tesis doctoral es el desarrollo de un sistema para la consulta de colec-

ciones digitales de arte no figurativo con software especializado en análisis de imágenes por contenido.

La hipótesis de partida es el convencimiento de que existen categorías semánticas en el conjunto de obras de un artista consolidado que son susceptibles de ser determinadas mediante métodos de visión computacional.

En esta investigación nos proponemos conseguir una clasificación automática de imágenes digitales de obras de artista abstractas basada únicamente en su contenido semántico, sin necesidad de anotación textual alguna, que sea robusta y considerada como significativa por expertos en arte.

El reto al que nos enfrentamos, a diferencia de los estudios encontrados en la literatura que abordan la agrupación automática por contenido visual de imágenes digitales bidimensionales de escenas naturales y objetos cotidianos, es que estas tienen un contenido semántico universalmente asumido y en cambio, las bases de datos de artista que utilizamos en nuestro análisis son colecciones de imágenes de formas que el artista creador vincula porque considera que entre ellas existen analogías de sentido, y que por tanto suponen un reto de más difícil validación. Las Fig. 1.3 y 1.4 muestran una selección de los tipos de imágenes a los que nos referimos.

Un sistema de agrupación y ordenación de este tipo, además de la propia catalogación, propiciaría la apreciación de nuevos valores, nuevas cualidades y características comunes entre las imágenes comparadas.

Así, en el presente proyecto concurren por un lado el artista, por otro su obra y por otro un observador, que en nuestro caso sería una máquina computadora especialmente programada para ello.

La recuperación de imágenes basada en contenido (Content-Based Image Retrieval, CBIR), tal como la vemos hoy, es cualquier tecnología que, en principio, ayude a organizar archivos digitales por su contenido visual. En consecuencia, cualquier procedimiento adecuado

Figura 1.3. Muestra de imágenes de Antoni Tàpies

tanto para medir el parecido de imágenes como para anotarlas cabe en el ámbito del CBIR.

El CBIR es una área de investigación que encontramos situada en una coyuntura multidisciplinar en la comunidad científica. A pesar de que el esfuerzo original radica en dar solución al problema fundamental de la interpretación de imágenes, vemos a personas de diferentes ámbitos, como por ejemplo, la visión por computador, el aprendizaje automático, la recuperación de información, la interacción hombre-computador, los sistemas de base de datos, la Web, la minería de datos, la teoría de la información, la estadística y la psicología, contribuir y formar parte de la comunidad del CBIR.

Los años 1994-2000 se consideran como la fase inicial de investigación y desarrollo en la recuperación de imágenes por contenido. Los progresos realizados durante esta fase han sido resumidos por Smeulders, Worring, Santini, Gupta & Jain (2000), que han tenido una clara influencia sobre los progresos realizados en la década posterior. Estos autores remarcan dos aspectos que definen y motivan las dos grandes líneas de trabajo implicadas en este ámbito de investigación.

1- Brecha sensorial (*Sensory Gap*): La brecha sensorial se refiere a la diferencia entre el objeto en el mundo real y la descripción computacional que hacemos derivada de una grabación de este objeto.

2- Brecha semántica (*Semantic Gap)*: La brecha semántica se refiere a la diferencia entre la información que se puede extraer de los datos visuales y la interpretación que de los mismos datos hace un usuario en una situación dada.

Mientras que el primero expresa que la recuperación de imágenes basada en contenido es una tarea compleja debido a las limitaciones que supone registrar las imágenes con un código abstracto, el segundo enfatiza la dificultad de elaborar interpretaciones a partir de las imágenes.

Por la naturaleza de su tarea, la tecnología CBIR se reduce a dos problemas intrínsecos:

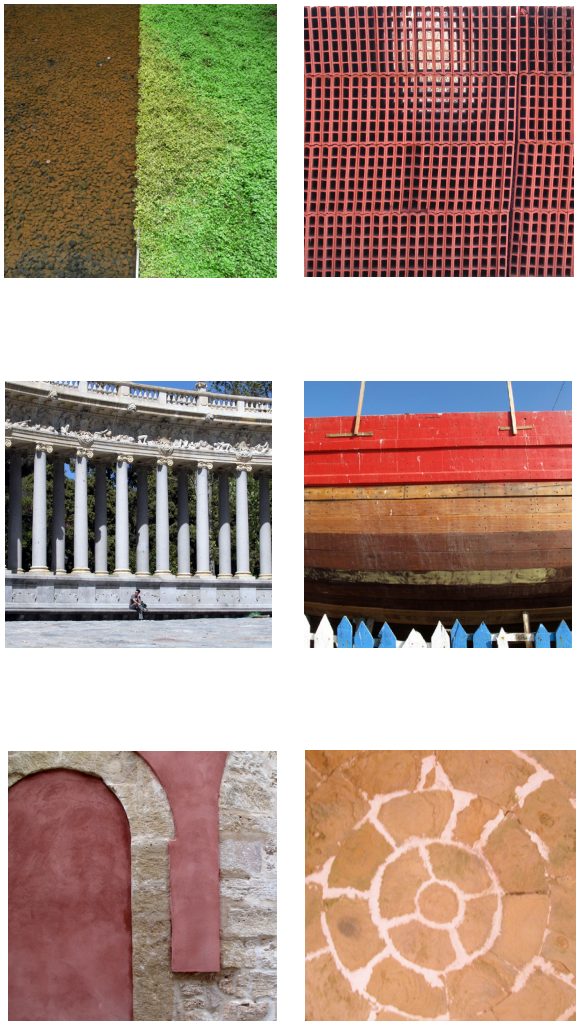


Figura 1.4. Muestra de imágenes de Miquel Planas

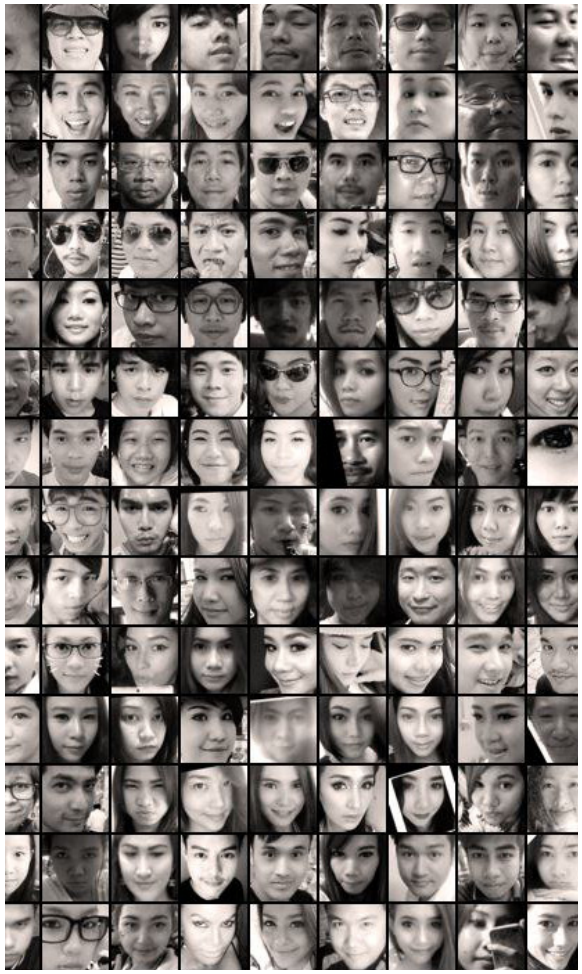


Figura 1.5. Lev Manovich 2014. Selfiecity. Aplicación web interactiva que explora un conjunto de 3200 fotos selfie realizadas por personas de 5 ciudades diferentes del mundo (Manovich, 2015)

1- Como describir matemáticamente una imagen, y
2- La forma de evaluar la similitud entre un par de imágenes basada en sus descripciones abstractas (matemáticas).

La primera cuestión se plantea porque la representación original de una imagen, que es una matriz de valores de los píxeles, corresponde pobremente a nuestra respuesta visual, por no hablar de la comprensión semántica de la imagen. Nos referimos a la descripción matemática de una imagen, con fines de recuperación, como su firma. Desde la perspectiva del diseño de un sistema CBIR, la extracción de las firmas y el cálculo de semejanzas de las imágenes no pueden estar netamente separados. La formulación de la firma determina en gran parte el ámbito de las medidas de similitud.

1.3.1 Construcción de la identidad de la imagen

La firma de una imagen se puede construir a partir de sus características. Una característica se define para describir una determinada propiedad visual de una imagen, ya sea globalmente para toda la imagen o local para un pequeño grupo de píxeles. Las características más utilizadas son aquellas que describen los colores, la textura, la forma y los puntos más destacados de una imagen.

En general, hay dos aproximaciones para obtener la firma de una imagen a partir de las características: Firmas basadas en regiones y firmas basadas en modelos matemáticos.

Las firmas a partir de regiones son más habituales, como argumenta Wang, Li, Gray, & Wiederhold (2001), esto se debe a que regiones homogéneas en color y textura en una imagen se corresponden probablemente con un objeto. Por lo tanto, la detección de regiones en una imagen equivale a detectar una colección de objetos con los que resultará más fácil establecer intuiciones para definir las medidas de similitud.

1.3.2 Similitud entre imágenes: Distancia

Una vez se dispone de las firmas de las imágenes, es decir, de una descripción matemática, la semejanza entre ellas viene dada por la semejanza entre las dos descripciones. El concepto de similitud se utiliza dado que, intuitivamente, datos similares tendrán grupos/clases similares. La distancia será inversa a la similitud; a mayor distancia menor parecido. Según el tipo de firmas utilizado tendremos que recurrir a una función de distancia que se adecue (Wang et al., 2001).

1.3.3 Agrupación y clasificación de imágenes : Aprendizaje máquina

Llegados a este punto podemos comentar que nos encontramos ante el problema de la extracción automática de conocimiento. Pero, ¿en que consiste aprender? Aprender, según la visión genérica de Mitchell, (1997) es mejorar el comportamiento a partir de la experiencia, según Kurzweil, (2013) es la identificación de patrones, de regularidades existentes en la evidencia o según las últimas investigaciones relacionadas con el funcionamiento del cerebro, las conexiones de nuestras neuronas se van configurando y reconfigurando según el tipo y la cantidad de información que transmiten en cada instante (Forbes, 2005).

En este proyecto, las técnicas de aprendizaje automático (Marsland, 2009) que se han utilizado pertenecen a la categoría de:

1- Aprendizaje supervisado: Se proporciona un conjunto de entrenamiento con ejemplos correctos y en base a este, el algoritmo es capaz de generalizar respuestas correctas a imágenes problema. En una fase inicial de este proyecto, un grupo de expertos en arte realizó una clasificación manual de 75 imágenes en 5 categorías (15 imágenes en cada categoría). El promedio de acierto del 70% en la posterior clasificación de la muestra problema resultó prometedor.

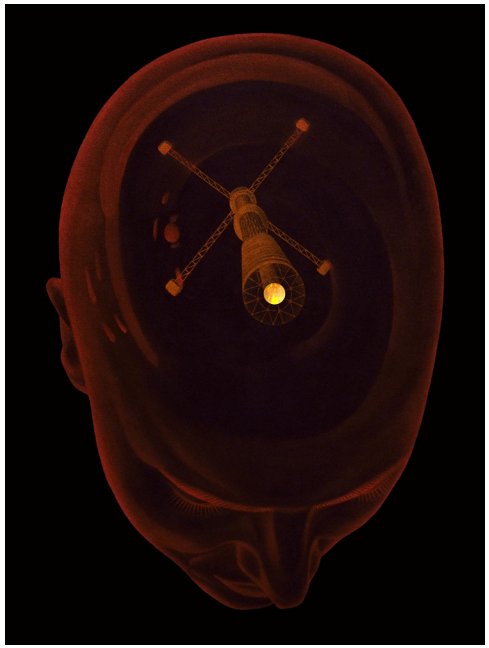


Figura 1.6. Marina Núñez, 2000. Sin título (ciencia ficción). Óleo sobre lienzo y luz. © (Núñez, 2008)

2- Aprendizaje no supervisado: No se proporcionan respuestas correctas, y el algoritmo intenta identificar similitudes entre el conjunto de entradas proporcionado. Las imágenes que tienen algo en común se categorizan juntas. En una posterior etapa de la investigación se aplicó este tipo de algoritmo para intentar encontrar agrupaciones en el conjunto de obra de los artistas, sin anotación textual ni intervención manual alguna, obteniendo también buenos resultados reconocidos por los expertos.

No hemos encontrado antecedentes en la literatura de aplicación de métodos no supervisados de clasificación automática de imágenes basada en contenido aplicadas a conjuntos de imágenes de obras de arte, por lo que esta tesis abre la posibilidad de explorar este territorio de forma extensiva.

Los resultados obtenidos han sido corroborados por expertos dedicados al análisis de obras de arte así como a la producción artística.

La pertinencia de la investigación planteada viene justificada por el convencimiento de que este trabajo permitirá mejorar el acceso a grandes colecciones de imágenes de artista atendiendo únicamente a su contenido visual y que favorecerá el establecimiento de relaciones novedosas entre dichas imágenes. Así mismo constituirá una herramienta positiva para generar análisis que servirán de punto de partida para estudiosos del arte y para los propios artistas, proporcionando a su vez nuevas formas de relacionarse con el patrimonio cultural digitalizado y de compartir el conocimiento.

En la actualidad disponemos de tecnologías poderosas y de grandes reservas de materiales digitalizados que anteriormente los estudiosos no tenían. El presente estudio de investigación supone una aportación a los trabajos que se están realizando en la actualidad sobre aplicación de técnicas de visión artificial junto con métodos estadísticos en la mejora del análisis y comprensión de imágenes artísticas. En nuestro proyecto nos interesan especialmente las novedades que puede revelar su aplicación sobre los problemas del arte y la historia del arte y la forma en la que puede cambiar el análisis de imágenes por ordenador nuestra comprensión del arte.

Estos nuevos métodos, guiados por el conocimiento histórico del arte, pueden arrojar nueva luz sobre las obras de arte y la praxis artística.

1.4 CONTEXTUALIZACIÓN

En un ensayo sobre cómo la inteligencia artificial puede parecerse a la humana, Kurzweil [3](2013) nos describe muy sintéticamente la forma en que el relato de la inteligencia humana comienza con un universo que es capaz de codificar información. Este es el factor que ha hecho posible que la evolución tenga lugar a pesar de ser un hecho increíblemente poco probable. Los átomos formaron moléculas cada vez más complejas y miles de millones de años después, surgió una molécula compleja llamada ADN que podía codificar con precisión largas cadenas de información y generar organismos descritos mediante dichos programas. Los organismos desarrollaron redes de comunicación y de decisión llamadas sistemas nerviosos. Sus componentes, las neuronas se juntaron en cerebros cada vez más inteligentes que se encuentran en la vanguardia del almacenamiento y manipulación de la información. El cerebro de mamíferos posee una capacidad no encontrada en ninguna otra clase animal: la capacidad de pensar jerárquicamente, de comprender una estructura compuesta de diferentes elementos ordenados según un patrón. En los humanos, el neocórtex ha alcanzado un nivel de sofisticación y capacidad tal que estos patrones han merecido el nombre de ideas. La inteligencia de nuestros cerebros junto con el pulgar oponible que nos permite manipular el medio y construir herramientas han hecho posible que la neurología de lugar a la tecnología.

3 Raymond Kurzweil es un inventor estadounidense, además de músico, empresario, escritor y científico especializado en Ciencias de la Computación e Inteligencia Artificial. Desde 2012 es director de ingeniería en Google. Experto tecnólogo de sistemas de Inteligencia Artificial y eminente futurista. Es actualmente presidente de la empresa informática Kurzweil Technologies, que se dedica a elaborar dispositivos electrónicos de conversación máquina-humano y aplicaciones para personas con discapacidad y es impulsor de la Universidad de la Singularidad de Silicon Valley.

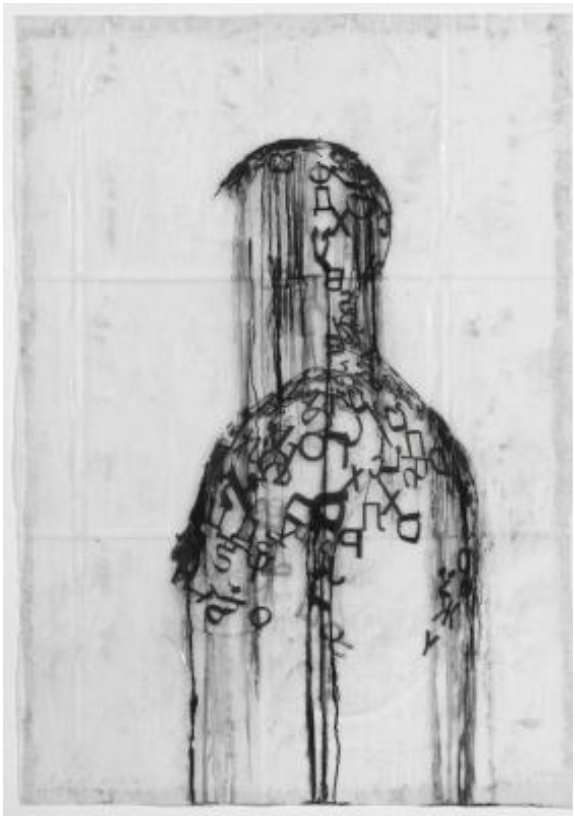


Figura 1.7. Jaume Plensa, 2012. Light Shadow XII. Técnica mixta en papel. © (Plensa, 2008).

La apasionante cuestión para el futuro es si la humanidad podrá o no encontrar un algoritmo que convierta un ordenador en una entidad que sea equivalente a un cerebro humano.

1.4.1 Texto versus Imagen

Nuestra primera invención fue el relato hablado que nos permitió representar conceptos mediante vocablos diferenciados. Posteriormente con la invención del lenguaje escrito, desarrollamos diferentes formas de simbolizar nuestros pensamientos. Así, mediante ideas recursivamente estructuradas, las bibliotecas aumentaron enormemente la capacidad de nuestros cerebros de retener y expandir nuestra base de conocimientos. El proceso evolutivo de la tecnología desembocó inevitablemente en el ordenador, que a su vez ha supuesto una enorme expansión de nuestra base de conocimientos y ha permitido aumentar extraordinariamente la capacidad de comunicación entre las diferentes áreas del conocimiento. Kurzweil explica en detalle en este libro su teoría de la mente basada en el reconocimiento de patrones. Explica que IBM trabaja en el diseño de un ordenador que será capaz de leer inmensas cantidades de literatura médica en lenguaje natural para convertirse en un maestro del diagnóstico y la consulta médica, pero claro, que realmente el ordenador no "entenderá" lo que lea porque se limitará a realizar un análisis estadístico. Afirma que las matemáticas que se han desarrollado en el campo de la inteligencia artificial son muy similares a los métodos que la biología desarrolló bajo la forma del neocórtex.

Proponemos, como Flusser (2009), que en la cultura humana se han producido dos acontecimientos fundamentales; el primero la "invención de la escritura lineal" alrededor de la mitad del segundo milenio antes de Cristo, y el segundo la "invención de las imágenes técnicas", en el momento actual. Esta visión de la historia nos sitúa en la pugna entre escritura e imagen, entre dos "contenedores de significados" que codifican y contienen el tiempo de manera diferencial; Flusser (2009) habla del tiempo circular de la magia en las imágenes y del tiempo lineal de la historia en los escritos:

> Las imágenes son superficies con significado. Normalmente señalan algo ubicado *afue-*

> *ra* en el espacio-tiempo, que han de hacer concebible en forma de abstracciones (reducciones de las cuatro dimensiones de espacio y tiempo a las dos de la superficie). Esta capacidad específica de abstraer superficies del espacio-tiempo y de reproyectarlas al espacio-tiempo la llamaremos *imaginación*. Ella es indispensable para la generación y el desciframiento de imágenes; o, dicho de otro modo: para la capacidad de cifrar fenómenos en símbolos bidimensionales y de leer esos símbolos. El significado de las imágenes se encuentra en su superficie. Se aprehende con una sola mirada, si bien así permanece superficial. Si nos proponemos profundizar en el significado, es decir, reconstruir las dimensiones abstraídas, tendremos que pasear la mirada por la superficie, dejar que la explore. Esta exploración de la superficie de la imagen con la mirada la llamaremos "escaneo". Al escanear, la mirada sigue un rumbo complejo marcado, por una parte, por la estructura de la imagen y, por otra, por las intenciones del contemplador. ...Este espacio-tiempo propio de la imagen no es otra cosa que el mundo de la magia, un mundo en el que todo se repite y todo participa de un contexto significativo . Este mundo se distingue estructuralmente de la linealidad histórica, en la que nada se repite y todo tiene causas y acarrea consecuencias (p.11-12).

Visualizar es la capacidad de formar imágenes mentales. El pensamiento en conceptos probablemente surgió del pensamiento en imágenes y mediante la abstracción permitió la simbolización y la escritura fonética. La evolución del lenguaje comenzó en las imágenes pero hoy en día existen numerosos indicios de que es necesario un retorno hacia la imagen, en el sentido de que es importante encontrar analogías con el lenguaje que puedan aplicarse a la información visual. Esto no es nada fácil dado que para conocer el significado de las palabras es necesario conocer las definiciones comunes que comparten, y este proceso trasladado a las imágenes corre el peligro de la sobredefinición. Pero si ha sido posible descomponer el lenguaje en elementos y estructuras ¿sería posible hacerlo también con las imágenes?.

La presente investigación propone realizarlo a través de las matemáticas y aplica un método extraído del análisis automático de textos que es capaz de representar, como veremos más adelante, el contenido de las imágenes en unidades discretas de información denominadas "palabras visuales".

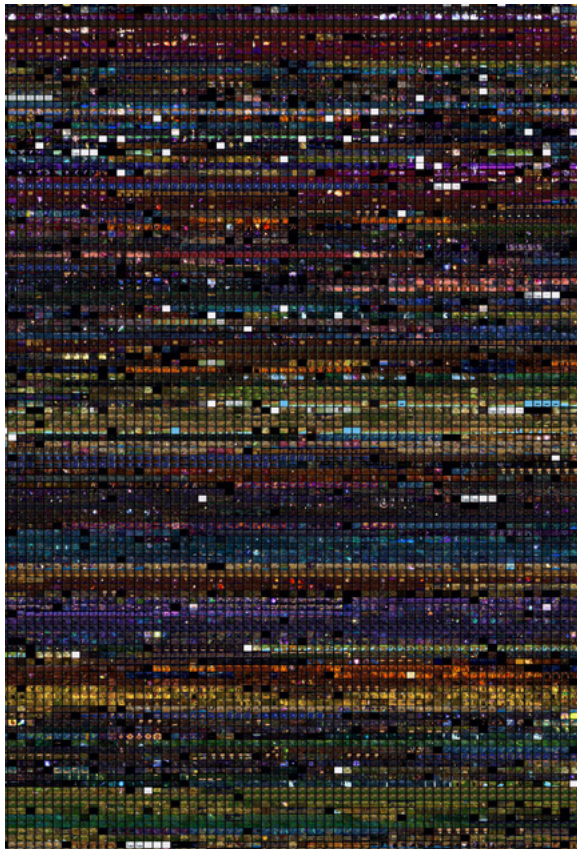


Figura 1.8. Lev Manovich y William Huber ©. 2010. 22500 frames tomados del videojuego japonés *Kingdom Hearts II* cada 6 segundos de juego. 62,5 horas de juego representadas en una imagen. Permite visualizar la estética general del mundo del juego y también los cambios visuales y narrativos dentro de la progresión de un solo juego (Manovich ,2015).

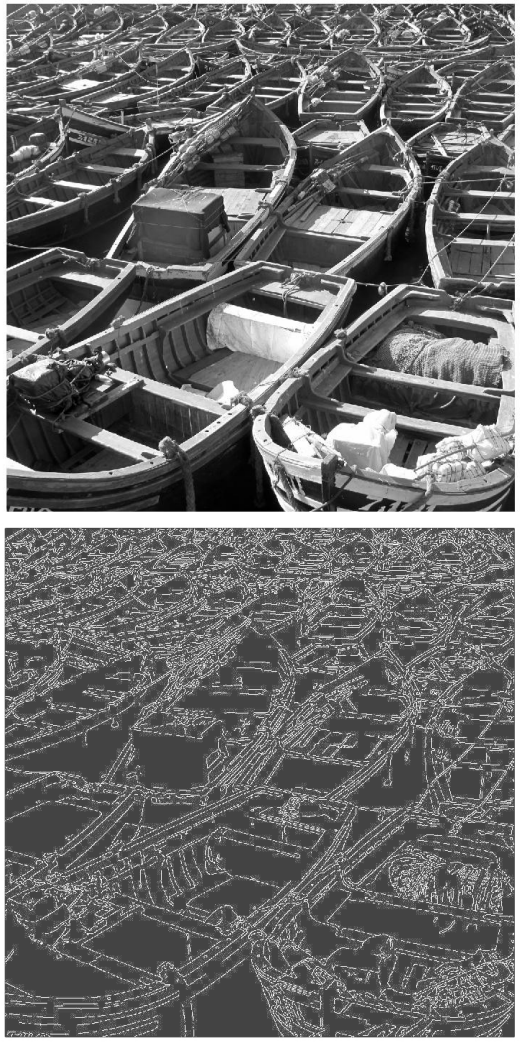


Figura 1.9. Imagen original a l izquierda e imagen de bordes físicos a la derecha.

1.4.2 Visión por computador

La visión por computador es un campo de la inteligencia artificial que se ha desarrollado en los últimos 30 años cuyos principales objetivos son:

1/ Desarrollar sistemas de comprensión de imágenes que automáticamente puedan proporcionar descripción de escenas reales.

2/ Entender la visión biológica.

David Mumford en la década de los 80 ya se enfrentaba al problema de describir matemáticamente la habilidad humana para comprender una imagen. Realmente no somos conscientes, cuando entramos en una habitación y enseguida entendemos lo que vemos, de la dificultad que esto entraña. Pero cuando se intenta construir un robot que realice esta operación constatamos de que se trata de un problema muy difícil. Una de las aportaciones de Mumford es considerar que el cerebro trabaja integrando lo que percibe en cada momento con información previa: si estas caminando por la ciudad y oyes un rugido, sabes que es muy poco probable que se trate de un tigre, así que reconoces que se trata del ruido que produce el motor de un camión. Por su parte, Mumford (2002) aplicó herramientas de cálculo a variaciones de la teoría de la visión y desarrolló modelos estadísticos en imagen y reconocimiento de patrones.

Desde los trabajos de David Marr (Marr, 1982; Marr and Poggio, 1976) es habitual considerar la visión como un sistema de procesamiento de la información que podría dividirse en diversos módulos en diferentes niveles teóricos, al menos como primera aproximación. En particular Marr sugiere que el objetivo del primer paso de la visión es obtener descripciones de las propiedades físicas de superficies tridimensionales obtenidas del entorno del observador tales como distancia, orientación, textura y reflectancia.

En la visión humana, la luz (constituida por energía electromagnética) incide en el ojo y es transformada en impulsos nerviosos por la retina. En la retina encontramos dos tipos de

células especializadas en captar la luz; los conos y los bastones, A través del nervio óptico, el impulso nervioso es transmitido al cerebro y allí finalmente, el córtex visual se encarga de dar forma y sentido a la imagen. El procesamiento de imágenes trata las instantáneas capturadas por dispositivos electrónicos para extraer de ellas mejor información y con su análisis dar solución a posibles problemas que se planteen.

La visión computacional, al modelar matemáticamente los procesos de percepción visual en los seres vivos también permite simular los procesos de estas capacidades visuales para su estudio. El enfoque de la Neurociencia Computacional (Aznar & Moreno, 2011), trata de integrar de modo coherente, en un modelo explicativo, los hallazgos convergentes provenientes de la Neurofisiología, la Psicofísica y la Inteligencia Artificial (aplicada a la visión), procurando mantener la plausibilidad biológica.

Marr (1982) estaba interesado en la realización de programas de ordenador que fueran capaces de analizar escenas de modo eficaz, haciendo uso de los procedimientos que se supone utiliza el sistema visual humano. La teoría de la visión que postula tiene como meta explicar qué etapas tienen lugar para lograr reconocer una imagen o interpretar una escena.

Desde el punto de vista computacional de D. Marr, la visión es el cálculo (realizado por diversos módulos del sistema de visión) de representaciones simbólicas sucesivas de la escena presentada al observador. Dichas representaciones deben entenderse en el sentido de descripciones explícitas de la imagen en cuestión.

Según la Teoría de la visión de D. Marr el cálculo (procesamiento) se realiza a través de dos etapas sucesivas y sólo en la segunda etapa intervienen los sistemas de conocimiento (memoria, razonamiento, etc.). Estas etapas son:

1- Procesamiento inicial o temprano, que consiste en un conjunto de procesos que intentan recuperar las propiedades físicas de la escena 3-D visible a partir de la matriz de intensidades de luminancia de la imagen digitalizada. En esta etapa se producen dos tipos de representaciones:

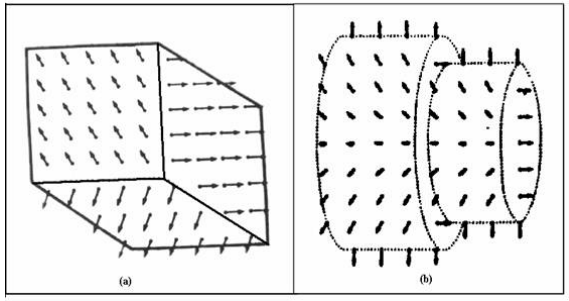


Figura 1.10. Esbozo 2 ?-D a) de un cubo y b) de dos cilindros acoplados (Tomado de Marr y Nishihara, 1978, Fig. 2).

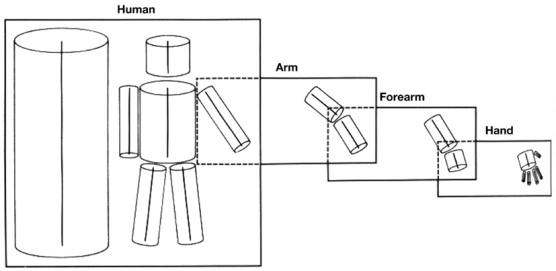
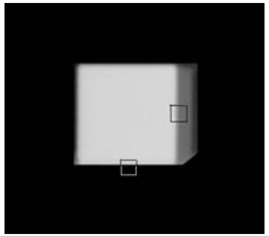


Figura 1.11. Representación 3D de una figura humana (Tomado de Marr y Nishihara, 1978, Fig. 3).



MATRIZ DE LUMINANCIA (TROZO DERECHO)

211	212	213	208	203	189	167	149	145	137	122	109	107	108	105	100
212	210	207	205	199	185	168	154	145	133	120	112	111	110	108	102
213	209	208	206	200	187	171	155	145	137	126	113	113	114	106	103
214	210	209	207	199	184	167	153	146	137	122	111	110	113	104	101
214	213	211	208	200	181	163	154	147	134	122	116	114	113	105	100
210	210	211	206	198	182	163	149	142	132	121	113	111	110	107	103
214	213	210	204	196	183	167	153	146	137	127	117	113	111	106	102
210	211	208	204	199	185	168	151	142	134	123	114	112	111	105	101
213	210	208	208	200	182	164	150	144	136	123	110	109	111	104	101
211	208	207	206	197	180	162	150	145	137	122	107	110	112	105	102
214	212	212	207	198	182	166	154	149	139	124	113	109	111	106	100
213	212	210	206	199	184	165	152	146	138	123	114	108	112	104	102
212	211	209	205	198	183	167	150	145	136	122	113	109	113	105	103
214	210	208	206	198	182	165	151	144	136	124	115	110	110	104	101
211	212	207	205	199	183	167	153	146	137	122	114	109	111	106	102
213	211	209	204	197	185	167	150	145	135	125	115	110	112	104	102

MATRIZ DE LUMINANCIA (TROZO INFERIOR)

210	212	215	214	213	214	212	211	215	215	214	213	214	213	213	212
208	211	212	210	209	211	211	210	210	210	210	211	209	208	211	210
211	213	213	209	210	214	214	210	208	210	210	209	210	213	214	212
211	213	211	210	210	211	211	209	211	208	209	210	209	209	210	211
211	212	210	210	210	211	210	209	209	211	211	211	210	208	209	210
207	208	208	208	209	207	205	205	207	208	208	207	207	207	208	206
207	206	208	208	207	207	206	206	208	211	209	209	208	207	208	206
205	206	205	205	205	205	205	204	205	207	206	205	206	205	205	204
182	180	181	181	181	181	183	184	183	181	181	183	183	183	182	181
093	093	093	093	093	094	093	093	093	093	093	093	093	092	092	093
004	001	003	006	004	007	005	000	002	001	004	003	005	002	004	003
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000

Figura 1.12. Imagen de un cubo en el que se han marcado dos áreas que contienen 2 áreas de bordes físicos y las matrices numéricas de luminancia que corresponden a estas áreas.

a- *Esbozo primario* (o bosquejo primario). Consiste en lograr una descripción constituida por un amplio número de características (líneas, bordes, manchas) tal como harían los analizadores descubiertos por Hubel y Wiesel (1959) (Fig. 1.9).

b- *Esbozo 2 ?-D* (o casi tridimensional), obtenido agrupando los elementos del esbozo primario (líneas, puntos, bordes, manchas) a fin de descubrir las propiedades de las superficies que forman la imagen o escena (Fig. 1.10).

2- *Procesamiento tardío*, que transforma el esbozo 2 ?-D en una representación identificable del objeto y sus partes constitutivas. El objetivo de esta etapa es la obtención de una representación 3D de la imagen bidimensional original, haciendo uso del procesamiento de alto nivel (Fig. 1.11).

En otros términos, el modelo teórico propuesto por Marr parece partir de un procesamiento guiado por los datos (*bottom-up*) en sus fases iniciales, para finalmente admitir el procesamiento guiado conceptualmente (*top-down*), premisa que ‘a priori’ no parece universal a algunos autores.

Aquí nos vamos a centrar en el procesamiento inicial, cuyo fin es obtener el *esbozo primario en bruto* de la imagen estimular. En dicho esbozo se representan los *bordes* físicos y su geometría, mediante la localización y caracterización de los cambios bruscos y significativos de luminancia presentes en la imagen.

Aunque este punto de vista de Marr de la visión temprana (desde imágenes a superficies) ha sido adoptado ampliamente (Barrow and Tennenbaum 1981; Brady 1982; Brown 1984; Poggio 1984), es importante destacar que su exactitud aún no está probada. En particular todavía no está clara cuál es la naturaleza del esbozo 2 ?-D, la forma en que actúan los diferentes módulos visuales, como se fusiona el resultado ni que papel desempeña el conocimiento de alto nivel en los procesos de la visión temprana. El problema critico de la organización de la visión y del control del flujo de información desde los diferentes módulos, y la forma como se utiliza el conocimiento de alto nivel sigue siendo en gran medida un problema abierto (Marroquin J. , Mitter S.; Poggio T., 1987).

1.4.3 Descripción física de la imagen

La luz que llega a nuestros ojos y es captada por los sistemas de conos y bastones, no es mas que energía electromagnética, cuyas longitudes de onda se hallan dentro del rango denominado *espectro cromático visible*. Es preciso enfatizar que no estamos refiriéndonos a una masa informe de energía lumínica, sino a una organización de la energía según una determinada configuración espacio-temporal.

Una imagen puede describirse como una distribución bidimensional de pequeños elementos puntuales, adyacentes unos de otros, y que cada uno de ellos emite una cierta intensidad luminosa. Por consiguiente, desde un punto de vista formal (matemático), la imagen estimular se define como una función bidimensional en el conjunto de los números reales, en la que a cada punto del plano con coordenadas (x, y) se le asigna el valor de luminancia emitido por dicho punto.

Una imagen digitalizada es una matriz de números enteros positivos, en la que los subíndices (fila y columna) de cada elemento indican la localización del punto en la imagen. Y los valores de los elementos representan el nivel de gris, el cual indica la intensidad luminosa en un determinado punto de la imagen.

Por tanto, cualquier imagen puede ser digitalizada y representada mediante una matriz de números, similar a la que mostramos en la Fig. 1.12. Y, recíprocamente, cualquier matriz de números (o una transformación de éstos) puede representarse como una distribución de luminancias para lograr la apariencia de una imagen. Esta matriz bidimensional de intensidades luminosas, es la primera representación que capta cualquier sistema de captura de imágenes digitales (cámara, escáner, etc.) y por tratarse de una representación cuantitativa es susceptible de ser procesada.

Cada fotorreceptor puede considerarse como un diminuto fotómetro que mide la intensidad de luz que incide sobre él y lo cuantifica en una patrón de impulsos bioeléctricos por unidad de tiempo. A partir de la matriz bidimensional de luminancia registrada por los

fotorreceptores, el sistema de visión humano es capaz de detectar, discriminar, reconocer e identificar los objetos, e incluso de dar significado a la escena. En resumen, la señal de salida de las células fotorreceptoras está discretizada tanto en sus coordenadas espaciales como en la intensidad de luz y es espacialmente variante, resultando codificada la intensidad en un patrón temporal de impulsos bioeléctricos (Fig. 1.13 y 1.14).

La definición estándar de la visión computacional es que es inversa a la óptica. El problema de la óptica clásica o del diseño gráfico es determinar las imágenes de objetos tridimensionales. La visión por computador se enfrenta al problemas inverso de recuperación de superficies a partir de imágenes. Mucha información se pierde durante el proceso de crear una imagen; proyectar un mundo tridimensional en matrices bidimensionales . Por consiguiente la visión debe ser consciente de sus limitaciones naturales.

Así, a partir de la comprensión de las primeras etapas de la visión, se pone de manifiesto la idea original de D. Marr de que la visión es, esencialmente calculo, operaciones aplicadas a representaciones, que transforman sucesivamente la imagen que captó la retina, es decir: neuro-computación. Sin olvidar, que en el llamado 'procesamiento visual tardío' se añadirá un plus de información a la generada en la etapa de procesamiento inicial o temprano, también referida como 'etapa de bajo nivel'.

Así, el reconocimiento de objetos visualmente es una función clave del cerebro de los primates. Se toleran cambios considerables en las imágenes, tales como los producidos a causa de la iluminación variable, por diferentes ángulos de visión y rotaciones del objeto; también es capaz de realizar generalizaciones automáticas.

Los mecanismos neurales de reconocimiento de objetos visuales se han investigado tanto en estudios de pacientes con daño cerebral como en monos con lesiones experimentales. El desarrollo reciente de técnicas de medición no invasivas para examinar el cerebro humano ha proporcionado nuevas y potentes herramientas, que han aumentado la velocidad de exploración.

Tanaka (1993) revisa las fronteras de la investigación en este campo, a partir de estudios

con monos y después de pasar a estudios en humanos. Concluye (Tanaka, 1997) que hay bastantes evidencias de que el reconocimiento de objetos se realiza en las neuronas del córtex temporal inferior y que se reconocen características de complejidad intermedia. Estas características son típicamente invariantes a una amplia gama de cambios de ubicación, escala, e iluminación, mientras que son particularmente sensibles a combinaciones de forma local, color y propiedades de textura.

Lowe en el año 2000 describe un sistema de visión por computador para realizar el reconocimiento de objetos que también hace uso de las características locales de complejidad intermedia de la imagen que son invariantes a muchos parámetros. Los denomina descriptores *SIFT* (Scale Invariant Feature Transform); características invariantes a transformaciones de escala y con ellos consigue transformar una imagen en una representación que no se ve afectada por los cambios de escala y otras transformaciones similares.

Este proceso consigue la integración de las características de una manera similar al proceso de atención visual en serie que se ha demostrado que desempeña un papel importante en el reconocimiento de objetos en la visión humana.

El propio Lowe, en 2004, encuentra que la mejor solución de compromiso entre rendimiento y rapidez se obtiene usando una cuadrícula de muestreo de gradientes de 16 x 16 y agrupando los histogramas en 4 x 4. El descriptor final propuesto en esta formulación es 128 dimensional (4 x 4 x 8).

Los descriptores *SIFT* descritos por Lowe, que serán comentados con más detalle en el capítulo 2 de la metodología de la presente tesis y también, a nivel más técnico, en el Anexo A, son unos de los más destacados descriptores locales de puntos de interés de una imagen utilizados en visión por computador, y por este motivo son los que se ha decidido utilizar principalmente en la presente investigación para lograr los objetivos propuestos.

Esta manera de representar la imagen numéricamente da paso a los próximos comentarios acerca de la relación entre la matemática y las formas.

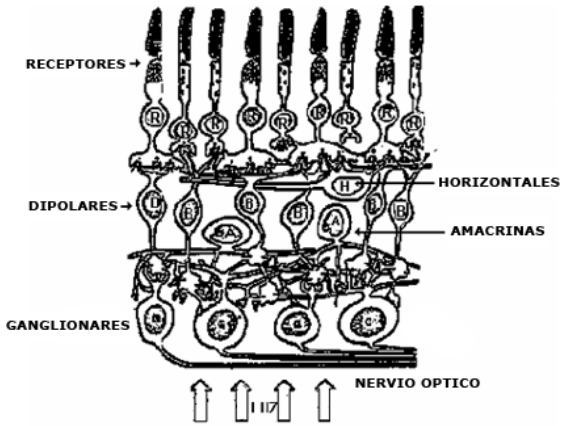


Figura 1.13. Histología de la retina. Células que la componen.

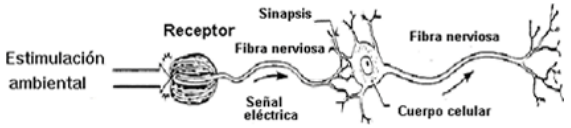


Figura 1.14. Transmisión del impulso nervioso desde la célula receptora al córtex.

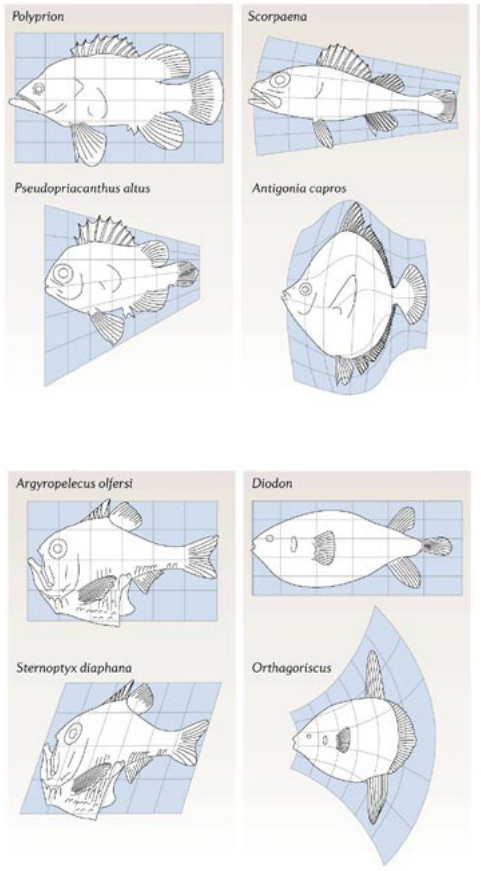


Figura 1.15. Permutaciones y transformaciones de formas de peces que pueden explicarse con cambios de sistemas de coordenadas, algunos sencillo y otros sorprendentes e inesperados (Wentworth, D. ,1980) © Nature Publishing Group (Nature Reviews. Genetics)

1.5 LA FORMA

Para abordar el problema de la forma y el contenido, tal y como interesa en el presente estudio, recurriremos a diversos antecedentes como el biólogo y matemático D'Arcy Wentworth, el estructuralismo, la Gestalt o la entropía de Shannon:

1.5.1 Matemática y Forma

A D'Arcy Wentworth Thompson (Wentworth, D. (1980) le interesaba principalmente la explicación del crecimiento y la forma biológica en términos físico-matemáticos. D'Arcy se diferenciaba del resto de científicos experimentales, como los embriólogos quienes intentaban comprender la forma a partir de la configuración de los antecesores genealógicos inmediatos, porque él se conformaba con una descripción matemática o una analogía física, independientemente de la línea evolutiva que hubiesen seguido las especies comparadas. Le interesaba la relación entre las matemáticas y la forma, no sólo las formas adoptadas por la materia en todos los aspectos, sino incluía a todas las formas teóricamente imaginables (Fig. 1.15).

El estudio de la forma se puede abordar desde lo meramente descriptivo o puede ser analítico: comenzar por describir la forma de un objeto con palabras sencillas para terminar definiéndolo en lenguaje matemático. Por ejemplo, la forma de la tierra o la de una gota de lluvia pueden describirse más o menos con palabras usuales, pero cuando hemos aprendido a definir la esfera hemos realizado un avance considerable. En palabras de Wentworth:

> La definición matemática de una forma tiene una cualidad de precisión que faltaba por completo en nuestra primera descripción; está expresada en pocas palabras o en símbolos aún más breves, y estas palabras o símbolos están tan repletas de significado que se ahorra esfuerzo mental...
> Podría pensarse que las definiciones matemáticas son demasiado estrictas y rígidas

> para el uso corriente, pero su rigor está combinado con una libertad casi infinita. La definición de una elipse nos introduce a todas las elipses del mundo (p.260).

Para D'Arcy Wentworth, mediante la acción combinada de las fuerzas apropiadas, cualquier forma material puede transformarse en cualquier otra.

En el proceso que nos ocupa de estudio de obras de arte, interesa especialmente de este autor su idea de que debemos aprender de los matemáticos a eliminar y descartar, a mantener en la mente el arquetipo, porque en ese sacrificio de lo superfluo podemos encontrar la esencia de la analogía entre las formas y nos puede ser útil para establecer relaciones, a pesar de las diferencias entre las formas comparadas.

1.5.2 La Sintaxis Visual

Es seguro que existe una sintaxis visual, líneas generales de construcción de composiciones, elementos básicos y mensajes visuales que se pueden comprender y aprender, seas artista o no (Dondis, 1984). Captamos información visual de muchas formas y a ello le afecta tanto la fisiología perceptiva como nuestro propio movimiento o estado de ánimo. A pesar de que existen diferencias a nivel individual y colectivo (cultural), existe un sistema perceptivo visual que todos los seres humanos compartimos.

Muchas disciplinas han abordado el problema del significado en las artes visuales. Artistas, filósofos, historiadores del arte y otros especialistas de las ciencias humanas y sociales han abordado el problema de cómo y qué comunican las artes. A finales del siglo XIX los psicólogos de la *Gestalt* realizaron aportaciones muy interesantes en este campo. Frente al estructuralismo atomista que proponía un análisis del estímulo en sus elementos constituyentes, el holismo de la Gestalt propone que el todo es superior y no reducible a la suma de las partes constitutivas (Köhler, 1947; Koffka, 1967).

El término *Gestalt* se traduce literalmente como *forma*; sin embargo tiene la connotación

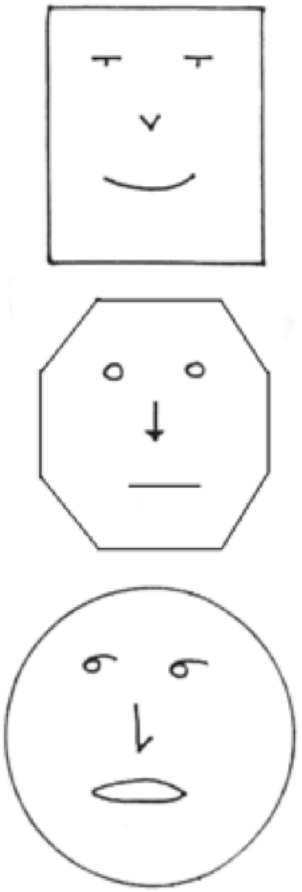


Figura 1.16. Esquemas diferentes de rostros que ilustran la teoría de la Gestalt porque, a pesar de ser diferentes tienen en común el percepto "caricatura de rostro".

de *estructura configuracional* y define el enfoque adoptado por esta escuela, que se centró en el problema de la organización perceptual, desarrollando ingeniosos experimentos y demostraciones originales de numerosos fenómenos perceptuales.

La Psicología de la Gestalt propuso que la experiencia perceptiva tiene un carácter organizado y constituye una estructura de elementos ordenados jerárquicamente, de modo que, en función de dicha jerarquía, quedan determinadas las características de configuración, actualidad y significado. El principio básico de la organización perceptual es que el todo es más que la suma de las partes, es decir, que las propiedades de la totalidad no resultan de los elementos constituyentes, sino que emergen de las relaciones espacio-temporales del todo.

En la Fig. 1.16 se muestra cómo, a pesar de que los elementos del estímulo difieren en los tres casos, en todos ellos emerge el percepto "caricatura de rostro". Este tipo de propiedades de las formas son fundamentales a la hora de analizar los resultados del presente trabajo de investigación, pues están íntimamente relacionados con los aspectos latentes que se explicarán en detalle más adelante (apartado 4 del Anexo A).

1.5.3 La organización perceptiva

Pasaremos a comentar la organización perceptiva según Aznar (2015). Es una preocupación antigua discernir si percibimos desde el detalle a la totalidad o viceversa. En definitiva el problema es determinar si la unidad básica de la percepción es el todo o son las partes.

Los individuos no tenemos conciencia de percibir los estímulos de forma aislada (luz, sonido, temperatura, etc.), sino organizados en estructuras perceptuales (formas, objetos, escenas, secuencias, etc.). Aquí surge el problema relativo a como la estimulación, a través de determinados procesos, se organiza formando un todo significativo.

1.5.3.1 El estructuralismo

El iniciador de la psicología científica, W. Wundt (1832-1920) distinguía tres contenidos de la conciencia: 1) Sensaciones; fenómenos mentales resultantes de la elaboración posterior a la estimulación de los órganos de los sentidos, 2) Imágenes; sensaciones experimentadas sin presencia del estímulo ni estimulación sensorial y 3) Sentimientos; emociones referidas al mundo subjetivo.

Las sensaciones se combinan mediante la atención y determinados principios de conexión sensorial formando agregados. Las imágenes procedentes de experiencias previas también forman parte del agregado. Distinguía en la conciencia el campo y el foco, los cuales determinaban los estados de conciencia, según los contenidos cayesen bajo el foco de la conciencia (reciben atención) o estuviesen fuera de él. Toda percepción siempre posee un significado para el sujeto. Para Wundt, la conciencia era un flujo permanente en continua actividad y cambio.

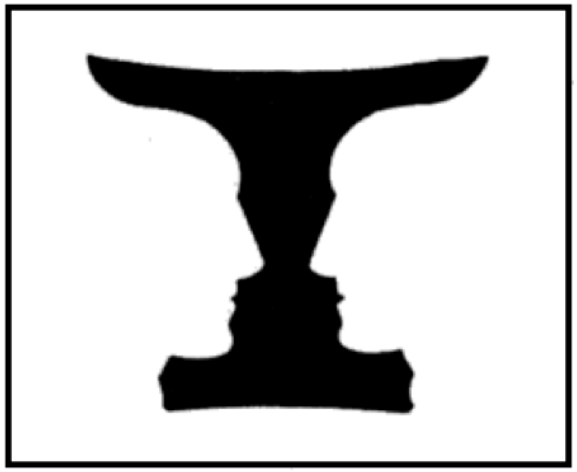


Figura 1.17. Figuras ambiguas (reversibles fondo-figura) Creyeron demostrar que toda percepción se basa en la organización figura-fondo; puesto que los sujetos no pueden percibir, en dicha figura ambigua, las dos figuras a la vez.

1.5.3.2 La psicología de la Gestalt

La percepción se organiza y estructura de modo innato, entre la forma subyacente a los procesos neurofisiológicos y las experiencias perceptuales. Dichos procesos son entendidos como 'campos de fuerza', que interactúan y mantienen un equilibrio del que resulta una totalidad o configuración; el cambio de una parte modifica a las demás. La acción de estas fuerzas organizativas determinaban que el todo sea algo más y distinto de la suma de las partes. Este campo perceptivo queda determinado por una serie de leyes (más de 100) de las que citaremos algunas:

1- Figura-Fondo: En síntesis, los psicólogos de la Gestalt señalaron que la percepción está organizada, que no percibimos elementos independientes unos de otros, El primer estadio en la organización perceptual era la configuración de totalidades, que constan de dos

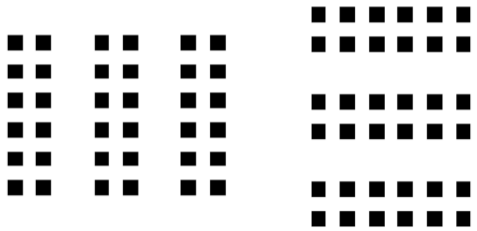


Figura 1.18. Principio de proximidad espacial.

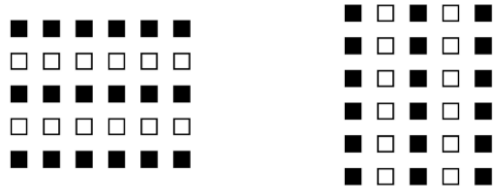


Figura 1.19. Principio de similitud (acromática).

componentes:

a- Una parte más estructurada y bien delimitada, denominada figura.
b- Otra parte indiferenciada y periférica que captamos de modo difuso, denominada fondo.

Figura y fondo presentan un contraste. Esta configuración figura-fondo se halla necesariamente presente en cualquier percepción y conduce a la percepción de objetos que se destacan de un fondo (Fig. 1.17). Esta es la ley más básica que conduce a que un objeto (figura) se destaque sobre un fondo difuso. Los experimentos de Rubin en 1921 demostraban las relaciones entre ambas. Las principales características que destacó Rubin las sintetizamos en la siguiente tabla:

Figura	*Fondo*
- Tiene forma, contorno.	- Es difuso, informe.
- Sobresale en primer plano	- Queda en segundo plano
- Adquiere significado	- No es significativa
- Colores densos y sólidos	- Colores diluidos
- Se recuerda mejor.	- El recuerdo es menor.

2- Primacía: Afirma que el todo es más originario, primario y se manifiesta antes que las partes.

3- Autonomía: Afirma que el todo queda determinado por factores internos más que por factores externos.

4- Buena forma: La percepción se organiza de modo que las figuras aparezcan lo más simples, regulares y simétricas que sea posible. Cuando la figura permite descripciones alternativas se percibe la más simple.

5- Principio de proximidad: En igualdad de condiciones, los estímulos más próximos (en el espacio y en el tiempo) tienden a percibirse formando parte de un mismo "todo" perceptual (Fig. 1.18).

6- Principio de similitud: En igualdad de condiciones, los elementos estimulares más semejantes tiende a percibirse formando parte de un mismo "todo perceptual" (Fig. 1.19).

7- Principio de la buena continuación o dirección: En igualdad de condiciones, tendemos a percibir, formando parte de una misma figura, los estímulos que guardan entre sí una continuidad. Es decir, se agrupa según una continuidad suave, más que según cambios bruscos (Fig. 1.20).

8- Principio de cierre o clausura: Como consecuencia de la anterior, una figura incompleta se tiende a percibir como si fuese completa (Fig. 1.21).

9- Principio del tamaño relativo: En igualdad de condiciones, el área estimular más pequeña tiende a articularse como figura (Fig. 1.22.A).

10- Principio del área envolvente y envuelta: El área envolvente suele articularse como fondo, y el área envuelta como figura (Fig. 1.22.B).

11- Principio de simetría: En igualdad de condiciones, las áreas simétricas tienden a articularse como figuras y las asimétricas como fondo (Fig. 1.22.C).

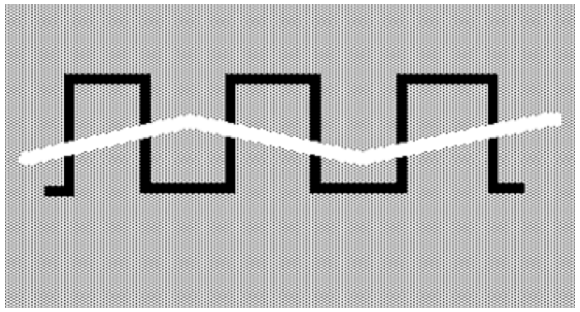


Figura 1.20. Principio de la buena continuación.

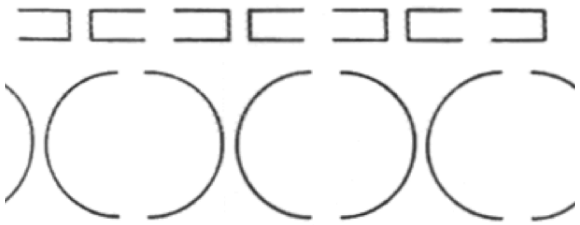


Figura 1.21. Principio de la buena continuación.

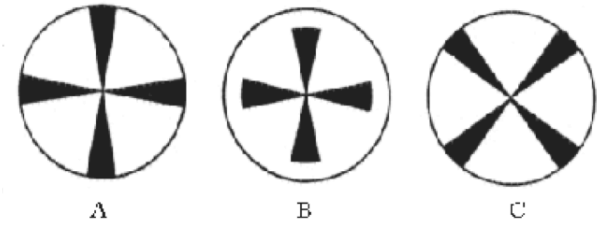


Figura 1.22. A. Principio del tamaño relativo. B. Principio del área envolvente y envuelta, C. Principio de simetría.

1.5.3.3 Estructuralismo versus Gestalt

Para los estructuralistas los datos primarios son los elementos y lo secundario, obtenido por aprendizaje asociativo, es la constancia del tamaño, forma, color, etc.; mientras que, para la psicología de la Gestalt la experiencia fenomenológica de la constancia es el dato primario organizado y estructurado, los elementos son derivaciones secundarias segrega-

das por abstracción.

Las principales críticas a la teoría de la Gestalt son:

1- No probaron sus leyes experimentalmente, ya que utilizaban un método fenomenológico consistente en mirar la figura y verla por sí misma. Esta técnica no informa del proceso perceptivo ni de los parámetros de las leyes.

2- Sólo aplicaron las leyes a representaciones bidimensionales de estímulos geométricos, no a objetos tridimensionales de la vida real.

3- Experimentalmente, desde la neurofisiología no se han encontrado evidencias de los supuestos responsables de esta organización perceptual.

4- La cuarta crítica se dirige hacia la idea de que sólo intervienen las experiencias pasadas y las interpretaciones de los datos sensoriales cuando el estímulo no se halla claramente estructurado o es ambiguo (nubes, manchas de tinta, etc.). Tampoco considera la influencia del contexto en la percepción, informándonos de qué patrones son más probables de hallar en cada contexto particular.

A pesar de las críticas, existe un creciente y renovado interés por el punto de vista gestáltico y la problemática que plantearon, aunque ahora se abordan desde una metodología experimental.

1.5.4 Forma e información: Entropía de Shannon

Entendiendo el conjunto de imágenes digitalizadas de la obra de un artista como un mensaje, como el conjunto de información que el artista envía al espectador a través del medio de su obra, haremos referencia a la entropía de Shannon (ver apartado 8 del Anexo A para detalles técnicos sobre su cálculo) que se utiliza en la presente tesis para establecer dife-

rencias en el grupo de obras analizadas.

En 1948, Claude Shannon y Warren Weaver proponen una teoría de la información sobre las leyes matemáticas que actúan en la transmisión y el procesamiento de la información. Esta teoría es una rama de la teoría matemática y de las ciencias de la computación que, entre otras cosas se ocupa de la medición de la información y de la representación de la misma. La idea es garantizar que el transporte masivo de datos no suponga una pérdida de calidad, incluso si los datos se comprimen de alguna manera.

Un concepto importante en la teoría de la información es que la cantidad de información que contiene un mensaje es un valor matemático medible. La cantidad no se refiere a la cuantía de datos, sino a la probabilidad de que un mensaje, dentro de un conjunto de mensajes posibles, sea recibido. El valor más alto se le asigna al mensaje que menos probabilidades tiene de ser recibido. Si se sabe con certeza que un mensaje va a ser recibido, la cantidad de información que contiene es cero.

> Por su naturaleza, los mensajes son una forma y una organización. Efectivamente es posible considerar que su conjunto tiene una entropía como la que tienen los conjuntos de los estados particulares del universo exterior. Así como la entropía es una medida de desorganización, la información que suministra un conjunto de mensajes, es una medida de organización. De hecho puede estimarse la información que aporta uno de ellos como el negativo de su entropía cuanto más probable es el mensaje, menos información contiene (Wiener, 1988, p. 21).

El nivel de información de una fuente se puede medir según la entropía de la misma. El índice de Shannon o Shannon-Wiener fue creado originalmente para ser usado como medida de entropía en cadenas de caracteres (por ejemplo de texto), en el contexto de la teoría de la información y se utiliza en ciencias como indicador de biodiversidad. Este índice se utiliza en la presente tesis para organizar el conjunto de imágenes estudiadas. Se aplica, como se verá más adelante, a la representación de las imágenes en forma de distribución de probabilidad de los aspectos visuales que contienen (Fig. 3.19 y 3.20) y resulta de gran utilidad para ordenar la colección de cara a su procesamiento.

Figura 1.23. Gilbert Garcin. 1996. Le parvenu. Collage fotográfico. © (Gilbert, 2013)

Para un teórico de la información toda regularidad predecible es redundante porque a él lo que le preocupa es la economía, pero no se puede decir lo mismo en arte. El arte constantemente tiene en cuenta la estructura, y en este contexto, la regularidad de la forma no es una redundancia, no disminuye la información. Pensemos por ejemplo en una exposición de Andy Warhol; una fotografía en filas de reproducciones idénticas con el objeto de explorar las connotaciones de la multiplicación mecánica como fenómeno de la vida moderna. La palabra información, en sentido literal significa "dar forma" y la forma necesita estructura (Arnheim,1980). Por tanto, somos conscientes de que la aplicación de la teoría de la información a las artes visuales reduciendo la forma estética a mediciones cuantitativas sin tener en cuenta relaciones estructurales debe hacerse con suma cautela. En la presente tesis se utilizan los resultados del cálculo del índice de Shannon con el ánimo de comprender y ordenar, con la intención de hacer más accesibles al observador la estructura organizada de formas y colores, y de esta forma poner de manifiesto el orden subyacente.

1.6 EL ARTISTA

Dando por sentado que no es la intención del presente trabajo abarcar la riqueza comunicacional de las obras de los artistas, tema muy complejo que ya hemos avanzado que estaría condicionado por múltiples factores como la psicología de la percepción estética y la relación que se establece entre la obra y el espectador, entre el que mira y el que es mirado (Hildebrand, 1988), de lo que sí estamos convencidos a la vista de los resultados obtenidos es de que estos sofisticados sistemas informáticos nos pueden ayudar a vislumbrar la cantidad y complejidad de procesos que lleva a cabo la mente del creador visual que persigue un objetivo estético a la hora de decidir si una imagen o una obra acabada pertenece o no a su vocabulario visual.

Considerando que la mirada hacia el mundo es un juego de equilibrio entre las propiedades del objeto observado y la naturaleza del sujeto que observa, topamos con la preocu-

pación acerca de la psicología de la percepción. La complejidad para distinguir entre lo que realmente vemos y lo que inferimos a través de la inteligencia es tan pretérita que ya Plinio dejó constancia del pensamiento en la antigüedad clásica al escribir: "la mente es el verdadero instrumento de la visión y la observación, y los ojos sirven como una especie de vasija que recibe y transmite la porción visible de la conciencia" (Gombrich, 1998, p. 12).

Si a esta consideración sumamos que nuestro objeto de análisis lo constituyen imágenes producidas por artistas en su proceso de creación, por lo tanto, que ya son interpretaciones del mundo en los términos de los esquemas que ellos conocen, y añadimos también que al analizarlas nos situamos en el lugar del observador, que en la lectura de estas imágenes empatiza y colabora con el artista en transformar un pedazo de pantalla con píxeles coloreados en un parecido con el mundo visible, podemos concluir de antemano que la problemática es inabordable en un proyecto de investigación como el que aquí se propone.

Los artistas son generadores especializados de imágenes, en su proceso creativo las producen constantemente. Estas pueden formar parte, tanto de las fases más iniciales y previas de este proceso como convertirse en el resultado final de su trayecto artístico; pueden ser tanto la herramienta como el producto de su trabajo.

El punto de partida de nuestro análisis es el convencimiento de que entre las obras de las colecciones de imágenes de artistas existen unos lazos de unión, de parentescos formales que hacen que constituyan una familia de significado común. El artista visual hace uso de sus categorías formales para capturar desde lo particular aquello universalmente significativo, de una forma necesariamente personal. Las palabras textuales de Josep Mª Jori (Reverter et al. 2013) ilustran a la perfección la importancia que tiene para un artista visual su colección estética:

> Son vitales para todo artista creador, las colecciones de formas taxonómicamente vinculadas por analogías de sentido, es decir por un principio intuitivo que relacione expresiones formales diferentes y las vincule por intuición estética o sea, sensible a los significados, que no se reduce a lo sensitivo sino a un intelecto sensible que en el artista creador se reconoce por resonancia visual y que interrelaciona instantáneamente y en

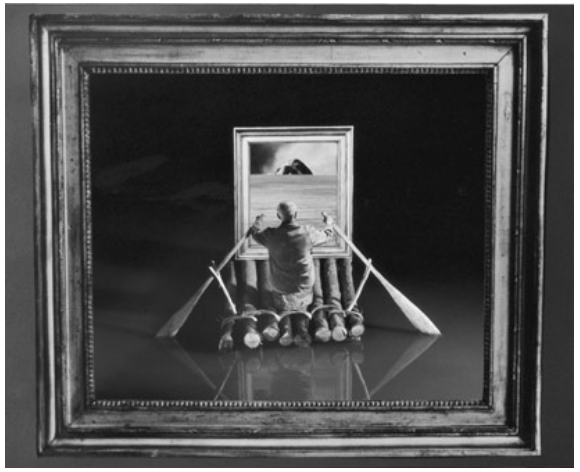


Figura 1.24. Gilbert Garcin. 1998. Le Cap de Bonne Espérance. Collage fotográfico. © (Gilbert, 2013).

Figura 1.25. Gilbert Garcin. 1999. Au musée Collage fotográfico. © (Gilbert, 2013).

Figura 1.26. Alberto Durero 1494-5. Estudio de tres manos. Pluma y tinta marrón sobre un dibujo preliminar a pluma sobre papel. © Graphische Sammlung Albertina (GuggheimBilbao, 2015).

> sincronía las formas e imágenes del entorno con las necesidades y pulsiones de su ser interno en un solo cuerpo de percepción (p. 9).

Kandinsky (1987), además de ser un artista apasionado, siempre se mostró deseoso de explicar las razones primeras y profundas de la creación artística y realizó aportaciones fundamentales que, además de contribuir a esclarecer el análisis de los elementos esenciales del quehacer pictórico, contribuyeron a la búsqueda de un método genérico para las investigaciones de las ciencias artísticas (Kandinsky, 1996).

> También existe la posibilidad de penetrar en la obra, participar en ella y vivir sus pulsaciones con sentido pleno. Y aunque no se tenga en cuenta su valor científico, que depende de un minucioso examen, el análisis de los elementos artísticos es un puente hacia la pulsación interior de la obra de arte (p. 15-16).

Indagando en el testimonio de los propios artistas referido a la belleza y a su creación, encontramos también las declaraciones de Mark Rothko[4] (2004, p.110) en relación a los testimonios que dan en sus escritos ya Leonardo y Durero de que la pintura es básicamente un espejo, dando a entender que el espejo no sólo puede retratar las apariencias, sino que retrata los aspectos más profundos de las apariencias.

Según Rothko, estos maestros tienen que estudiar todos los objetos que retratan para así retratarlos correctamente. Una de las estrategias de Leonardo para identificar los principios claves era hacer una analogía entre los elementos de la estructura profunda de diferentes sistemas. Así el artista, dependiendo de su talento, es capaz de retratar las verdades internas al registrar y representar las profundas sutilezas de esas apariencias externas. Y tanto Leonardo como Durero escriben que para construir una figura deben examinar numerosas figuras y elegir entre ellas aquellos miembros y órganos que les parezcan más perfectos, y

4 Marcus Rothkowitz (Daugavpils, Letonia, 25 de septiembre de 1903-Nueva York, Estados Unidos, 25 de febrero de 1970), conocido como Mark Rothko, fue un pintor y grabador nacido en Letonia, que vivió la mayor parte de su vida en los Estados Unidos. Ha sido asociado con el movimiento contemporáneo del expresionismo abstracto, a pesar de que en varias ocasiones expresó su rechazo a la categoría «alienante» de pintor abstracto.

entonces combinarlos en una figura sintética que sea símbolo de perfección. Rothko ya es consciente, y así lo declara en sus escritos, de que con esta forma de trabajar introducen un valor estadístico en su noción de perfección, pues afirman que la consideración de perfección en esos miembros debe resultar de un consenso general. La posición de Leonardo y Durero sobre lo que se considera generalmente bello contiene los gérmenes del método estadístico. Así cada artista hace sus propias transformaciones, o digamos sus propias distorsiones para alcanzar la perfección.

De la mano de las vanguardias y particularmente de Duchamp la obra de arte renuncia a ser espejo del mundo y se manifiesta como lugar de acontecimiento, materia sobre la que ocurren cosas. El arte entonces es como la vida porque es materia de sucesos, así rescata la verdad latente, originaria del acto creador del azar (Méndez, 1992). El artista ya no ejecuta encargos siguiendo las reglas establecidas por su cliente y su propio arte. Ahora ya no hay reglas. El artista las va estableciendo conforme las crea, y también por esto la obra puede ser vista como un acontecimiento.

Recurriremos también a las declaraciones de otro artista, Joan Fontcuberta[5] (2010) que también se cuestiona acerca de los patrones de mímesis gráfica y su posible relación con la matemática:

> Tanto desde la filosofía del arte como desde la semiótica se ha producido un esfuerzo para diagnosticar los rasgos que en una imagen permiten identificar al objeto representado. ¿Se trata de patrones basados en una mímesis gráfica objetiva y universal o por el contrario dependen de sistemas de representación culturales y objetivos? Una multiplicidad de hipótesis ha dado respuesta a estas cuestiones que en el fondo vienen impregnadas de una incertidumbre más profunda: la que atañe a nuestros modelos de construcción de la realidad... ¿Seria hoy posible diseccionar el concepto de semejanza según un criterio de lógica matemática? (p. 83).

Figura 1.27. Leonardo da Vinci 1505. Estudio de caballos. © Real colección del Castillo de Windsor. (WahooArt, n.d).

5 Joan Fontcuberta i Villà (24 de febrero de 1955, Barcelona) es un artista, docente, ensayista, crítico y promotor de arte español especializado en fotografía, premio David Octavious Hill por la Fotografisches Akademie GDL de Alemania en 1988, Chevalier de l'Ordre des Arts et des Lettres por el Ministerio de Cultura en Francia en 1994, Premio Nacional de Fotografía, otorgado por el Ministerio de Cultura de España en 1998 y Premio Nacional de Ensayo en 2011.

En nuestros días, a la complejidad del escenario que hemos descrito contribuye también la saturación de imágenes a la que nos vemos sometidos: la preeminencia de internet, las redes sociales, el abaratamiento de las cámaras digitales y su implantación en los teléfonos móviles, etc., contribuyen a la ubiquidad de la imagen en nuestra realidad.

Esta inmediatez y estas prisas favorecen también que nuestra mirada se haya vuelto más superficial. Como nos apunta el artista visual Àlex Nogué[6] (2013):

> L'omnipresència de les imatges a la vida moderna ens ha fet desenvolupar uns mecanismes perceptius capaços d'actuar amb gran rapidesa. Hem après a captar les imatges immediatament i, emprant un sistema que fa intervenir la nostra potent memòria visual, en poques fraccions de segon som capaços de discernir-ne algun significat. Això només indica que el que s'ha vist encaixa amb el que ja havia estat vist i que encara persisteix en l'arxiu visual personal, però no vol dir que realment hàgim vist la imatge[7] (p. 158).

Pero nuestro punto de partida son imágenes de artista. Considerando el arte como una forma de conocimiento que elabora imágenes a partir de los sucesos del mundo, los objetos de los que parte nuestro análisis encierran una enorme complejidad sobre la que se ha escrito mucho pero que aún está por conocer. "Considero el arte como una forma de conocimiento basado en el principio de comunicabilidad de complejidades no necesariamente inteligibles" (Wagensberg, 1985, p.110).

En este trabajo de investigación nuestro objetivo no consiste en asociar la realidad tridimensional del mundo a las obras que analizamos, sino que damos por supuesto que el artista, obsesionado por cierta complejidad o temática, actúa con la esperanza de capturar

6 Àlex Nogué (10 de marzo de 1953, Hostalets d'en Bas) es Doctor en Bellas Artes y Catedrático de pintura de la Universidad Barcelona.

7 La omnipresencia de las imágenes en la vida moderna nos ha hecho desarrollar unos mecanismos perceptivos capaces de actuar con gran rapidez. Hemos aprendido a captar las imágenes inmediatamente y, empleando un sistema en el que interviene nuestra potente memoria visual, en pocas fracciones de segundo somos capaces de discernir algún significado. Esto sólo indica que lo que se ha visto encaja con lo que ya había sido visto y que aún persiste en el archivo visual personal, pero no significa que realmente hayamos visto la imagen (traducido por la autora).

al menos una parte de este conocimiento y que posteriormente esta información o parte de ella, puede ser revelada al espectador mediante un análisis matemático.

Y otro tema es ¿Quién debe recibir la obra de arte? En principio el propio artista. Así para que exista cierta coherencia en el conjunto de la obra el propio artista debe reconocerla. El arte empieza comunicando al propio artista y luego se extiende a los demás. Para estos fines los artistas se inventan un lenguaje, lo cual sirve para complicar aún más las cosas. Somos conscientes de la audacia de nuestro propósito al pretender que el algoritmo que hemos construido será capaz de desvelar al menos una porción de toda esta complejidad, pero este es el motivo por el cual esta tesis se plantea desde el ámbito artístico y no desde el científico. Compartimos con el arte esta osadía, y como científicos reconocemos esta gran limitación que nos asegura de antemano modestos resultados, pero nos impulsa la emoción de descubrir aunque sea una pequeña parte del conocimiento que estamos convencidos que atesora el arte. Y lo que se sí se pretendería es abrir una linea de debate entorno a lo que esta categorización implicaría con respecto a la relación entre el artista, su obra y el espectador.

Nuestro propósito tiene que ver con el gozo de encontrar la inteligibilidad en la belleza (Wagensberg, 2007, p.137). Al redactar esta introducción a posteriori de realizar los experimentos, ya podemos transmitir al lector la emoción que proporciona el ver que realmente hay grupos de significados o clases, que las obras se agrupan en el conjunto de un artista consolidado. Y que sólo el intento que nos hemos propuesto ya significa un avance. Jorge Wagensberg (2007) lo describe muy bien con sus palabras:

> Clasificar es una forma de inteligibilidad siempre y cuando, eso sí, haya más piedras que clases (...) Supongamos que tenemos una buena clasificación de piedras (la teoría) y un río largo cuyo cauce es una fuente generosa de cantos rodados (la experiencia). Cada nueva piedra supone un chispazo entre teoría y experiencia. Pueden ocurrir 4 cosas:
> 1) una piedra encaja en una sola clase y la teoría vigente se confirma,
> 2) una piedra no encaja en ninguna clase (paradoja de incompletitud) y hay que ampliar la teoría vigente,
> 3) una piedra encaja, con igual derecho, en dos clases diferentes (paradoja de contradicción) y hay que corregir la teoría vigente,

Figura 1.28. Panel número 49 ('Sentimiento contenido del triunfo. Mantegna') del Atlas Mnemosyne de Aby Warburg (Warburg, 2010).

> 4) una clase permanece vacía
> (...) En cualquiera de estos casos se gana conocimiento. En cualquiera de estos casos se acelera el pulso del investigador (p.138).

1.7 EL DISCURSO

Las colecciones no son simples acumulaciones, sino más bien recolecciones de significados reconocidos del propio entorno. Von Goethe (1790) fue quizá el primero que relacionó metamorfosis con plantas y publicó un ensayo sobre su morfología intentando establecer una analogía entre ciertas formas. Aunque su preocupación era filosófica y lo que quería era encontrar la planta original, el arquetipo formal del que posteriormente surgen el resto de linajes.

Ya en el siglo XX, el historiador del arte y gran coleccionista de imágenes, Aby Warburg, con una visión premonitoria considera que las imágenes por sí mismas y en su relación mutua y cambiante generan un espacio de pensamiento. Su proyecto inacabado Atlas Mnemosyne (Warburg, 2003) es un esfuerzo por comprender el recorrido de los contenidos culturales lejos de la cronología y el formalismo de Wölfflin. Warburg organiza sobre tablones de madera cubiertos con un paño negro fotografías de imágenes, reproducciones de libros y materiales visuales de periódicos y/o de la vida diaria, de tal manera que ilustran una o varias áreas temáticas (Fig.1.28 y 1.29). Hoy en día, el estilo de trabajo de Warburg se calificaría como investigación mediante agrupaciones visuales. Las imágenes no están ordenadas de acuerdo a la similitud visual evidente, en el sentido de una historia iconográfica del estilo; sino más bien mediante relaciones de una afinidad entre sí y del principio de buena compañía.

> Fue la idea salvadora de Aby Warburg ante las dificultades de poner por escrito su complejísimo mundo. Cómo una historia del arte o historia de la cultura sin texto posibilita "verlas" examinando multitud de imágenes a la vez, ya con la idea revolucionaria además de que no es necesario observar originales. Fue su modo de localizar el pensar

> en un espacio visual dinámico siempre cambiante, mudable, en una aventura exegética siempre abierta, infinita, como un desafío también al supuesto orden del tiempo (Reguera, 2010).

Los museos virtuales, banco de datos o redes de mapas que primen el esquema o imagen sobre el lenguaje son en esencia warburgianos: el atlas del visionario Aby Warburg tiene ya una disposición similar a la de una página de Internet.

Pero muchas veces resulta difícil expresar con palabras lo que se percibe en una imagen. Es normal que podamos ver y sentir sus cualidades sin poder explicarlas. En cierta forma, la comprensión de las obras de arte se expresa a través de los discursos de miembros del mundo del arte (artistas, museos, coleccionistas, fundaciones, galeristas, curadores, mecenas, críticos, algunas instituciones docentes y políticas).

Una aportación importante a tratar en la presente investigación es que posibilita el establecimiento de nuevas relaciones entre las obras de arte; de distintos artistas, periodos y movimientos estilísticos, favoreciendo sin duda la construcción de nuevos discursos en torno a sus obras.

1.8 VISIÓN POR COMPUTADOR EN EL ANÁLISIS DE OBRAS DE ARTE

Las Humanidades digitales, ámbito en el que convergen las humanidades y la informática, combinan las metodologías propias de las disciplinas humanísticas tradicionales y de las ciencias sociales con el uso de herramientas informáticas y la edición digital (Berry, 2011; Alvaro, 2013).

Los nuevos miembros de esta generación de humanistas digitales abogan por explorar cómo la tecnología puede ayudar en la comprensión de las artes liberales. Las humanidades se ocupan de cuestiones que no son fáciles de medir, pero el análisis de cantidades

Figura 1.29. Panel número 2 ('Representación griega del cosmos') del Atlas Mnemosyne de Aby Warburg (Warburg, 2010).

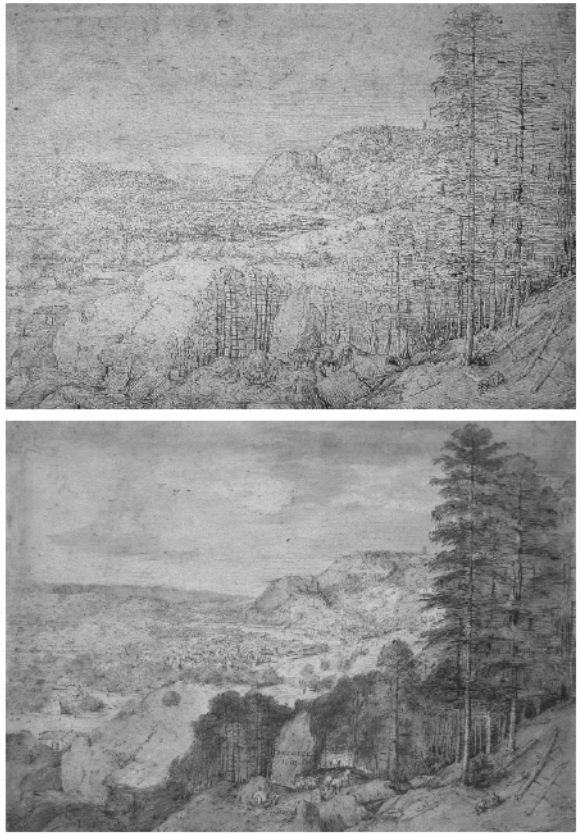


Figura 1.30. Arriba mostramos dibujo auténtico de Pieter Bruegel El viejo. Abajo la imitación (*Rockmore, Lyu, & Farid, 2006).*

sin precedentes de datos puede revelar patrones, tendencias y plantear preguntas inesperadas para nuevos estudios. Así, conservadores, curadores e historiadores del arte pueden encontrar en estos métodos informáticos herramientas valiosas para su trabajo, siempre teniendo en cuenta tanto las fortalezas que aportan como sus limitaciones.

En este contexto en los últimos años, un conjunto de estudiosos de todo el mundo, entrenados en visión artificial, reconocimiento de patrones, procesamiento de imágenes e historia del arte han aplicado técnicas de visión por computador y gráficos por ordenador a los problemas de la historia y la interpretación del arte. Los estudios relacionados en colecciones de obras de arte con métodos de visión artificial se pueden agrupar en categorías, en función de los objetivos subyacentes y los objetos de interés. Los objetos de estudio pueden abarcar desde pinturas históricas y dibujos, hasta ilustraciones de manuscritos históricos, fotografías contemporáneas u otros objetos de dos dimensiones (2D).

Según los objetivos que persiguen, los estudios se pueden agrupar como indicamos a continuación.

1.8.1 El objetivo es determinar la autenticidad de obras de arte o su atribución

Por ejemplo, los trabajos de Rockmore, Lyu, & Farid (2006) describen una técnica computacional para la autentificación de obras de arte (pinturas y dibujos de Perugino y Bruegel). Realizan el análisis a partir de imágenes digitales de alta resolución escaneadas de las obras originales. Este enfoque crea un modelo estadístico del artista explorando un conjunto de obras auténticas, contra el cual se comparan las obras de las que se desea comprobar la autenticidad (Fig. 1.30). Los autores dividen la imagen en múltiples regiones y aplican filtros a diferentes escalas y orientaciones con el fin de extraer vectores de características dimensionales. Entonces, la similitud de pares de las imágenes se calcula utilizando la distancia de Hausdorff (Huttenloche, Klanderman, & Rucklidge, 1993). Los autores comparan siete cuadros de Bruegel y sus imitadores y encuentran distancias más estrechas entre los cuadros de Bruegel que entre las pinturas de sus imitadores.

También aplican estas técnicas para el problema de determinar el número de artistas que pueden haber contribuido a una pintura atribuida a Perugino (Fig. 1.31) y contrastan los resultados con la opinión de expertos. En un primer experimento, los autores recortan los rostros de los cuadros de Perugino y comprueban que las caras pintadas por el mismo artista están más próximas entre sí, según esta distancia, que sus imitaciones.

Las limitaciones consisten en que los resultados se obtienen sobre un conjunto de datos muy limitado de pinturas y no está claro si el marco podría generalizarse a otros conjuntos de datos.

En este mismo sentido de autentificación, otros autores (Bajcsy & Moslemi, 2010) han trabajado el problema en la búsqueda de características sobresalientes de diferentes autores estudiando 48 imágenes de rostros extraídas de ilustraciones de las crónicas de Froissard (Besançon, Bibliothèque d'Etude et de Conservation MS 864 & MS 865). Se etiquetaron manualmente por historiadores de arte dos artistas con 48 imágenes pertenecientes a rostros pintados por cada artista. La Fig. 1.32 muestra en la parte superior dos ejemplos de caras dibujadas por diferentes artistas y sus representaciones y en la parte inferior sus histogramas de características. Estas diferencias se utilizan para determinar que están pintados por artistas diferentes.

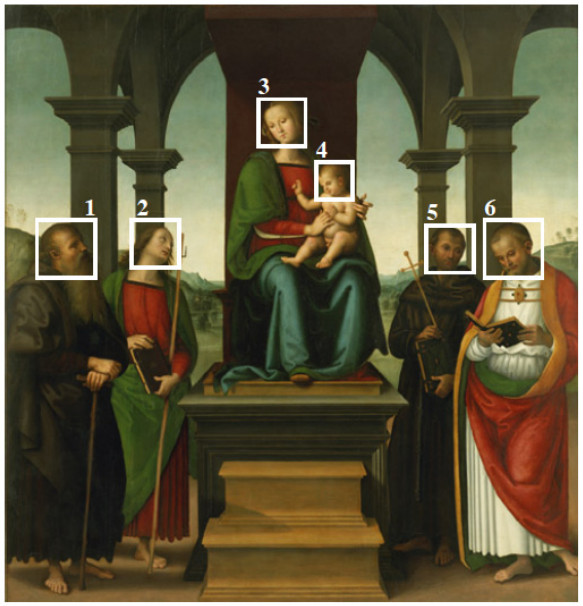


Figura 1.31. Madonna con Niño de Perugino. ¿Cuántas manos contribuyeron a esta pintura? (Rockmore, Lyu, & Farid, 2006).

1.8.2 El estudio de herramientas utilizadas en la pintura

Con herramientas utilizadas en la pintura nos referimos, por ejemplo, a pinceladas y movimientos del pincel. Los métodos de procesamiento de imágenes por ordenador pueden detectar variaciones sutiles de pinceladas que pasarían desapercibidas al ojo mejor entrenado (Melzer, Kammerer & Zolda, 1998).

Otro equipo de investigadores ha utilizado métodos altamente sofisticados como el análisis fractal de las pinturas de action painting de Jackson Pollock (Coddington, Elton, Rockmore & Wang, 2008), introduciendo nuevas medidas visuales nunca antes consideradas

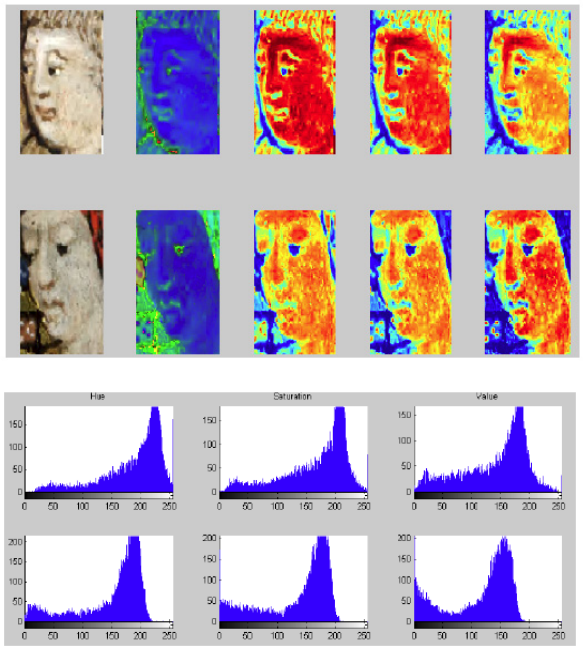


Figura 1.32. Los dos rostros de la izquierda están dibujados por dos artistas diferentes. Debajo se muestran las diferencias a nivel de histogramas; en la fila superior los de un artista y debajo los de otro.

por la comunidad artística.

1.8.3 Descubrir los métodos utilizados en la pintura

Por último, mediante reconstrucciones gráficas por ordenador de los estudios de artistas, los estudiosos pueden explorar los escenarios de creación de las obras, proporcionando información valiosa sobre los métodos de trabajo de algunos artistas. Los psicólogos visuales han demostrado que la mayoría de nosotros, incluidos artistas y estudiosos, no estamos especialmente dotados para juzgar la perspectiva o la ubicación de la iluminación en una fotografía y, por extensión, en una pintura, sin embargo los métodos de análisis de imágenes por ordenador son particularmente buenos en esta tarea.

Con métodos computacionales también se ha estudiado la deformación de imágenes que aparecen dentro de las pinturas distorsionadas en espejos curvos y estas investigaciones han proporcionado nuevos puntos de vista sobre los estudios de los artistas.

Cabe destacar el importante trabajo que está llevando a cabo David G. Stork[8] (2006) en este ámbito. Algunos autores han abordado el problema de la comprensión de los métodos utilizados en la creación artística (Stork, 2006; Stork & Johnson, 2006; Stork & Duarte, 2007). El equipo del Dr. Stork utiliza tecnologías para analizar pinturas con las cuales es posible inferir puntos de fuga, localizar inconsistencias de perspectiva y determinar si el artista utiliza algún tipo de herramienta para su construcción.

También existen sistemas que permiten inferir el número, color y posición de las fuentes de luz basados en la posición, el color y el desenfoque de sombras y luces a lo largo de

8 El Dr. David G. Stork es Director de Investigación en los laboratorios Rambus. Se graduó en física en el Instituto de Tecnología de Massachusetts (MIT) y la en la Universidad de Maryland (College Park), y estudió Historia del Arte en el Wellesley College. Es miembro de la Asociación Internacional para el Reconocimiento de Patrones y de SPIE (sociedad internacional de óptica y fotónica). Para ampliar información visitar la página de internet: http://www.diatrope.com/stork/FAQs.html

las fronteras de oclusión (Fig. 1.33) (Johnson, Stork, Biswas & Furuichi, 2008). De esta forma se pueden estimar los tamaños de los objetos representados en base a la perspectiva y los objetos o relaciones de referencia. Se pueden estimar parámetros de la habitación del artista (o sistema de imagen), tales como la ampliación efectiva, la distancia focal y en algunos casos aberraciones. Ante este tipo de problemas, estos métodos informáticos son más sensibles, más *perceptivos* incluso que un artista o historiador del arte formado y además, este tipo de análisis riguroso de la perspectiva permite detectar inconsistencias o falsificaciones.

1.8.4 Clasificación de pinturas en base al análisis de imagen

El proyecto que se ha desarrollado en esta investigación se enmarcaría en esta sección. Vamos a comentar algunos estudios que han desarrollado otros autores.

En el trabajo de Shen (2009) se utilizaron características globales y locales para clasificar pinturas occidentales. Color, textura, forma y la distribución de color son las características globales mientras que para las características locales utilizan el filtro de Gabor (Russ, 1999). Emplean una red neuronal para etiquetar y clasificar las pinturas. Se trata de un sistema supervisado que utiliza el 20% de los datos como conjunto de entrenamiento y el resto para evaluar la eficiencia de la clasificación, sin que haya solapamiento entre los datos de entrenamiento y los de evaluación. Para un conjunto de 25 artistas y un promedio de 30 imágenes por artista (1080 pinturas en total), los autores lograron una precisión de clasificación del 69,6%.

Los trabajos de Li J. & Wang J.Z. (2004) abordan el aprendizaje basado en la caracterización de diferentes estilos de pintura. Se centran en la comparación de los estilos de pintura de diferentes artistas. Para perfilar el estilo de un artista estiman una mezcla de modelos aleatorios utilizando imágenes de entrenamiento. Se utiliza en el experimento el método MHMM (bidimensional multirresolución modelo oculto de Markov) (Li,Gray & Olshen, 2000; Chellappa & Jain, 1993). En estos modelos se forma una identidad digital distinta

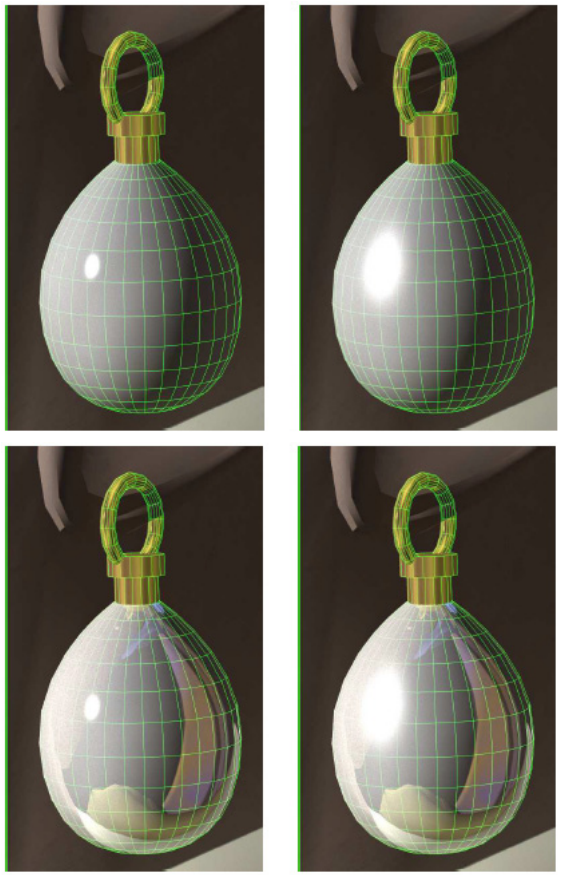


Figura 1.33. Los autores crean un modelo de gráfico por ordenador de la perla (del cuadro "Chica de la perla" de Vermeer) y van ajustando los parámetros para que se acerquen a la representación de la pintura (Johnson, Stork, Biswas & Furuichi, 2008).

Figura 1.34. Algunos de los resultado obtenidos por Li J. & Wang J.Z. (2004). Se trata de fragmentos de las pinturas estudiadas de la dinastía Zhang. En la primera columna los describen como aguada pesada y gruesa, en la segunda como aguada pálida y diluida, en la tercera como trazos pequeños y oscuros y en la cuarta como trazos finos y rápidos.

de cada artista. Han implementado y probado el sistema con fotografías digitales de alta resolución de algunos de los artistas más famosos de China (Fig. 1.34). Los experimentos han demostrado un buen potencial de este enfoque en el análisis automático de pinturas.

Las novedades que aporta nuestro planteamiento en este contexto sería por un lado la búsqueda de significados y no de estilo, en el conjunto de obra de artista y por el otro la aplicación de un método no supervisado para el establecimiento de las categorías semánticas. También la aplicación de la metodología sería novedosa ya que sólo hemos encontrado antecedentes de ella en la clasificación de escenas naturales y nada sobre colecciones de imágenes de arte para descubrir categorías semánticas en obras de artistas abstractos.

1.9 ANTECEDENTES DE LA METODOLOGÍA APLICADA EN LA TESIS

El problema de la clasificación automática en base al contenido semántico de una imagen ha sido abordado por diversos autores en diferentes bases de datos con el fin de categorizar escenas y objetos de distintas tipologías.

En el año 2000, Weber, Welling y Perona plantean un sistema capaz de distinguir entre rostros y coches sin ninguna información añadida. En la Fig. 1.35 encontramos una muestra de los patrones que obtienen en función de los puntos de interés encontrados para cada clase; rostros a la izquierda de la figura y coches a la derecha.

Sivic, Russell, Efros, Zisserman y Freeman, en 2005, buscaron descubrir categorías de objetos y su localización en conjuntos de imágenes sin etiquetar. Consiguieron su objetivo mediante la utilización de un modelo desarrollado en el análisis automático de textos: Análisis probabilístico de Aspetos Latentes (*pLSA*).

En el análisis de textos, el *pLSA* se utiliza para descubrir temas en un corpus, representando el documento como una *bolsa-de-palabras*. Para estos autores una categoría de objetos

equivaldría al concepto de tema de un documento. Así, una imagen que contenga varias categorías se modela como una mezcla de temas. El modelo se aplica a las imágenes mediante el uso de un análogo visual a la palabra formado por el vector de cuantificación de una región de imagen. Este enfoque del tema traducido al dominio visual resulta exitoso pues demuestra buenos resultados, para un pequeño conjunto de objetos, tanto su categoría como su disposición espacial son halladas sin supervisión alguna. Comparando el rendimiento obtenido en su experimento con este método no supervisado, con otros autores que utilizan para un solo objeto, un método supervisado, concluyen que la categorización es satisfactoria. Utilizan dos bases de datos de objetos: una de Caltech y la otra del MIT. En la Fig. 1.36 podemos ver ejemplos de algunos de los resultados.

Quelhas, Monay, Odobez, Gatica-Perez, Tuytelaars y Van Gool en 2005 analizan con descriptores locales y aspectos latentes diferentes escenas de paisajes urbanos y paisajes campestres. En la Fig. 1.37 podemos ver ejemplos de algunos de los resultados obtenidos.

Lazebnik, Schmid & Ponce, en 2006, implementan un método de categorización de escenas que funciona dividiendo la imagen en sub-regiones cada vez más finas y computan las características locales de los histogramas que se encuentran dentro de cada sub-región (Fig. 1.38). Esta *pirámide espacial* es una extensión computacionalmente eficiente de la representación de la imagen como *bolsa-de-características* y demuestra progresos significativos en la categorización de escenas. Su método mejora los resultados hasta el momento para la base de datos Caltech-101 y logra una alta precisión en una gran base de datos con quince categorías de escenas naturales que presentan una gran variabilidad dentro de una clase determinada (Fig. 1.39). El contexto de la *pirámide espacial* también nos da una idea del éxito de varios descripciones de imagen novedosos en el momento; incluyendo descriptores GIST (Torralba et al., 2003) y descriptores *SIFT* (Lowe, 2004).

Bosch, Zisserman & Muñoz en 2006, dado un conjunto de imágenes de escenas que contienen múltiples categorías de objetos se proponen descubrir estas categorías de manera no supervisada, y posteriormente usar esta distribución de objetos para realizar la clasificación de escenas de forma supervisada. Utiliza *pLSA* aplicado a la representación de las imágenes como *BoW*. El resultado obtenido se compara con el de otros autores

Figura 1.35. Muestra de los patrones obtenidos por Weber et al. (2000) para caras en la parte izquierda y para coches en la parte derecha.

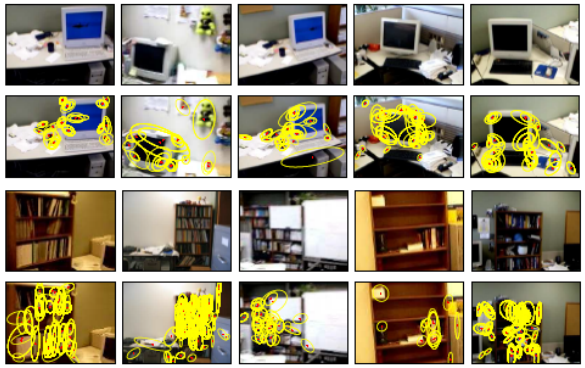


Figura 1.36. Ejemplos de dos de los temas encontrados por Sivic et al. en 2005 en la base de datos del MIT. La fila superior muestra las imágenes originales y la fila de abajo muestra las palabras visuales que corresponden al tema particular de esa imagen. Es importante observar que se puede dar una interpretación semántica a esos temas; en las dos filas superiores se trata de ordenadores y en las dos inferiores de estanterías.

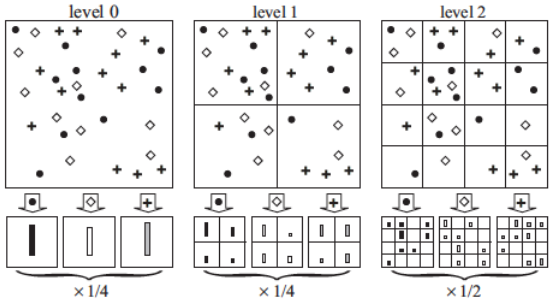


Figura 1.38. Ejemplo ilustrativo de la construcción de una pirámide de tres niveles. La imagen tiene tres tipos de entidades, indicadas por círculos, diamantes, y cruces. En la parte superior, se subdivide la imagen en tres niveles diferentes de resolución. A continuación, para cada nivel de resolución y cada canal, contamos las características que se encuentran en cada bin espacial. Esquema extraido de Lazebnik et al. (2006).

Figura 1.39. Imágenes recuperadas de la base de datos Caltech-101 que es probablemente la base de datos de objetos y escenas naturales mas diversa disponible hoy en día.

que han trabajado de forma supervisada (Oliva & Torralba, 2001; Vogel & Schiele, 2004) y semi-supervisada (Fei-Fei & Perona, 2005) . En todos los casos la combinación de *pLSA* (no supervisado), seguido de clasificación (supervisada) logra mejores resultados.

1.10 LOS DATOS

La dificultad para la categorización de los objetos y escenas naturales está vinculada a:

1- La gran variabilidad dentro de la misma clase y al grado de desorden; un mismo objeto puede presentarse en una pose estereotipada o puede estar solapado, girado o a diferente escala, aumentando con ello los problemas de clasificación.

2- Variabilidad entre distintas clases: es importante que no se produzcan confusiones de una clase con otra.

3- Invariancia a la escala. Significa que una misma escena u objeto no cambie de clase por aumentar o disminuir su escala.

4- Condiciones de iluminación: los sistemas deberían ser capaces de reconocer una misma escena u objeto bajo diferentes condiciones de iluminación.

5- Rotaciones, oclusiones y variaciones del punto de vista.

Los autores que trabajan con obras de arte utilizan bases de datos, como hemos visto ,muy heterogéneas, mientras que los autores comentados en referencia al análisis de escenas naturales trabajan con bases de datos de acceso público y suelen utilizar las mismas que sus colegas para validar sus resultados ya que esto les permite comparar los rendimientos de clasificación sobre las mismas imágenes.

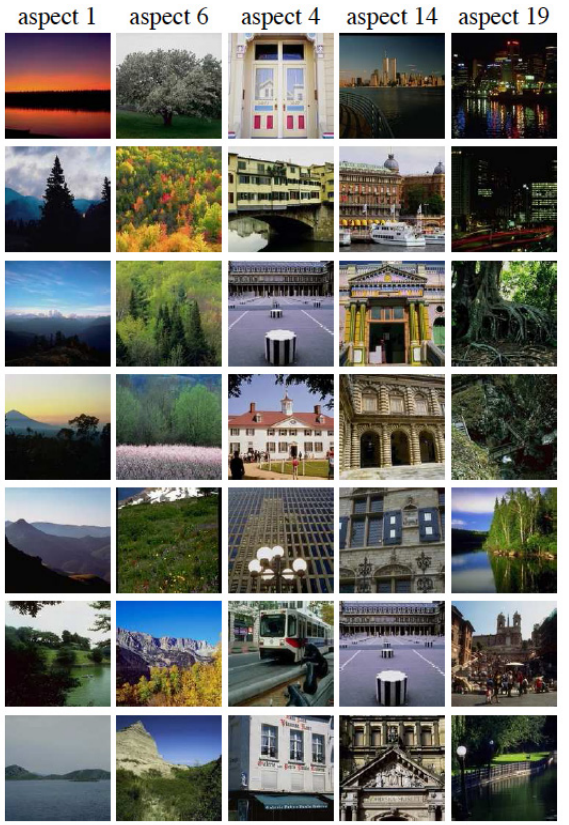


Figura 1.37. Muestra las imágenes más probables de algunos de los aspectos obtenidos por Quelhas et al. en 2005. El aspecto 1 correspondería a la categoría semántica de paisaje panorámico, el aspecto 6 a paisaje boscoso y el aspecto 14 correspondería a la categoría de ciudad. Imagen extraída de Quelhas et al. (2005).

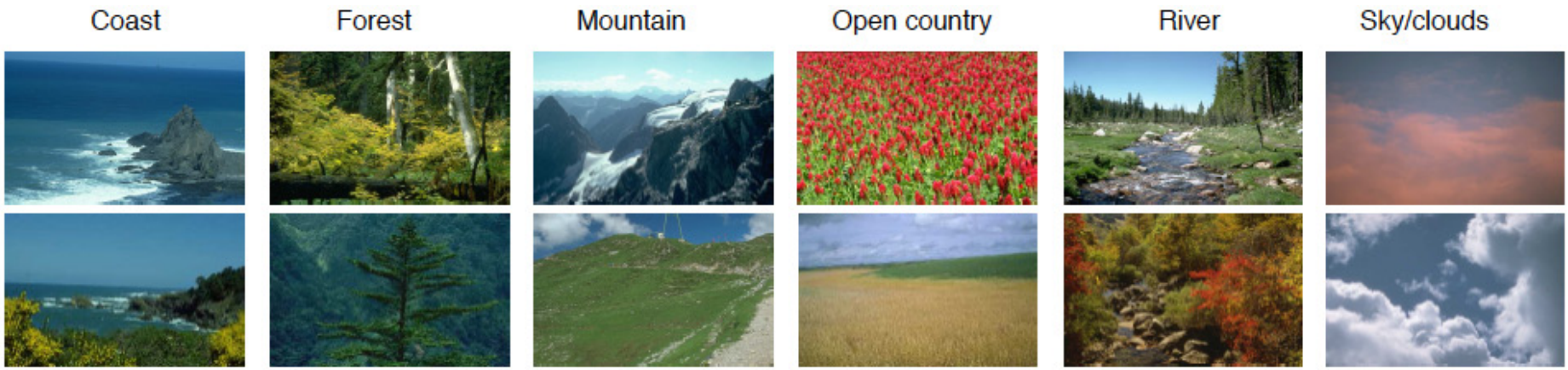


Figura 1.40. Conjunto de datos de Vogel et al. (2004)

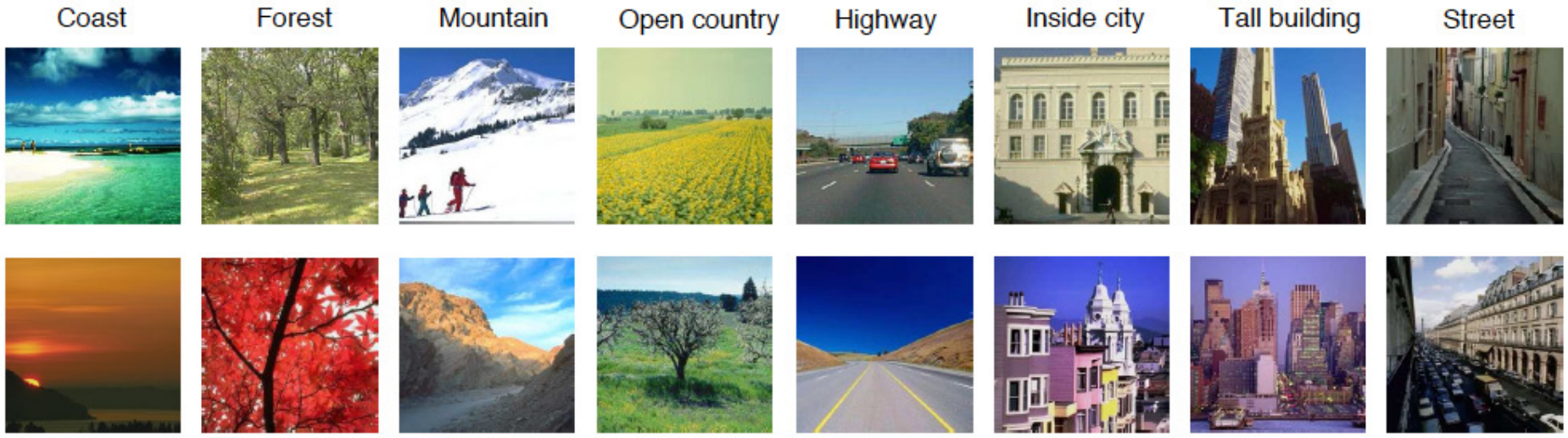


Figura 1.41. Conjunto de datos de Oliva et al. (2001)

Figura 1.42. Conjunto de datos de Fei Fei et al. (2005).

Figura 1.43. Conjunto de datos Caltech-101.

1.10.1 Bases de datos utilizadas para clasificación de escenas

Conjunto de datos de Vogel et al. (2004):

Incluye 702 imágenes de escenas naturales correspondientes a 6 categorías: 144 costas, 103 bosques, 179 montañas 131 montañas, campo abierto, 111 río y 34 cielo con nubes (Fig. 1.40).

Conjunto de datos de Oliva et al. (2001):

Incluye 2688 imágenes clasificadas como 8 categorías: 360 costas, 328 bosques, 374 montañas, 410 campo abierto, 260 carretera, 308 interior de las ciudades, 356 edificios altos y 292 calles (Fig. 1.41).

Conjunto de datos de Fei Fei et al. (2005):

Contiene 13 categorías y sólo está disponible en escala de grises. Este conjunto de datos consta de las 2.688 imágenes de Oliva et al. (2001) más 241 barrio residencial, 174 dormitorios, 151 cocinas, 289 sala de estar y 216 oficinas (Fig. 1.42).

1.10.2 Bases de datos utilizadas para clasificación de objetos

Conjunto de datos Caltech-101:

Consta de imágenes de 101 categorías de objetos. Esta base de datos contiene entre 40 y 800 imágenes por categoría, sin embargo la mayoría de las categorías tienen cerca de 50 imágenes (Fig. 1.43).

Procederemos a comentar las características diferenciales entre estas imágenes y las que hemos utilizado en nuestro análisis dado que una de las principales aportaciones de nues-

tro trabajo es ampliar la aplicación de la metodología al estudio de bases de datos de categorización más difícil, que corresponden a imágenes bastante abstractas y cuya dificultad radica en establecer a que clase pertenecen. La decisión de no abordar el análisis de imágenes de pinturas más figurativas se tomó en base a que son cercanas a las escenas naturales y por ello los resultados eran más predecibles.

Conviene aclarar en este punto que, al referirnos al concepto "abstracto", no nos referimos al hecho de que las imágenes "extraen elementos comunes a cierto número de casos particulares y los presenta como una nueva suma o configuración" (Arnheim, 1980, p. 35), sino más bien a imágenes de contenido no objetual, como explica mejor el propio Arnheim:

> Es necesario considerar dos puntos de partida opuestos: por un lado, el material estimular del objeto; por otro, la forma, prerrequisito indispensable de la comprensión visual. Percibir una cosa, lo mismo que representarla, significa encontrar una forma en su estructura. Los esquemas del arte "no objetivo", considerados desde el punto de vista del mundo de las cosas naturales, son extremadamente abstractos. Reducen la representación de la realidad a un equivalente visual de las fuerzas físicas y psicológicas universales que están en la base de la naturaleza y la vida, y de su interacción. Expresan de esta manera armonía y disonancia, dominio y coordinación, contraste y semejanza, movimiento y reposo, equilibrio y desequilibrio. Sin embargo, desde el otro punto de vista, es decir, desde el punto de vista de la forma, los esquemas básicos no objetivos, no son abstractos. Son los elementos mismos de la comprensión visual, el material de construcción de la composición que el artista crea para representar la estructura del mundo de la manera en que su temperamento le hace verlo (p. 45).

1.10.3 Bases de datos de obra de artista utilizadas en la tesis

En el caso que abordamos en nuestra investigación los conceptos semánticos son más abstractos y la complejidad para reconocer las clases es grande dado que no siempre es fácil encontrar las palabras o las ideas a las que se refieren las agrupaciones. El problema no sólo reside en el lenguaje, sino en que para poder plasmar esas cualidades percibidas en

las categorías adecuadas, previamente debemos haberlas visto, oído, pensado o sentido (Arnheim, 1983, p.15). Por ello hemos recurrido al asesoramiento de expertos en el ámbito del análisis del arte y de la creación artística que han podido corroborar el sentido de las agrupaciones que realizaba el sistema de forma automática.

Antoni Tàpies Puig[9] (Barcelona, 1923-2012) artista catalán, uno de los principales exponentes a escala mundial del informalismo y está considerado como uno de los artistas catalanes mas importantes del siglo XX. Existe un centro de estudio y conservación de su obra en la Fundación Antoni Tàpies de Barcelona (Tàpies, 2001). En su obra se combinan tradición e innovación en un estilo abstracto pero lleno de simbolismo que otorga gran relevancia al sustrato matérico. Hay que subrayar el marcado estilo espiritual que inspira su obra, en la que el soporte material transciende su estado para analizar de forma profunda la condición humana (Fig.44).

Miquel Planas (Planas, 2014) empezó a trabajar con imágenes digitales en el año 2000; esta labor fotográfica, se había iniciado ya con anterioridad pero la consolidación y difusión de la imagen digital, juntamente por su facilidad de manipulación y almacenamiento contribuyeron a que su trabajo fotográfico se viese reforzado. El conjunto de su colección se caracteriza por una tendencia a captar formas, objetos o elementos diversos, que ve y observa, con los que establece una conexión y que finalmente captura; este proceso no consiste en un trabajo de campo previsto ni elaborado, no responde a metodología o estructura alguna; forma parte de una predisposición a la observación, a la contemplación, con la finalidad de plasmar aquellos elementos, con los que de forma inicialmente inconsciente se establece una relación de carácter emocional, en que se perciben unos vínculos, inicialmente inconexos y tal vez azarosos con el sujeto contemplado. Todo este grupo de imágenes generado, se fue consolidando como un conjunto visual que permite al artista proponer e iniciar futuros procesos de creación (Fig. 45).

Los conjuntos de imágenes de partida de nuestro análisis los constituyen;

Figura 1.44. Conjunto de imágenes digitales de obras de Antoni Tàpies.

9 En la página web de su fundación (Tàpies, 2001) se puede encontrar una selección de bibliografía, tanto de escritos del propio artista como de otros autores que han escrito sobre él y su obra.

1- Por un lado la colección de 2846 imágenes fotográficas que el escultor Miquel Planas utiliza como fondo de ideación artística. Son fotografías capturadas por el propio artista, la mayoría de exteriores y muestran detalles desde diferentes ángulos (llegando a ser fragmentos y particularidades que se pueden considerar como elementos abstractos y/o texturados). El tamaño de las imágenes del artista está entre 480 x 480 píxeles y 1400 x 1400 píxeles. La interacción con el propio artista a la hora de valorar los resultados ha resultado fundamental pues su testimonio enriquece el análisis.

2- Por otro lado la colección de 434 imágenes digitalizadas de pintura y obra gráfica (gran parte perteneciente a libros de artista) de Antoni Tàpies que posee su Fundación en Barcelona (Tàpies, 2001). Las imágenes digitalizadas de obra escultórica tridimensional han sido descartadas pues las representaciones en dos dimensiones dependen del punto de vista desde el que se ha tomado la instantánea y pueden distorsionar la construcción de los descriptores totales del conjunto.

El proceso reescala a 480 píxeles todas las imágenes que superan este tamaño.

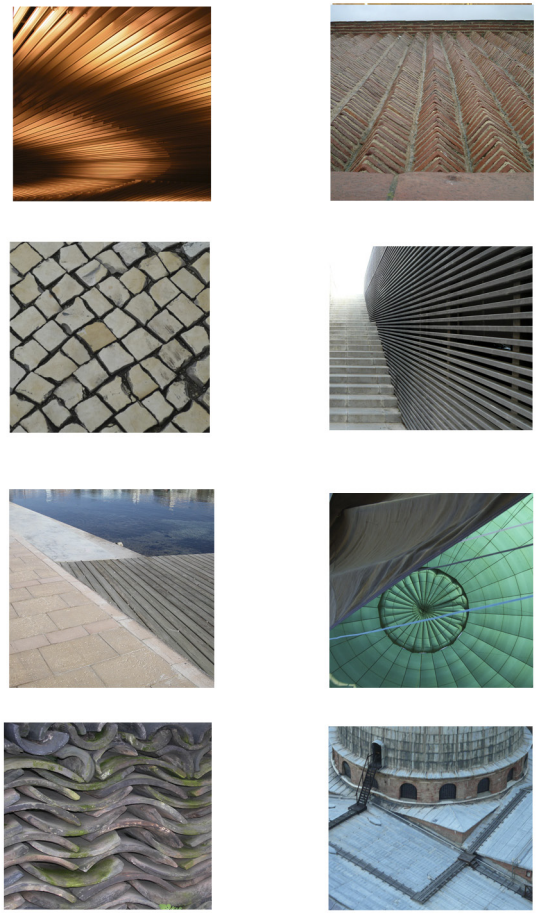


Figura 1.45. Conjunto de imágenes digitales realizadas por Miquel Planas a partir de 2003.

2- METODOLOGÍA

2.1 REPRESENTACIÓN DE LA IMAGEN

La representación de la imagen digital es un elemento clave para su clasificación, anotación, segmentación o recuperación. Casi todos los métodos de visión por computador, cuando se enfrentan al problema del análisis del contenido de una imagen, recurren a funciones adecuadas para describirlo de forma compacta. Este sería el caso de los procedimientos basados en características locales que producen una representación de la imagen versátil y sólida capaz de mostrar el contenido global y local al mismo tiempo, y a la vez hacen robusta la descripción ante la oclusión parcial de objetos contenidos y la transformación de la propia imagen.

En visión artificial existen muchas formas de representar el contenido de una imagen. Se podría decir que existen tres métodos principales:

1- Los procedimientos que extraen directamente características de bajo nivel de las imágenes.

2- Los métodos que utilizan una representación semántica de la imagen.

3- Los métodos que utilizan regiones locales como representación de la imagen.

2.1.1 Modelos de representación de la imagen de bajo nivel

Estos modelos representan las imágenes usando características de bajo nivel como texturas, bordes o histogramas de color. Por ejemplo; la presencia de rectas y bordes verticales-horizontales puede ser un indicio de que se trata de una escena urbana, o si la imagen contiene mucho color azul, puede que se trate de un paisaje de mar. En el método se pueden a su vez distinguir dos planteamientos:

a)- Representaciones globales (Fig. 2.1.a), donde las características de bajo nivel se calculan sobre la toda la imagen. Debido a la complejidad del contenido visual, en los sistemas de clasificación se suelen obtener mejores rendimientos utilizando varias características globales combinadas.

b)- Representaciones locales (Fig. 2.1.b), donde la imagen se divide primero en varios bloques, y después se extraen las características de cada uno de ellos. El sistema de clasificación obtiene primero una categoría para cada bloque y posteriormente estos resultados se combinan para obtener una categoría total de la imagen. La principal ventaja de estos métodos es que proporcionan una representación de la imagen muy simple. El principal inconveniente es que, si las imágenes tienen un notable desorden o hay mucha variabilidad intra-clase, esta representación no es suficiente para discriminar entre diferentes categorías.

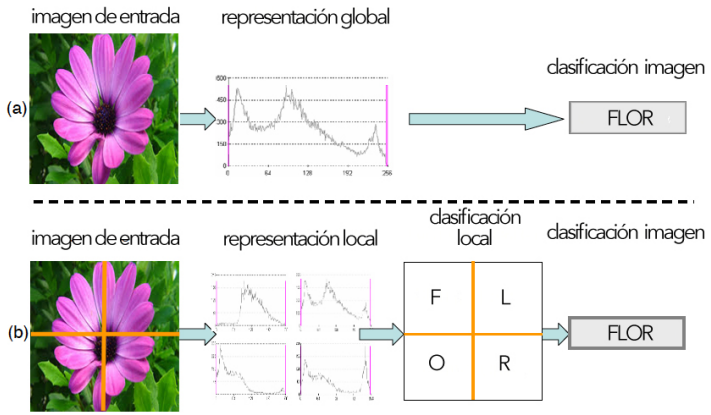


Figura 2.1. Ejemplo de representación de imagen mediante el uso de características de bajo nivel, por ejemplo, un histograma de color. (a) representación de la imagen global y (b) la representación local de la imagen mediante el uso de un histograma de color en cada sub-bloque. (Bosch, 2007)

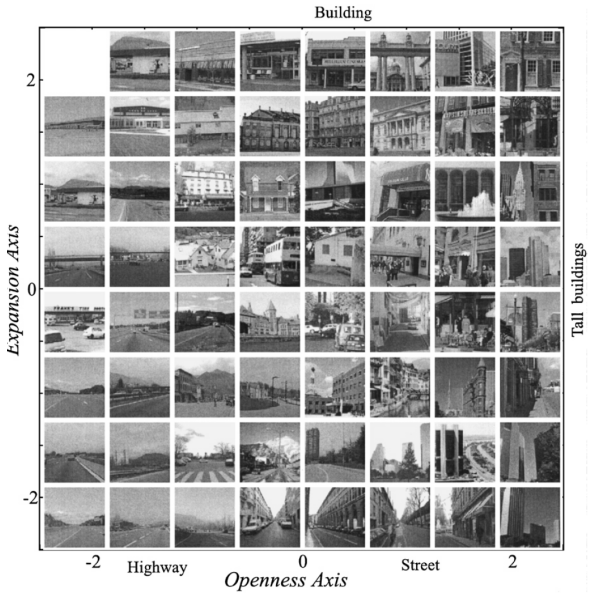


Figura 2.2. Representación semántica usando modelos globales. Organización de entornos artificiales de acuerdo con los grados de apertura y expansión. (Oliva & Torralba, 2001).

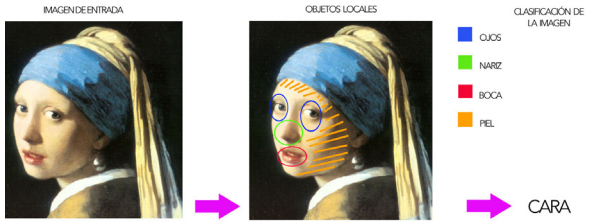


Figura 2.3. Esquema de representación semántica usando modelo locales (Bosch, 2007).

2.1.2 Representación semántica de la imagen

Podemos distinguir entre dos modelos:

a)- Modelos globales: Se realiza la descripción semántica utilizando las propiedades estadísticas del total de la imagen. Introducen un nivel semántico intermedio relacionado con configuraciones globales y estructura de la imagen. Por tanto la imagen se describe por las propiedades visuales, que son compartidas por las imágenes de una misma categoría. Oliva y Torralba (2001) propusieron un modelo computacional para el reconocimiento de escenas del mundo real (4 escenas naturales y 4 escenas artificiales). El procedimientos se basa en cinco cualidades perceptivas: el carácter natural (se refiere a que esté construido por el hombre), la apertura (se refiere a la presencia de una línea de horizonte), rugosidad (complejidad fractal), la expansión (perspectiva en escenas construidas por el hombre) y la desviación del horizonte en escenarios naturales. El modelo genera un espacio multidimensional en el que las escenas que comparten la pertenencia a categorías semánticas se proyectan juntas (Fig. 2.2).

b)- Modelos locales: El contenido semántico local de las imágenes puede ser utilizado como una representación intermedia para la clasificación de imágenes que permita hacer frente a la brecha entre las características de bajo y de alto nivel. Estos métodos se basan principalmente en la localización inicial de las diferentes regiones de la imagen (Fig. 2.3). Entonces se utilizan clasificadores locales para etiquetar estas regiones como pertenecientes a una determinada clase de objeto (por ejemplo, cielo, gente, piedra). A veces se introducen también algunas relaciones espaciales entre los objetos de las imágenes (por ejemplo, el cielo está por encima de una montaña o los ojos están por encima de la nariz). Finalmente se clasifica la imagen global en función de esta información local. Recientemente se han propuesto diferentes formas de llevar a cabo esta estrategia.

1b- Mojsilovic, Gomes, y Rogowitz (2002) inicialmente segmentan la imagen en base a la información de color y textura para encontrar los indicadores semánticos (por ejemplo, la piel, cielo, agua). A continuación, se utilizan estos objetos para identificar las categorías

semánticas (por ejemplo, personas, coches, paisajes).

2b- Barnard, Duygulu, Forsyth, Freitas, Blei y Jordan (2003): presentan una aproximación para modelar conjuntos de datos, centrándose en el caso específico de imágenes segmentadas con texto asociado. Consideran en detalle la predicción de palabras asociadas con las imágenes completas (auto-anotación) y que corresponden a áreas de imagen particulares (región de nomenclatura).

3b- Vogel y Schiele (2007), en contraste con los métodos anteriores que inicialmente segmentan la imagen, ellos trazan una cuadrícula espacial que la divide en subregiones regulares. La técnica utiliza el color y la textura para realizar clasificación de paisajes y recuperación de imágenes basada en un sistema de dos etapas; en primer lugar, la imagen se divide en subregiones de 10 x 10 y cada una se clasifica. El sistema puede aprender para cada categoría de la escena una representación prototípica. En una segunda fase se lleva a cabo la clasificación de imágenes a partir de estos prototipos. (Fig. 2.4)

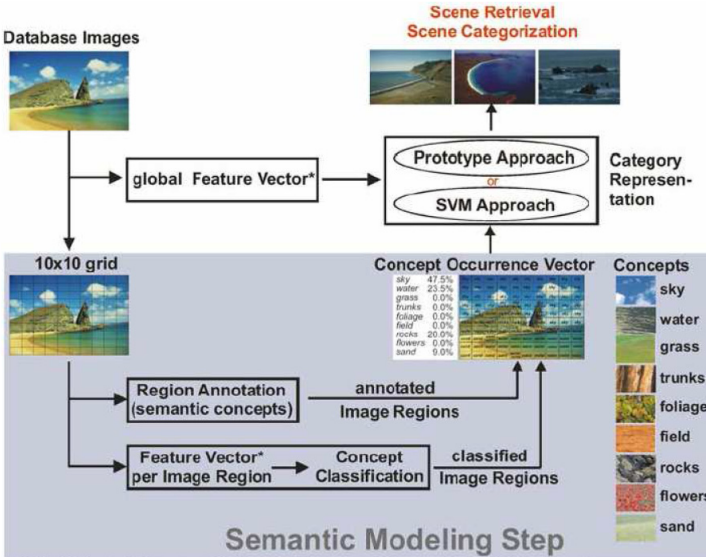


Figura 2.4. Visión general del enfoque de Vogel y Schiele (2007)

La principal ventaja de estos métodos es que utilizan significados humanos para clasificar primero los objetos y después la imagen. Son bastante discriminativos y se han aplicado para clasificar imágenes en un mayor número de categorías que con los métodos de bajo nivel. El principal inconveniente es que la mayoría de ellos se basan en la inicial segmentación de la imagen y esto puede causar algunos problemas cuando se trabaja con imágenes complejas, ya que, si el método de segmentación no es exacto, se puede fusionar algunas partes de los objetos y provocar una descripción de la imagen errónea.

Además, Thorpe, Fize, & Marlot en 1996 encontraron que los humanos son capaces de categorizar de forma muy rápida imágenes naturales complejas que contienen animales o vehículos. Fei-Fei, VanRullen, Koch & Perona en 2002 mostraron que se necesita poca o ninguna atención para esta rápida categorización de imágenes naturales. Por tanto, según Bosch (2007) ambos estudios plantean un serio desafío a la opinión actualmente aceptada de que para entender el contexto de una escena

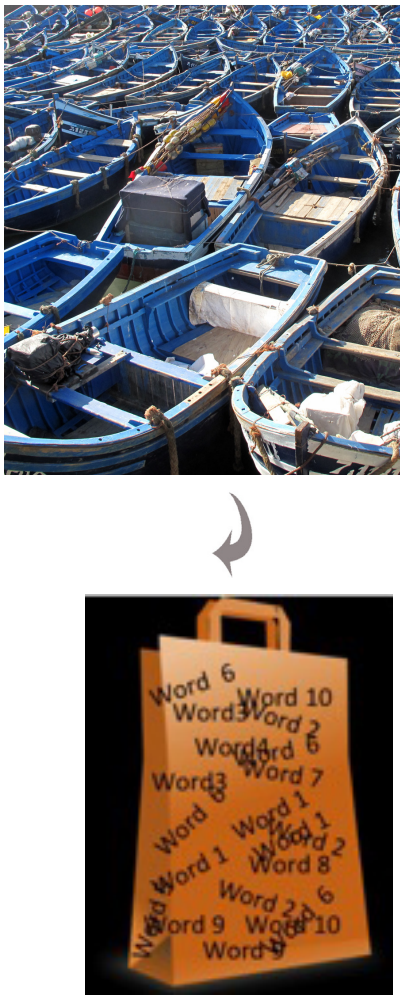


Figura 2.5. La representación *BoW* de una imagen no contiene información acerca de las relaciones espaciales entre palabras visuales que la componen.

compleja, es necesario primero reconocer los objetos para después reconocer la categoría de la imagen (Treisman & Gelade, 1980).

2.1.3 Representación de la imagen por *patches* locales

En este caso las imágenes se representan por cientos de *patches* locales. Utilizan un detector de región para encontrar un conjunto de zonas características de la imagen y luego las representan mediante algún tipo de descriptor. El modelo *Bag-of-Words* constituye una exitosa representación de este tipo.

2.2 MODELO *BAG-OF-WORDS (BoW)*

En los últimos años podemos encontrar en la literatura sobre clasificación de imágenes, gran cantidad de propuestas que utilizan la representación de la imagen mediante el modelo de *Bag-of-Words*, utilizando histogramas de fragmentos locales. Todas estas propuestas representan el contenido de la imagen mediante descriptores locales como por ejemplo, palabras visuales.

La metodología *Bag-of-Words* (a veces también llamada *bag-of-features* o *bag-of-visual-terms*) fue propuesta por primera vez para el análisis de documentos de texto y más tarde adaptada para aplicaciones de visión por ordenador (Leung & Malik, 2001; Sivic & Zisserman, 2003). El modelo se aplica a las imágenes mediante el uso de un análogo visual de la palabra, constituido por el vector de cuantización de características visuales (color, textura, etc.) que actúa como descriptor de región. En el caso de documentos, estos quedan representados como una distribución de frecuencias de las palabras presentes en el texto, sin tener en cuenta las relaciones sintácticas existentes entre ellas y lo mismo ocurre en la representación de imágenes con este modelo. (Fig. 2.5)

Trabajos recientes han demostrado que la representación de características locales como *BoW* es adecuada para la clasificación de imágenes y demuestra impresionantes niveles de rendimiento (Fei-Fei & Perona, 2005; Lazebnik, Schmid & Ponce, 2006; Quelhas, Monay, Odobez, Gatica-Perez, Tuytelaars & Van Gool, 2005).

La construcción del *BoW* a partir de las imágenes implica las siguientes etapas, descritas esquemáticamente en los cuatro pasos que se detallan en la Fig. 2.6 y en el apartado 3 del Anexo A:

1- Detección automática de regiones / puntos de interés (patches locales).

2- Cálculo de descriptores locales sobre estas regiones / puntos.

3- Cuantizar los descriptores en palabras para formar el vocabulario visual.

4- Contabilizar las veces que ocurre en la imagen cada palabra específica del vocabulario con el objetivo de construir el *BoW* (histograma de palabras).

2.2.1 Detección automática de puntos de interés: descriptores *SIFT*

En la fase inicial del proceso es fundamental caracterizar cada imagen mediante un conjunto de descriptores locales extraídos de ciertos puntos de interés para, en lugar de comparar las imágenes entre sí, poder comparan los conjuntos de descriptores. Estos puntos de interés deben ser relevantes tanto por su estabilidad como por la cantidad de información de su entorno que nos pueden proporcionar.

En la presente investigación vamos a trabajar con dos tipos de descriptores: los Scale Invariant Feature Transform (*SIFT*) definidos por Lowe (Lowe,2004) y los descriptores de textura de Haralick definidos en 1973 (Haralick, 1973).

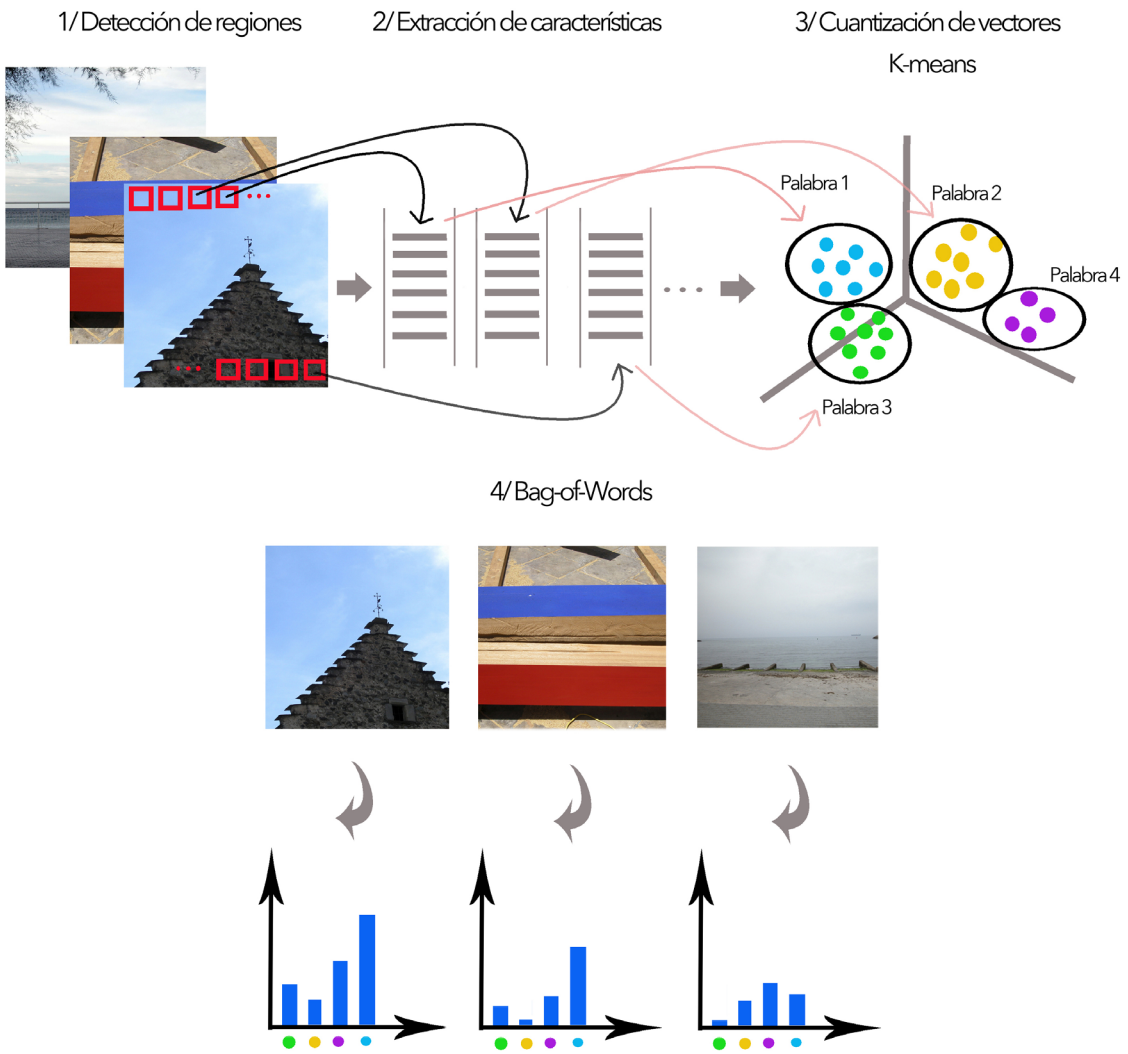


Figura 2.6. Cuatro pasos para calcular el modelo *Bag-of-Words* con imágenes. 1. Detección de regiones, 2. Extracción de características. 3. Cuantización de vectores (*K-means* y 4. Cálculo de los histogramas *BoW*.

El descriptor Scale Invariant Feature Transform (*SIFT*) fue desarrollado por (Lowe, 2004) como un algoritmo capaz de detectar puntos característicos (*keypoints*) estables en una imagen. Estos puntos son invariantes frente a diferentes transformaciones como traslación, escala, rotación, iluminación y transformaciones afines. Originalmente fue desarrollado para el reconocimiento de objetos en general y para realizar la alineación de imágenes. El algoritmo *SIFT* se compone principalmente de cuatro etapas que se describen siguiendo la implementación de (Lowe , 2004):

1- Detección de extremos en el Espacio de Escala: La primera etapa del algoritmo realiza una búsqueda sobre las diferentes escalas y dimensiones de la imagen identificando los candidatos a *keypoints*. Esto se lleva a cabo mediante la función DoG (*Difference-of-Gaussian)* (Fig. 2.7).

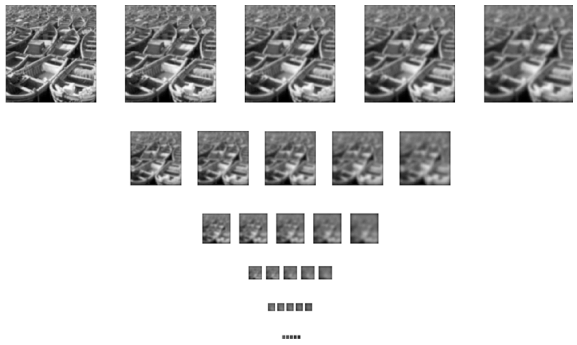


Figura 2.7. Pirámide Gausiana.

2- Localización de los keypoints: Se seleccionan los *keypoints* a partir del conjunto de candidatos encontrados, aplicando una medida de estabilidad sobre todos ellos para descartar los que no sean adecuados (Fig. 2.8).

3- Asignación de la orientación: Se asignan una o más orientaciones a cada *keypoint* basándose en las direcciones locales presentes en la imagen gradiente. Todas las operaciones posteriores serán realizadas sobre los datos transformados según la orientación, escala y localización dentro de la imagen, lo que nos proporcionará la invariancia parcial a distorsiones de forma así como a cambios de iluminación (Fig. 2.9).

4- Descriptor del keypoint: La última etapa hace referencia a la representación de los *keypoints* como una medida de los gradientes locales de la imagen en las proximidades de dichos puntos clave y respecto de una determinada escala. Cada punto de interés corresponde a un vector de características compuesto por 128 elementos (Fig. 2.10).

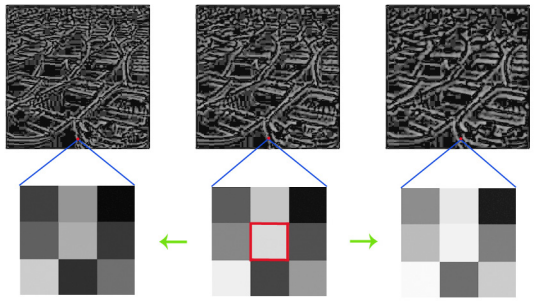


Figura 2.8. Localización de keypoints.

Las etapas anteriores están detalladas de una forma más exhaustiva en el apartado 1 del Anexo A.

En algunos estudios (Lazebnik et al., 2006; Fei-Fei & Perona, 2005) el cálculo de los descrip-

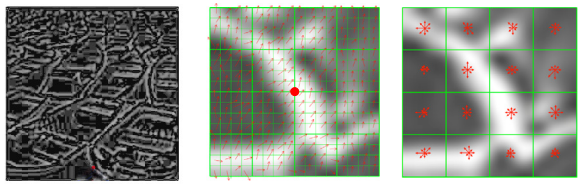


Figura 2.9. a) Keypoint. b) Región de 16 x 16 píxeles alrededor del keypoint y gradientes. c) Subregiones de 4 x 4 píxeles con histogramas de sólo 8 orientaciones.

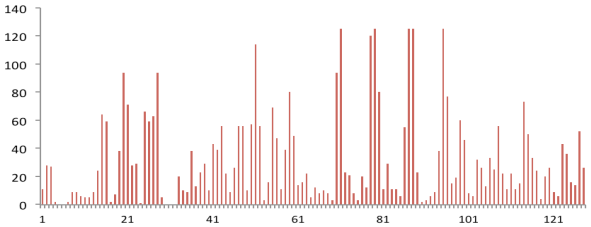


Figura 2.10. Descriptor. La orientación del keypoint corresponde al valor máximo.

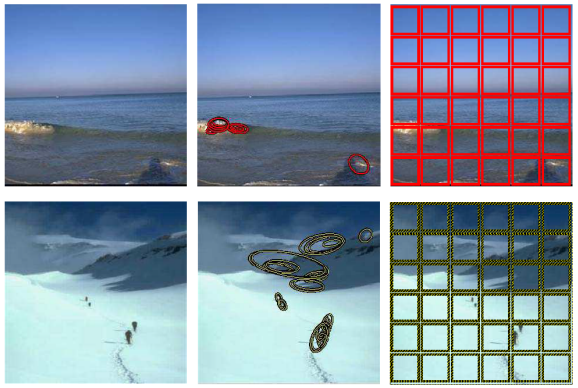


Figura 2.11. a) imagen natural. b) puntos de interés. c) malla regular. (Bosch,*2007)*

tores locales *SIFT*, en lugar de realizarse únicamente en los puntos de interés, se efectúa en los nodos de una malla regular sobreimpuesta en la imagen (Fig. 2.12). Este enfoque es preferible con el fin de mejorar la capacidad de discriminación en implementaciones orientadas a la clasificación de escenas, dado que, para determinados tipos de imágenes puede resultar poco representativo utilizar únicamente los puntos destacados. Como vemos en la Fig. 2.11, si solo utilizamos los puntos de interés la imagen no queda bien representada. Para resolver este inconveniente muchos autores superponen una rejilla regular sobre la imagen, de esta manera se puede tener información del total y no sólo de aquellas partes donde se detectan objetos. Este método ha demostrado un mejor rendimiento que cuando se utilizan regiones dispersas. En este trabajo adoptaremos este enfoque dado que, al tratarse nuestro estudio de imágenes abstractas, los detalles del total son significativos.

Para determinar el tipo concreto de descriptor *SIFT* que se utiliza son fundamentales parámetros como su densidad en la rejilla, el tamaño de los patches y el grado de solapamiento. Bosch (2007) demuestra mejor rendimiento de los descriptores *SIFT* en los siguientes casos:

1- mejores resultados en los descriptores *SIFT* densos que en los espaciados, siendo bastante lógico dado que con esta distribución de descriptores se posee más información sobre la imagen.

2- con solapamiento de los patches, un 6% más de rendimiento que si no hay solapamiento.

3- respecto a descriptores *SIFT* en gris o teniendo en cuenta la información sobre el color, se obtienen mejores rendimientos en la clasificación de escenas naturales teniendo en cuenta el color, sin embargo no son significativas cuando se trata de escenas construidas por el hombre o de interior.

Teniendo en cuenta estos resultados, en el presente estudio se han utilizado dos tipos de descriptores *SIFT* densos (Vedaldi & Fulkerson, 2008):

1- SIFT en gris: Los descriptores *SIFT* se calculan en los nodos de una cuadrícula regular con espaciado entre nodos de *M* píxeles; en este caso *M* = 5,10 y 15. En cada nodo de la rejilla se calculan los descriptores *SIFT* sobre patches de apoyo circulares con radios *r* = 4, 8, 12 y 16 píxeles. En consecuencia, todos los nodos se representa por *n* descriptores *SIFT* (donde *n* es el número de soportes circulares), cada uno es 128-dimensional. Se calculan múltiples descriptores para permitir la variación de escala entre imágenes. Los patches de radios 8, 12 y 16 se superponen. Los descriptores son invariantes a la rotación.

2- SIFT en color: Es como los anteriores pero ahora los descriptores *SIFT* se calculan para cada canal HSV, RGB u OPPONENT (dependiendo de la opción que se desee) de la imagen y se apilan. Este descriptor contiene información del gradientes de color de cada canal de la imagen.

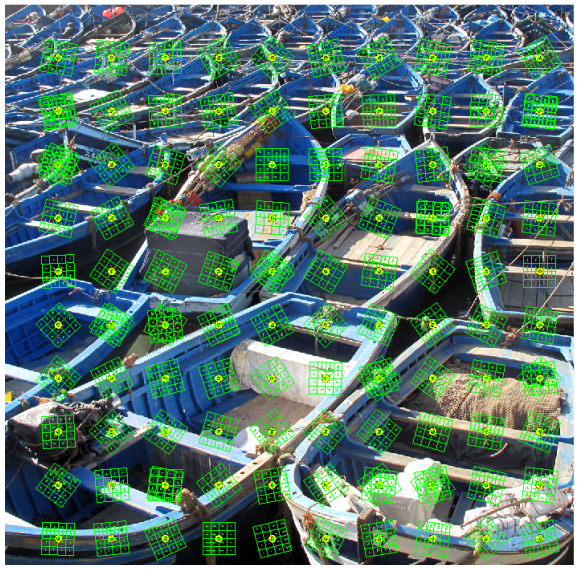


Figura 2.12. Imagen con una malla regular de 10 x 10 puntos de interés.

2.2.3 Construcción del Vocabulario visual

El modelo *BoW* define una metodología de trabajo para clasificar imágenes, si bien numerosos aspectos concretos de su aplicación quedan a la elección del analista, por ejemplo el tamaño y estructura del vocabulario, o la determinación del tipo de clasificador de que se va a utilizar.

El punto de partida para la construcción del vocabulario visual es el conjunto de descriptores calculados para la colección de imágenes, y el punto al cual queremos llegar es el vocabulario de palabras visuales (equivalente al término inglés *visual term* usado en la literatura). Cada imagen ha quedado descrita mediante los descriptores *SIFT*, por tanto disponemos de una gran colección de descriptores. La construcción del vocabulario se realiza mediante agrupación (*clustering*). Concretamente aplicamos el algoritmo *K-means* a un conjunto representativo de descriptores locales extraídos de la colección de imágenes (50000) y tomaremos como palabras visuales los vectores de medias de cada clúster. Usamos la distancia euclidia ordinaria en los procesos de agrupación y cuantización, y elegimos el número de clústers dependiendo del tamaño deseado de vocabulario. En el apar-

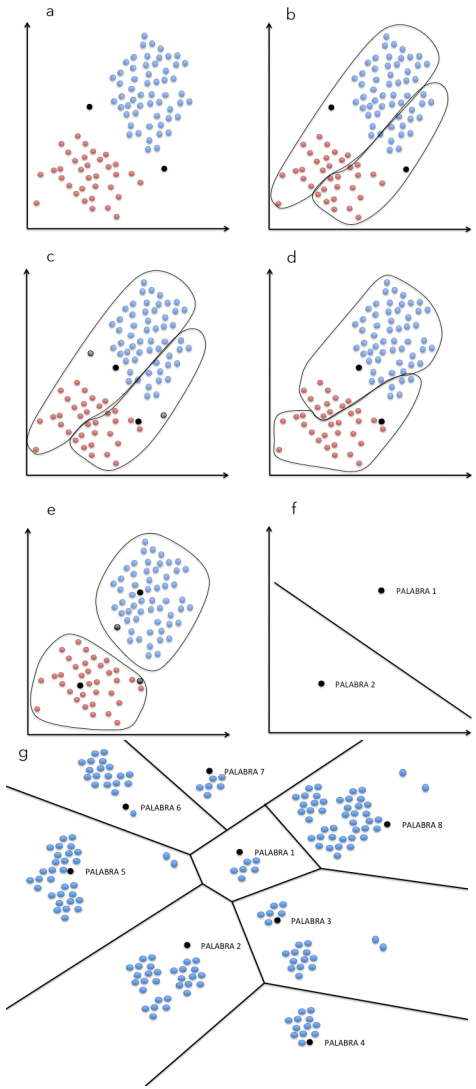


Figura 2.13. Etapas del algoritmo *K-means*.

tado 3 del Anexo A se explica con detalle el procedimiento de construcción del vocabulario visual y en el apartado 5 del Anexo A se amplia información sobre el algoritmo *K-means*.

El algoritmo *K-means* busca la partición mediante la iteración de dos etapas. La primera etapa consiste en asignar cada descriptor al centroide más cercano. En la segunda atapa se recalculan los centroides de cada región, calculando el vector de medias de los descriptores que han sido asignados a cada región. En las Fig. 2.13 se describe, a modo de ejemplo el caso de descriptores bidimensionales y de dos palabras visuales: el algoritmo *K-means* establecerá una partición del espacio en dos regiones, cada una asociada a una palabra como se describe a continuación:

a) Supongamos que los descriptores de la colección de imágenes configuran dos grupos separados (azul y rojo). El algoritmo empieza estableciendo dos centroides al azar (negro).

b) Asignamos cada descriptor al centroide más cercano.

c) Recalculamos los nuevos centroides de los grupos formados en la etapa anterior.

d) Repetimos la asignación de los descriptores al centroide más cercano.

e) El procedimiento prosigue recalculando los nuevos centroides.

f) El proceso iterativo se detiene cuando no se produce cambio apreciable en los centroides.

g) Ilustra la partición del espacio de descriptores en el caso de una vocabulario de más palabras. Dado un descriptor determinado, calcularemos el centroide más cercano, y le corresponderá la palabra representada por dicho centroide.

De esta manera, dada una imagen con un conjunto de descriptores, podemos usar los centroides obtenidos en el algoritmo *K-means* para atribuir la palabra visual a la que pertenece cada descriptor buscando el centroide más próximo.

Regresamos a la Fig. 2.6 para resumir el proceso de obtención de la representación *BoW* para un conjunto de imágenes:

1- Dada una colección de imágenes, se define una cuadrícula sobre las mismas.

2- Se calculan los descriptores.

3- Se cuantizan los descriptores en M clústeres, los cuales definirán un vocabulario visual de M palabras visuales. Una vez se dispone del vocabulario visual, los descriptores de cada imagen se asignan a la palabra visual más cercana.

4- Para obtener la representación *BoW* de una imagen dada, se calcula la frecuencia de cada palabra visual en la imagen.

Los métodos *BoW,* que representan una imagen como una colección desordenada de características locales, han demostrado excelente rendimiento en tareas de categorización de imágenes completas. Esta representación de la imagen, como ya hemos comentado anteriormente, no contiene información acerca de las relaciones espaciales entre palabras visuales, del mismo modo que la representación *BoW* de textos mezcla la información relativa al orden de la palabras en los documentos. Esta condición ha hecho que se vea limitada su capacidad descriptiva. Este método en particular, es incapaz de capturar formas o de separar un objeto de su fondo. La técnica habitual de *Bag-of-Words*, como se ha descrito anteriormente, no tiene en cuenta la información espacial. Sin embargo, información como las relaciones espaciales entre objetos vecinos o la posición absoluta de los objetos en ciertas escenas constituyen información muy valiosa muy útil para lograr mejores resultados de clasificación.

2.2.4 Añadiendo información espacial: *PHOW* (Pyramid Histogram Of visual Words)

Para superar las limitaciones del enfoque *BoW,* Lazebnik et al. (2006) proponen un mé-

todo basado en la pirámide espacial de coincidencias de Grauman & Darrel (2005), que incorpora con éxito información espacial al modelo *BoW*. Se denomina *PHOW* (Pyramid Histogram Of visual Words).

En nuestro trabajo hemos implementado esta metodología (Vedaldi et al., 2008) de histogramas en pirámide que consiste en la colocación de una secuencia de rejillas cada vez más finas sobre la imagen, y en la obtención de una suma ponderada del número de palabras visuales coincidentes que se producen en cada nivel de resolución (L).

Dada una resolución fija, se dice que dos puntos coinciden si están en el mismo cuadrante de la rejilla; las coincidencias encontradas en resoluciones más finas se ponderan más alto que las coincidencias encontradas en resoluciones más gruesas. La pirámide espacial resultante es una extensión de la representación de la imagen de *Bag-of-Words* (Fig. 2.14), equivaldría a un *Bag-of-Words* normal para L = 0. La imagen queda representada por un descriptor *PHOW* (Pyramid Histogram Of visual Words), un histograma en pirámide de palabras visuales (Fig. 2.15). Es posible ampliar información en el apartado 9 del Anexo A.

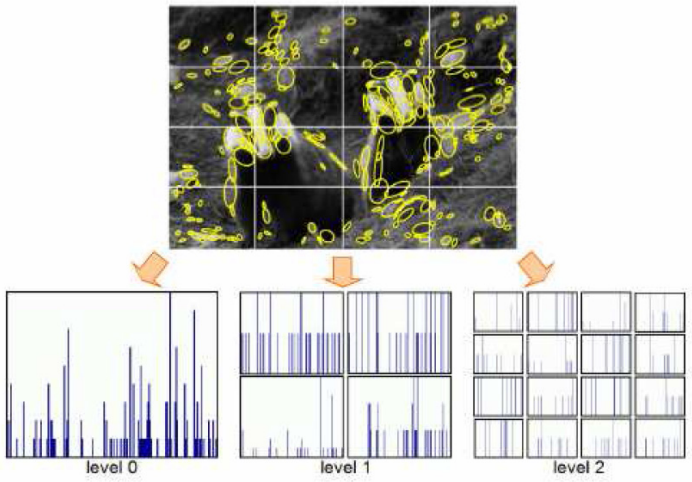


Figura 2.14. El método de Pirámide de coincidencias de Lazebnik et al. Conjunto de histogramas calculado sobre una descomposición de la imagen en pirámide de varios niveles (Bosch, 2007).

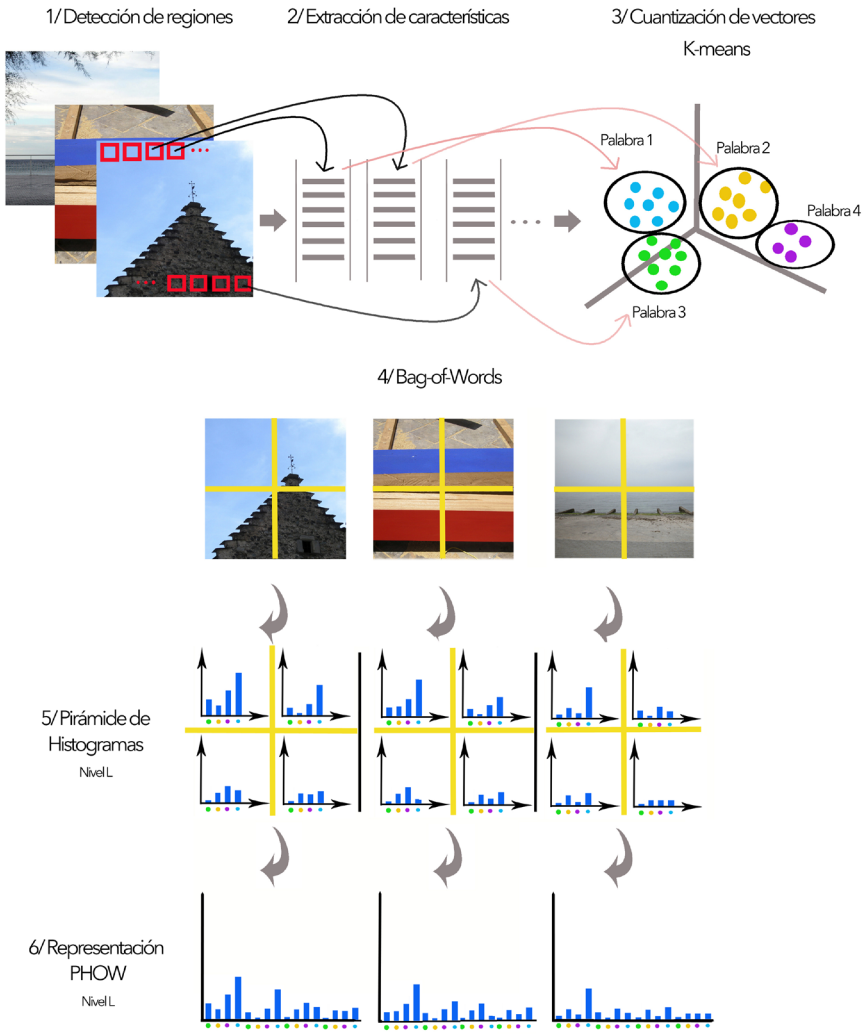
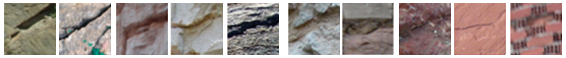


Figura 2.15. Esquema para calcular el modelo *PHOW* con imágenes. Añadimos dos pasos a la representación *BoW*. 5/ Pirámide de Histogramas de L niveles. 6/ Representación *PHOW* concatenando los histogramas de los diferentes niveles.

Palabra 1



Palabra 2

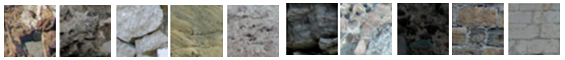


Figura 2.16. Muestras de regiones de imágenes correspondientes a dos palabras visuales de un vocabulario de 300 palabras. Podemos considerar que ambas palabras describen un contenido común; textura rocosa no homogénea, y en este sentido representan una sinonimia. Además, vemos que dentro de una misma palabra hay regiones que representan contenidos distintos, en unos casos el contenido es roca y en otros es muro, por tanto podríamos considerarlas polisémicas.

2.2.5 Problemas de polisemia y sinonimia en el vocabulario visual *BoW*

Cuando nos enfrentamos al análisis de un texto, una característica conocida de los vocabularios es la existencia de polisemia: una misma palabra puede tener varios significados, y sinonimia: varias palabras pueden referirse al mismo significado.

La representación *BoW* es fácil de construir. Sin embargo, adolece de dos inconvenientes (Fig. 2.16): polisemia (una sola palabra visual puede representar diferentes contenidos de la escena) y sinonimia (varias palabras visuales pueden caracterizar el mismo contenido de la imagen).

Para el análisis de la naturaleza "semántica" de las palabras visuales, Quelhas (2007) primero estudia comparativamente las veces que ocurren las palabras en las categorías de imágenes de ciudad y paisaje, partiendo de su representación *BoW* (Fig. 2.17). Observa que hay una gran mayoría de palabras visuales que aparecen en ambas clases: todos los términos están substancialmente presentes en la clase ciudad; sólo algunos de ellos no aparecen en la clase paisaje. Esto contrasta con lo que ocurre en los documentos de texto, en los que las palabras están, en general, vinculadas más específicamente a una determinada categoría. Estas observaciones reflejan que el contenido semántico contenido en las palabras visuales está fuertemente relacionado con cuestiones de sinonimia y polisemia. En la Fig. 2.18 podemos ver ejemplos de palabras visuales utilizadas por Quelhas (2007) y en la Fig. 2.16 ejemplos de palabras visuales encontradas en la colección de imágenes de Planas (2014).

Un factor que puede afectar en este sentido es el tamaño del vocabulario: la polisemia de palabras visuales puede ser más importante cuando se utiliza un vocabulario pequeño que cuando se utiliza un gran vocabulario. Por el contrario, con un vocabulario amplio, hay más posibilidades de encontrar muchos sinónimos que con uno pequeño.

En el ámbito de la representación de documentos de texto se ha introducido el modelado de aspectos latentes con el fin de hacer frente a estos problemas que plantean la sinoni-

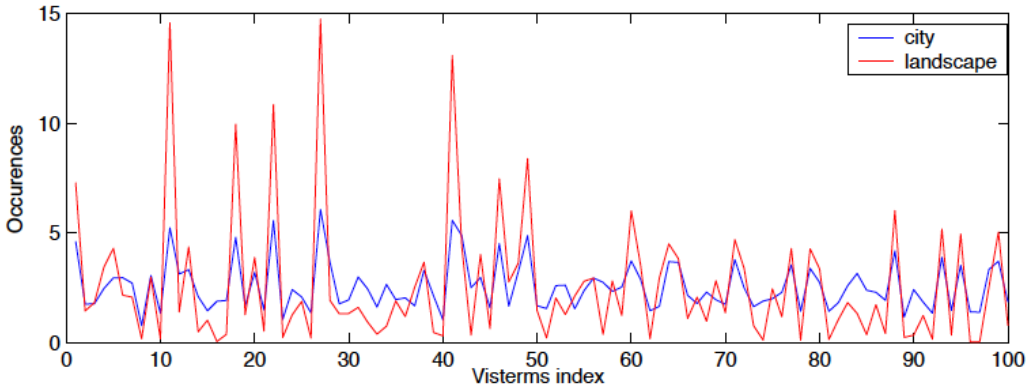


Figura 2.17. Tasa de ocurrencia por clase en una representación *BoW* de imágenes de la clase paisaje y de la clase ciudad. (Quelhas, 2007)

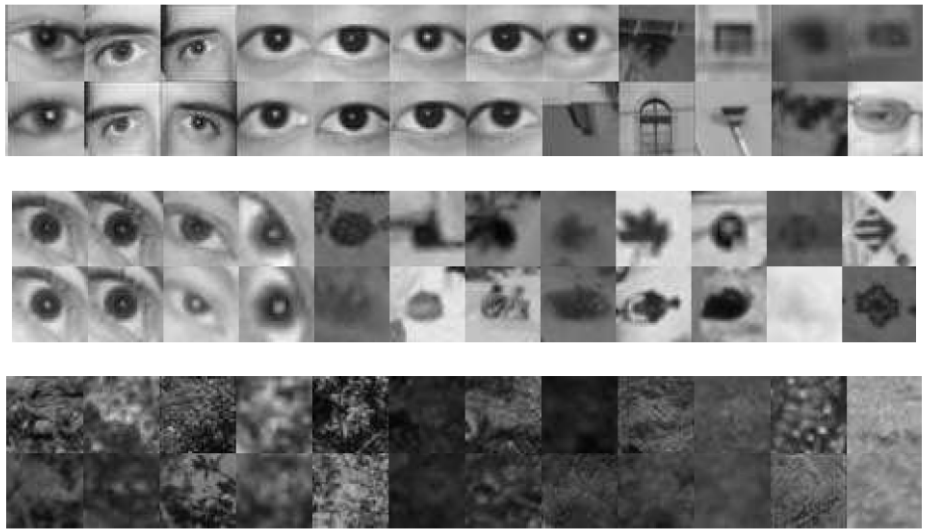


Figura 2.18. Muestra de 3 palabras de un vocabulario de 1000 palabras. La fila 1 (palabra 1) y la fila 2 (palabra 2) son ejemplos de sinonimia porque las dos representan el mismo significado; ojos humanos y a la vez son ejemplos de polisemia dentro de la propia palabra porque algunos patches, además de ojos, contienen puertas, ventanas y otro objetos. La fila 3 (palabra 3) es un ejemplo de polisemia porque las muestras de fondo se relacionan con textura de grano fino con diferente orígenes (roca, árboles, carreteras o textura de pared), que pueden provenir de muchos contextos. (Quelhas, 2007)

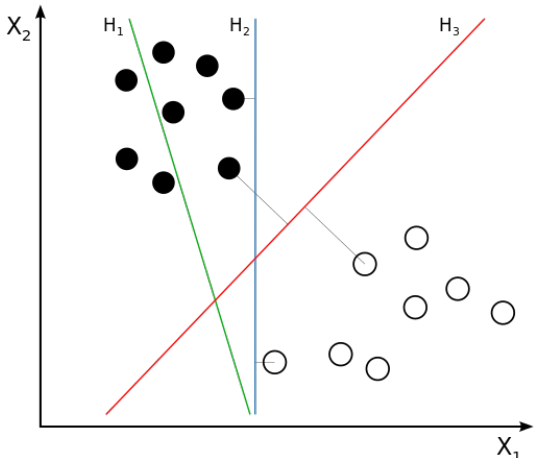


Figura 2.19. Esquema de *SVM*. H1 no separa bien las dos clases, por tanto no sería un buen hiperplano. H2 las separa, pero no es el más indicado, ya que la distancia de los puntos de las clases al plano es muy pequeña. La mejor opción es H3 (rojo) que está más espaciada de las dos clases. © ZackWeinberg (Màquina de vector de suport, 2014)

mia y la polisemia a la hora de asignar contenido semántico a un grupo de textos. De la misma forma se han aplicado con éxito estas técnicas para el análisis de imágenes. Una representación de aspectos latentes podría ser más estable para diferentes tamaños de vocabulario. En la próxima sección explicaremos con más detalle esta técnica de análisis probabilístico de aspectos latentes (*pLSA*).

2.3 CLASIFICACIÓN DE ESCENAS USANDO MODELOS ESTADÍSTICOS

Una vez tenemos construido el vocabulario visual de la colección de imágenes y hemos asignado a cada descriptor de la imagen la palabra que le corresponde, es posible obtener un nivel más de información si utilizamos modelos estadísticos que, convenientemente entrenados, sean capaces de discriminar patrones de distribución entre estas palabras (*SVM*) o modelos probabilísticos generativos (*pLSA*) que, de forma totalmente no supervisada, detecten aspectos semánticos latentes en nuestro conjunto de imágenes.

Por ejemplo, si en una determinada escena hemos determinado que se encuentran las palabras agua, arena y cielo distribuidas de una forma concreta (cielo en la parte superior, agua en la parte intermedia y arena abajo) podríamos categorizarla como paisaje de playa, o si encontramos coches y edificios, se trataría de una escena urbana. En el caso que nos ocupa de análisis de conjunto de obras de artista abstractas, nuestra esperanza es encontrar constantes en la obra o aspectos latentes significativos. Pasaremos a comentar con más detalle estos dos posibles modelos de clasificación.

2.3.1 Support Vector Machines (*SVM*)

Una máquina de vector de soporte (*SVM* o Support Vector Machines) (Boser, Guyon & Vapnik, 1992) consiste en un conjunto de algoritmos capaces de analizar datos y reconocer pa-

trones a través del aprendizaje supervisado. Estos métodos son utilizados principalmente en problemas de clasificación.

Una máquina de vector de soporte toma un conjunto de datos y predice, para cada una de estas entradas, a cual de las dos posibles clases pertenece. Mediante el entrenamiento con datos de entrada previamente clasificados, se establece un modelo que separa las dos clases entrantes. Este modelo establece una frontera entre las dos tipologías establecidas, esta se sitúa en el punto en el cual la diferencia entre clases sea la mayor posible y el margen de error sea cero (conjunto de datos separable) o mínimo (conjunto de datos no separable). Se llaman vectores de soporte a los puntos que conforman las dos líneas paralelas al modelo, siendo esta distancia la mayor posible (margen) (Fig. 2.19).

En nuestro caso, una imagen de entrada representada por su vector *BoW* podría ser clasificada empleando *SVM* (Fig. 2.20). Se implementan los descriptores de histogramas en pirámide *PHOW* para tener en cuenta la información espacial. Ver apartado 6 del Anexo A para ampliar la información.

2.3.2 Representación de aspectos latentes: Probabilistic Latent Sema*ntic Analysis (pLSA)*

Quelhas et al. (2005) proporcionan un enfoque mediante *Bag-of-Words* para modelar escenas visuales en colecciones de imágenes, basado en características locales invariantes y *pLSA*.

El *pLSA* es un modelo generativo que proviene del análisis estadístico de textos (Hofmann, 2001). En este tipo de análisis de texto se utiliza para descubrir los temas de un documento mediante su representación como *Bag-of-Words*. En este caso, hay "imágenes" en lugar de "documentos" y en lugar de "temas" se descubren "categorías de objetos". De esta forma una imagen que contiene diferentes tipos de objetos se modela como una mezcla de temas.

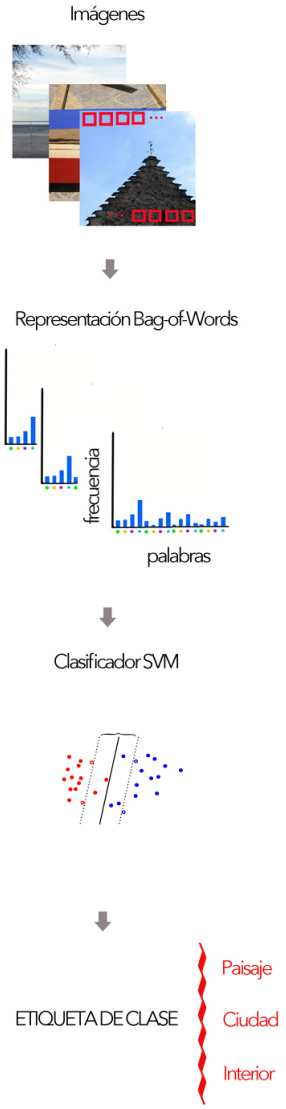
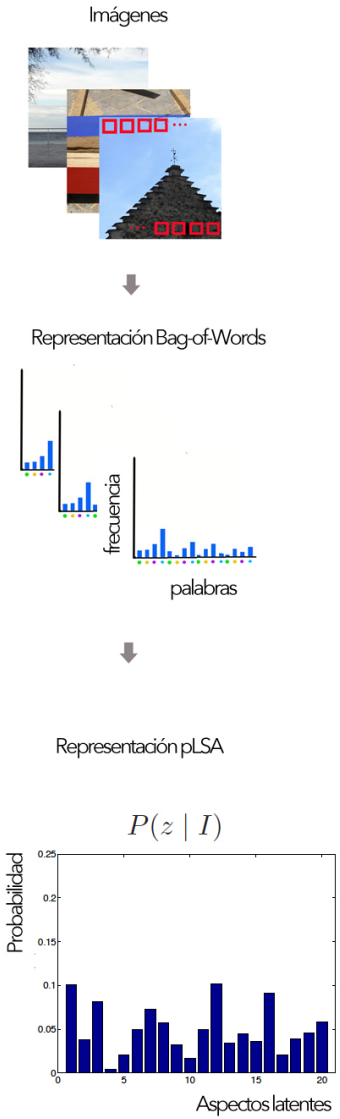


Figura 2.20. Esquema de modelo de clasificación utilizando representación de la imagen por *BoW* y *SVM*.



Este modelo tiene la doble capacidad de generar una representación de escena bajo-dimensional robusta, y también de capturar automáticamente los aspectos significativos de la escena.

Las aplicaciones del *pLSA* en el análisis estadístico de textos están orientadas a descubrir automáticamente los temas tratados en un documento, tomando como punto de partida la representación *BoW* de documentos.

La extensión del *pLSA* hacia el análisis de imágenes pasa por considerar las imágenes como documentos en un vocabulario visual establecido a partir de un proceso de cuantización como se ha señalado anteriormente. El método detectará en las imágenes categorías de objetos, patrones formales, de modo que una imagen que contiene varias tipos de objetos se modela como una mezcla de temas (Fig. 2.22).

Vamos a explicar el modelo en términos de imágenes, palabras visuales y aspectos. Disponemos de una colección de imágenes y de un vocabulario de palabras visuales. Podemos resumir las observaciones en una tabla de frecuencias, donde indicamos la frecuencia con que cada palabra visual ocurre en cada imagen .

El p*LSA* es un modelo estadístico generativo que asocia una variable latente con cada observación, entendiendo por observación la ocurrencia de una palabra visual en una imagen dada. Estas variables, normalmente llamadas aspectos, se utilizan para construir un modelo de probabilidad conjunta sobre las imágenes y las palabras visuales. (Ver detalles en el apartado 4 del Anexo A).

Con el *pLSA* finalmente obtenemos una nueva representación para las imágenes de la colección basada en la distribución de aspectos (Fig. 2.21). De hecho, también es posible hallar la distribución de aspectos para una imagen cualquiera que no forme parte de la colección inicial (Quelhas, Monay, Odobez, Gatica-Perez, Tuytelaars & Van Gool, 2005; Bosch, Zisserman & Muñoz, 2006).

Figura 2.21. Esquema de modelo de clasificación utilizando representación de la imagen por *BoW* y *pLSA*.

2.4 MEDIDA DE SIMILITUD: DISTANCIA DE BHATTACHARYYA

El *pLSA* nos proporciona una representación de la imagen basada en su distribución de aspectos. Para realizar una agrupación basada en estos aspecto, podemos seguir dos estrategias:

1- Agrupación en función del aspecto más probable: los grupos se establecen en función de la mayor probabilidad de un aspecto u otro, estableciendo una ordenación de mayor a menor probabilidad de contener el aspecto concreto.

2- Considerar la distribución de probabilidad de todos los aspectos. En la tesis se realiza un análisis de agrupación jerárquica que permite construir un dendograma (Fig. 2.23) con las imágenes de toda la colección. Este gráfico posee un notable interés a nivel informativo ya que constituye una herramienta visual que el artista o el experto puede utilizar para explorar y analizar la complejidad del conjunto de imágenes objeto de estudio.

En estadística, la distancia Bhattacharyya (Bhattacharyya, 1943) mide la similitud de dos distribuciones de probabilidad discretas o continuas (Información detallada en el apartado 7 del Anexo A). Mediante el cálculo de esta distancia entre los histogramas de aspectos se puede construir el dendograma.

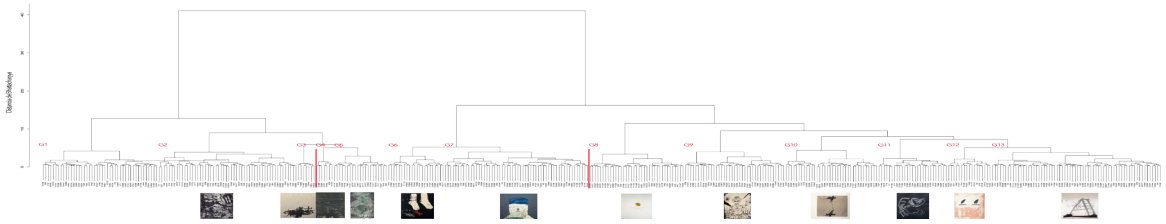


Figura 2.23. Dendograma basado en la distancia de Bhattacharyya.

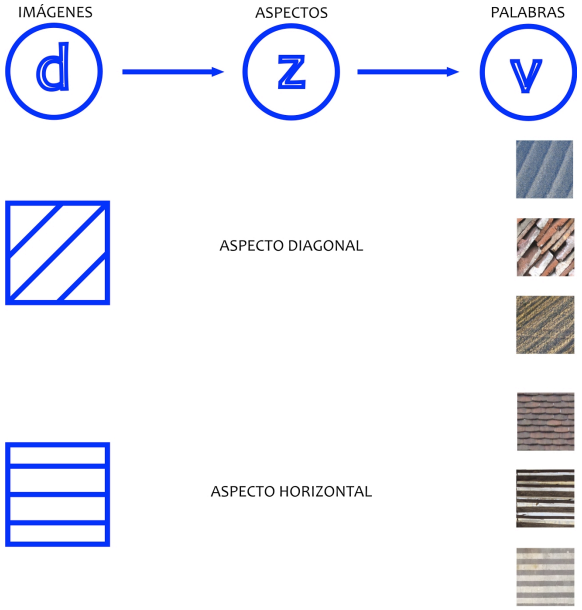


Figura 2.22. El método *pLSA* captura la co-ocurrencia de palabras visuales entre imágenes.

A modo de ejemplo, en la Fig. 2.24 mostramos las 16 imágenes que presentan el aspecto 7 de los 13 obtenidos en la colección de obras de Tàpies mediante *pLSA* y a la derecha mostramos la representación de estas mismas imágenes como histogramas de la distribución de estos aspectos. Aplicando la distancia de Bhattacharyya a estas imágenes obtenemos una agrupación por similitud que se muestra en la Fig. 2.25. Como puede observarse en los histogramas de la derecha de la Fig. 2.25, la agrupación se realiza por similitud de distribución de frecuencias en los histogramas.

2.5 OTRO TIPO DE DESCRIPTORES: TEXTURA DE HARALICK

La representación de la información contenida en las imágenes retinianas permite al sujeto obtener una descripción con significado de la escena observada. Es decir, le va a permitir describir lo que está presente en el entorno y dónde esta localizado. Pero antes de poder identificar un objeto, debemos establecer los limites que lo definen y separan del resto. El sistema visual humano segmenta el input óptico en regiones basándose en las diferencias de las propiedades que poseen áreas adyacentes de la superficie. Una importante fuente de información para este proceso son las texturas, ya que diferentes objetos suelen tener distintas propiedades de textura (Pérez, 1995).

En visión por computador, la textura es una de las características importantes utilizadas para la identificación de objetos o de zonas de interés en una imagen. Aunque intuitivamente se pueden asociar diversas propiedades de las imágenes, tales como suavidad, rugosidad, regularidad, etc. (Gonzalez & Woods, 2008; Bharati, Liu & MacGregor, 2004), realmente no existe una definición formal o completa de textura. Muchos investigadores describen la textura utilizando varias definiciones. Russ (1999) considera la textura de una imagen como la variación entre píxeles en una pequeña vecindad de una imagen. Alternativamente, la textura puede describirse como un atributo que representa la distribución espacial de los niveles de intensidad en una región dada de una imagen digital (Bharati et al.,2004). Existe, en ambas definiciones, el concepto de variación espacial en un entorno

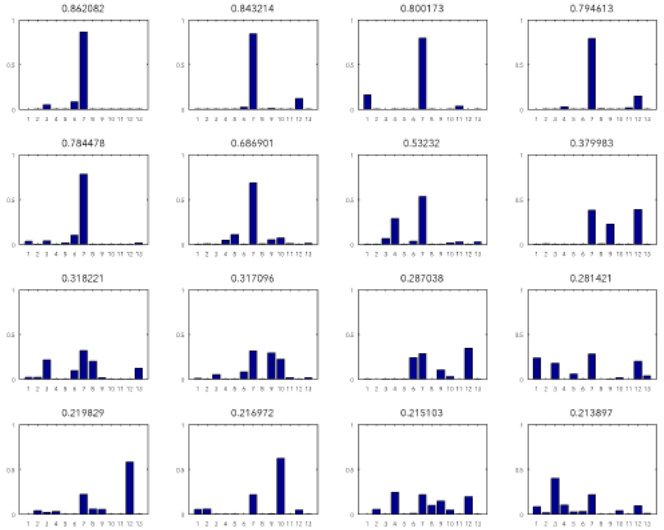


Figura 2.24. A la izquierda se muestran las 16 imágenes de la obra de Tàpies que presentan el aspecto 7 (Fondo Tramado) con mayor probabilidad y a la derecha su representación como histogramas de aspectos.

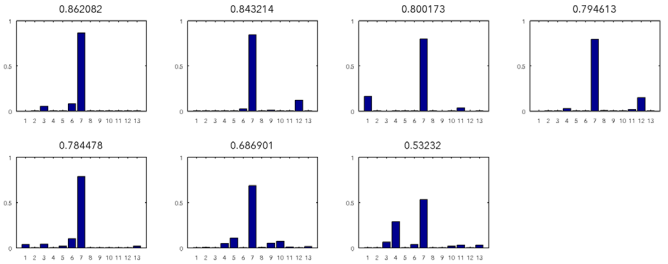


Figura 2.25. A la izquierda se muestran las 7 imágenes de la obra de Tàpies cuyas distribuciones de probabilidad presentan mayor similitud en base a la distancia de Bhattacharyya y a la derecha su representación como histogramas de aspectos. Distancia de Bhattacharyya: Grupo 3

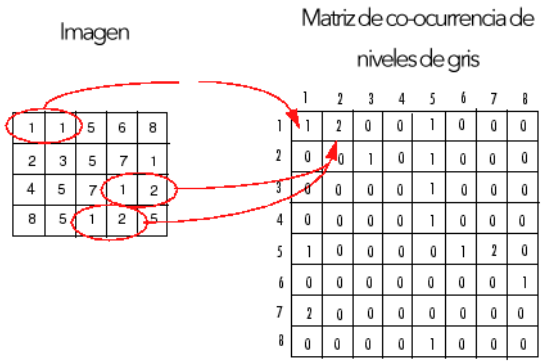


Figura 2.26. Cálculo de la matriz de coocurrencia de niveles de gris de una imagen de 4 x 5 píxeles.

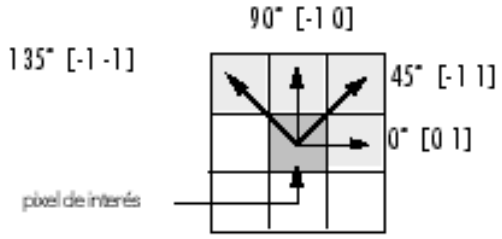


Figura 2.27. Ilustra el cálculo del desplazamiento para un sólo píxel.

de vecindad (Ruiz, 2011).

Los descriptores de textura definidos por Haralick (Haralick, 1973) son un conjunto de medidas de textura basadas en la matriz de co-ocurrencia. Se puede ampliar información sobre estos descriptores en el apartado 2 del Anexo A.

Son de naturaleza estadística y para su cálculo, es necesario asumir que la totalidad de la información textural de una imagen está contenida en las relaciones espaciales que se dan entre los distintos niveles de gris de un objeto. Incorporan información espacial en forma de posición relativa entre niveles de intensidad dentro de la textura.

Para cada imagen se crea una matriz de co-ocurrencia de niveles de gris (*GLCM,* Gray-Level Co-occurrence Matrix) mediante el cálculo de la frecuencia con la que un píxel con un nivel de gris determinado *i* se presenta horizontalmente adyacente a un pixel con el valor de *j*. Cada elemento *(i, j)* de la matriz de co-ocurrencia especifica el número de veces que el píxel con valor *i* ocurrió horizontalmente adyacente a un píxel con valor *j*. Se puede especificar el número de niveles de gris (en nuestro caso hemos utilizado 256 niveles).

La Fig. 2.26 muestra la forma en que se calcularían algunos valores de la matriz de co-ocurrencia para una imagen de 4 x 5 píxeles.

La Fig. 2.27 ilustra la distancia D, en número de filas y columnas, entre el píxel de interés y su vecino. En este caso concreto D es igual a 1, pero se podría considerar para mayores distancias (0 D, -D D, -D 0, -D -D). En el caso que nos ocupa hemos considerado 4 distancias diferentes: 1,10, 20 y 50 píxeles. Así, para cada imagen se han calculado en total 16 descriptores de Haralick (basados en 256 niveles de gris) correspondientes a; las 4 distancias consideradas multiplicadas por los 4 ángulos de cada distancia.

A partir de estos descriptores, se pueden inferir unas propiedades estadísticas que proporcionan información acerca de la textura global de la imagen (Fig. 2.28):

Contraste: El contraste en una imagen se refiere a la diferencia relativa en la intensidad de

un punto o zona.

Correlación: Esta propiedad estadística indica la fuerza y la dirección de una relación lineal y proporcional entre dos variables.

Energía: También se conoce como uniformidad de la energía o segundo momento angular.

Homogeneidad: Se refiere a la falta de variabilidad.

En el presente estudio utilizaremos estos descriptores como contrapunto para comparar los resultados con los obtenidos a partir de los descriptores *SIFT*.

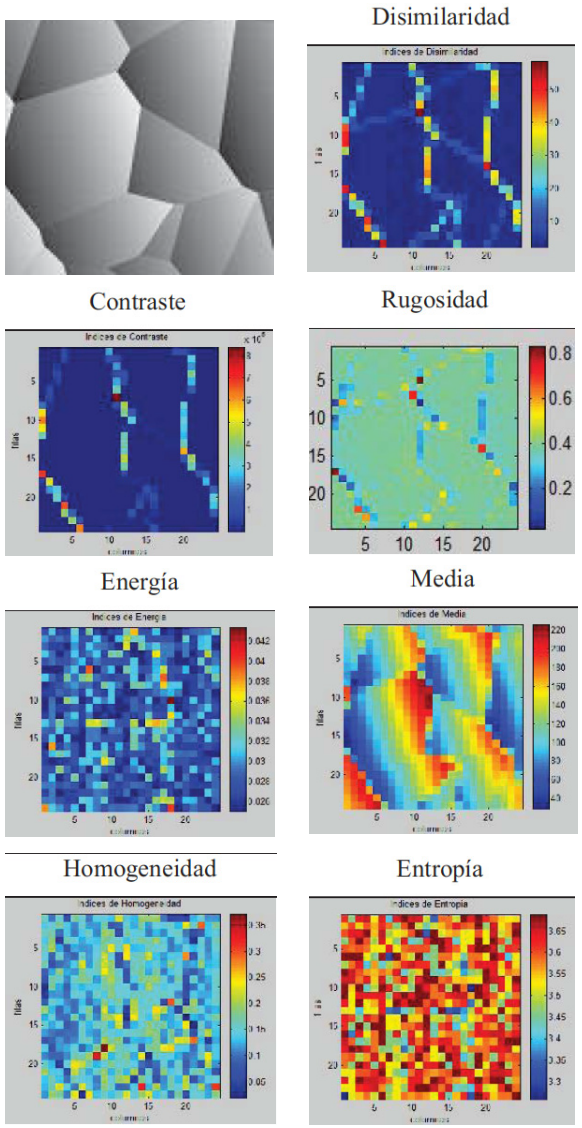


Figura 2.28. Ejemplo de algunas propiedades estadísticas que se pueden calcular sobre la textura de una imagen y después visualizar. En primer lugar, arriba a la izquierda, vemos la imagen sobre la que se realizan los cálculos y después propiedades como contraste, energía, homogeneidad, disimilaridad, rugosidad, media y entropía. Estas tres propiedades las hemos utilizado en nuestro análisis junto con la correlación.

3 - RESULTADOS

En resumen, podríamos decir que existen tres modelos principales para anotar imágenes:

1- Anotación totalmente manual: es un trabajo pesado ya que es necesario el esfuerzo humano para etiquetar cada imagen.

2- Anotación automática (métodos no supervisados): se realiza sin la ayuda del hombre pero, dadas las limitaciones que todavía existen en las técnicas de visión artificial y el procesado de imágenes, no se garantiza la precisión de los resultados.

3- Anotación semi-automática: divide la base de datos de imágenes a tratar en dos partes; una para entrenamiento y otra para validación. El conjunto de entrenamiento es etiquetado manualmente de modo que la relación entre imágenes y anotaciones es precisa. A continuación, se emplean las relaciones aprendidas entre las imágenes y las anotaciones para generar etiquetas en el conjunto que se desea validar.

Según el tipo de clasificador empleado, los métodos de anotación de imágenes pueden ser catalogados en dos categorías:

1- Métodos discriminativos: tratan las anotaciones como clases y emplean clasificadores entrenados para obtener fronteras que permitan discriminar entre aquellas imágenes en las que aparece un concepto y las que no. Dentro de este tipo de métodos discriminativos clásicos encontramos las redes neuronales o las máquinas de vector de soporte (*SVM*).

2- Métodos generativos: a diferencia de los anteriores, que sólo discriminan entre casos positivos y negativos, estos métodos tratan de inferir las probabilidades conjuntas entre imágenes y anotaciones. Esta información, si bien más compleja de obtener, proporciona un conocimiento extra sobre la generación de los datos. Introducen variables latentes que asocian a los conceptos semánticos de las etiquetas. Algunos de estos métodos que capturan información de co-ocurrencia son el *pLSA* (Probabilistic Latent Semantic Analysis) (Hofmann, 2001) y el Asignación de Dirichlet latente LDA (Latent Dirichlet Allocation) (Blei, Ng, Jordan & Lafferty, 2003).

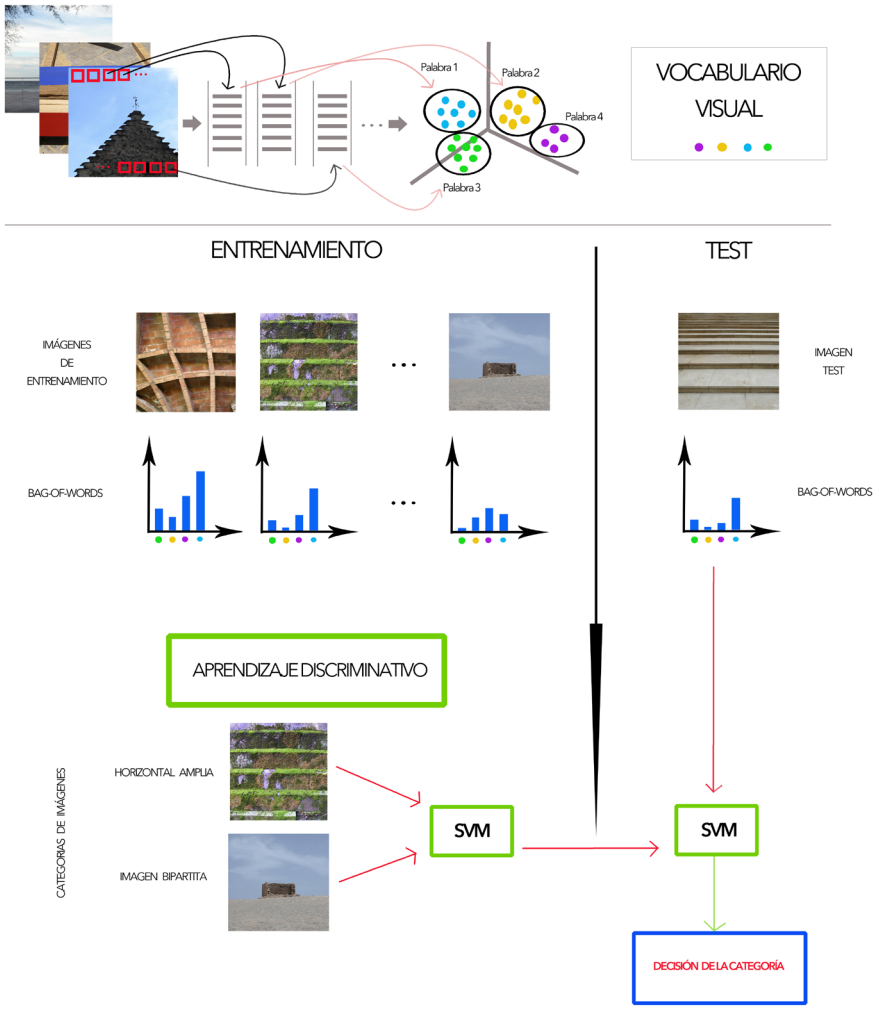


Figura 3.1. Esquema del programa de aprendizaje supervisado discriminativo.

En este capítulo se describen en detalle los programas informáticos y los experimentos que, en base a las metodologías descritas en el capítulo 2, se han realizado sobre los conjuntos de obras de artista detallados en el apartado 1.9.3, que constituyen el objeto de estudio que nos ocupa.

Los programas informáticos son 4:

1- Programa de aprendizaje supervisado discriminativo.

2- Programa de aprendizaje no supervisado generativo.

3- Programa de cálculo de distancias y elaboración del dendograma.

4- Programa de agrupación en base a descriptores de textura.

Los experimentos realizados con estos programas son los siguientes:

1- Primer experimento de Aprendizaje supervisado discriminativo.

2- Segundo experimento de Aprendizaje no supervisado generativo. En caso se ha trabajado la metodología sobre dos colecciones diferentes:

2.1 - Un primer caso con el conjunto de imágenes digitales de fotografías de Planas.

2.2- Un segundo caso sobre el conjunto de imágenes digitales de pinturas de Tàpies. Sobre este mismo conjunto de imágenes además se ha implementado una agrupación clúster basada en el cálculo de la distancia de Bhattacharyya entre los histogramas de aspectos resultantes del *pLSA*, y también se ha probado el rendimiento de clasificación utilizando los descriptores de textura de Haralick.

Dedicaremos el resto del capítulo a describir los resultados que se han obtenido en cada caso con mayor detalle.

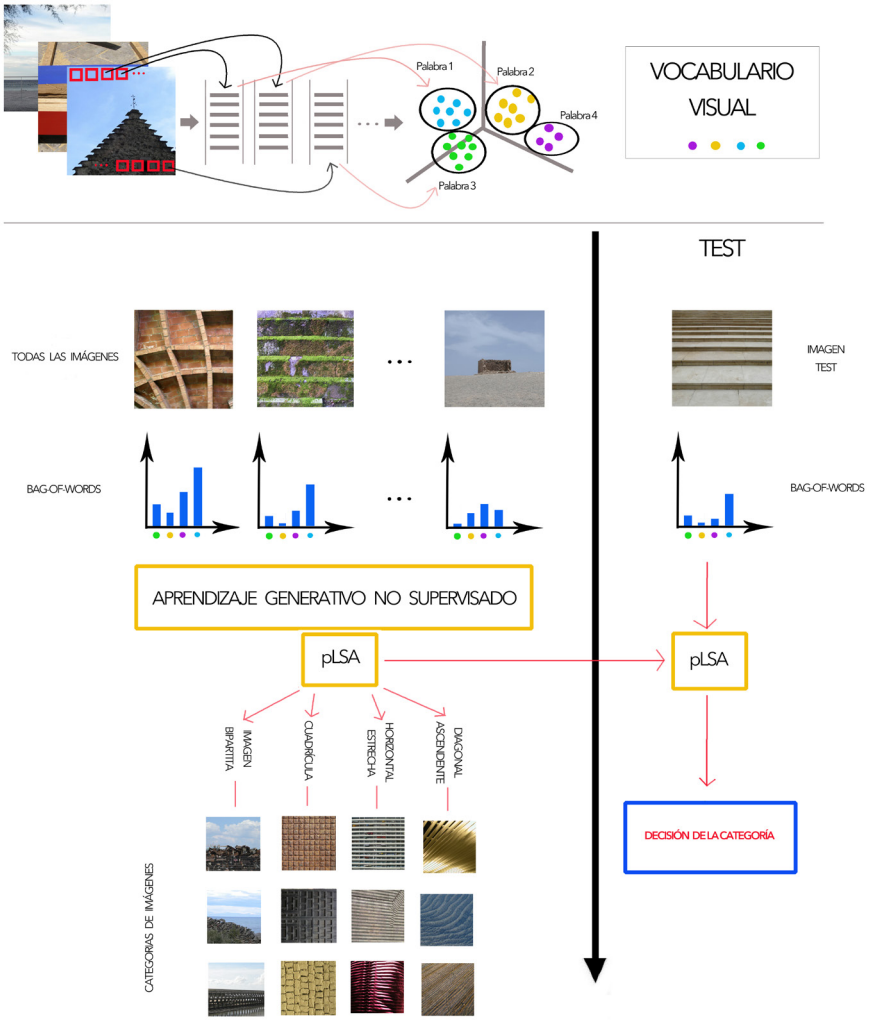


Figura 3.2. Esquema del programa de aprendizaje no supervisado generativo.

3.1 DESARROLLO DE PROGRAMAS INFORMÁTICOS

El desarrollo de estos programas se ha llevado a cabo mediante scripts escritos en *MATLAB*, versión 2013a (The MathWorks) (8.1.0.604). Los descriptores *SIFT* y el vocabulario de palabras visuales se han implementado mediante funciones disponibles en la biblioteca de código abierto VLFeat, versión 0.9.16 (Vedaldi & Fulkerson, 2008). El *pLSA* ha sido implementado mediante funciones desarrolladas.

El tiempo de procesado de la colección de 2846 imágenes con un ordenador de 2.4 GHz, es alrededor de 20 minutos.

3.1.1 Programa de aprendizaje supervisado discriminativo

En la Fig. 3.1 se muestra un esquema del programa desarrollado en *MATLAB*: con el conjunto de imágenes de entrenamiento se construye un vocabulario visual mediante el modelo *Bag-of-Words*. Después se computa la representación *PHOW* (Pyramid Histogram Of visual Words) de cada imagen. Por último, se calcula el mapa de características asociadas con la χ^2 -*kernel* y se estima el clasificador *SVM* multiclase. Existe código eficiente disponible para calcular estos mapas de características en la libreria VLFeat de código abierto (Vedaldi & Fulkerson, 2008).

3.1.2 Programa de aprendizaje no supervisado generativo

En la Fig. 3.2 se muestra un esquema del programa desarrollado en *MATLAB*: el proceso de este segundo programa consiste también en construir el vocabulario palabras visuales mediante el modelo *Bag-of-Words* y descriptores *PHOW,* pero en esta ocasión a partir del

total de las imágenes, para después computar la representación *PHOW* obtenida a través del modelo *pLSA* (Análisis Probabilístico de Aspectos Latentes*)* (Más detalles en el apartado 4 del Anexo A: Representación de aspectos latentes mediante *pLSA*). De esta forma finalmente se obtiene una distribución de aspectos latentes para cada imagen.

Por último, se pueden establecer un orden y clasificar las imágenes en función de la probabilidad de los aspectos que contienen. Con este sistema también es posible, dada una imagen problema, determinar a que categoría *pLSA* pertenece (Fig. 3.2).

3.1.3 Programa de cálculo de distancias y elaboración del dendograma

Este programa calcula la distancia de Bhattacharyya (Bhattacharyya, 1943) entre las distribuciones de probabilidad de los aspectos latentes entre pares de imágenes de las colecciónes. Como método de aglomeración se ha utilizado el de Ward (Gordon, 1999) .

3.1.4 Programa de agrupación en base a descriptores de textura

Este programa calcula los descriptores de textura de Haralick según se detalla en el apartado 2.5 de la Metodología y se amplia en el apartado 2 del Anexo A, para después agruparlos mediante un algoritmo *K-means* (más detalles en el apartado 5 del Anexo A).

3.2 EXPERIMENTO DE APRENDIZAJE SUPERVISADO DISCRIMINATIVO

Se utiliza el programa descrito en el apartado 3.1.1. En este experiencia de aprendizaje supervisado, se aplica un modelo discriminativo que utiliza el clasificador *SVM* sobre el

Figura 3.3. Imágenes de la colección de Planas.

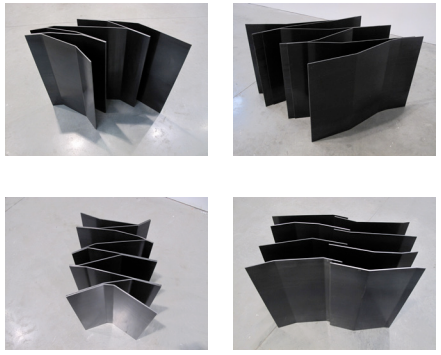


Figura 3.4. Imágenes de esculturas de Planas.

vocabulario *BoW* elaborado a partir de descriptores *SIFT* y representación *PHOW*. (Fig. 3.1)

La base de datos con la que se trabaja es la del artista Miquel Planas, constituida por 2846 imágenes digitales (Fig. 3.3). Después de capturar las instantáneas, el artista utiliza la colección para su observación y estudio; suponen una herramienta de conocimiento y de acceso a su poética personal (Fig. 3.4). Realiza sobre ellas una labor taxonómica, ordena y organiza estas imágenes para comprender las relaciones que él mismo, de manera inconsciente, ha establecido entre ellas. Al constituir una tarea personal, el artista percibió que con el paso del tiempo, las clasificaciones se iban repitiendo, provocando una cierta endogamia en los grupos definidos. Los resultados de las clasificaciones que realizaba venían en gran medida marcados o señalados por su propia narrativa, por su mirada subjetiva, se referían más al sujeto actor, que al propio objeto. Este hecho evidenció que las clasificaciones obtenidas no respondían a criterios objetivos y se centraban más en premisas o en valores ya consolidados y reconocidos por el propio creador, muchos de ellos determinados más por motivos cronológicos, vivenciales o conceptuales del autor, que por la propia esencia de las imágenes captadas, reduciendo o incluso eliminado en gran medida la posibilidad de aprendizaje y de conocimiento que podría aportar este cuerpo de imágenes.

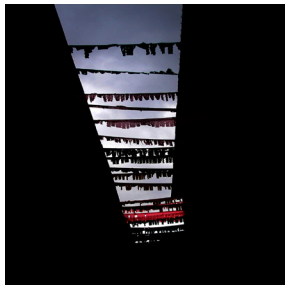
Paralelamente se añadió la dificultad cada vez mayor de ordenar y catalogar con criterios estables las imágenes que se habían generado día a día; el número de ellas iba aumentado exponencialmente, pero no así el número de grupos, de apartados, que permanecía similar, este hecho permitió concluir que las catalogaciones que se habían asignado hasta el momento no eran suficientemente sólidas, y que los resultados de carácter creativo que se obtenían eran escasos e incluso reiterativos.

Es en este momento cuando se esbozó la posibilidad de iniciar una investigación que permitiera que la referida búsqueda se plantease desde mecanismos o sistemas analíticos objetivos, que tendieran a superar maneras de mirar básicamente personales y parciales. En este sentido, una metodología basada en aspectos computacionales, propiciaba la objetividad reclamada, así como la posibilidad de obtener un mayor y diverso número de resultados.

Al proponer una investigación de estas características, dentro del ámbito de las bellas artes, se planteó que los resultados obtenidos se podrían extrapolar a todo tipo de acciones entorno a la creación, en las que la comparación entre imágenes fuera la característica principal, logrando aplicaciones encaminadas al aprendizaje, al conocimiento y a la investigación en imágenes, tal y como se expondrá más adelante. En el desarrollo de la presente investigación la sido de gran importancia el hecho de tener al alcance el fondo consolidado del artista Planas, constituido por 2846 imágenes digitales, todas ellas pertenecientes al mismo autor y con un perfil coherente, pero de estructura y características formales, conceptuales y temáticas diversas. La diversidad y cantidad documental al abasto, junto con la participación e implicación directa en este proceso del propio autor, permitiría, dentro de una metodología rigurosa, reaccionar ante futuros resultados parciales o poco precisos, así como comprobar y contrastar los resultados de forma inmediata y fiable con el creador.

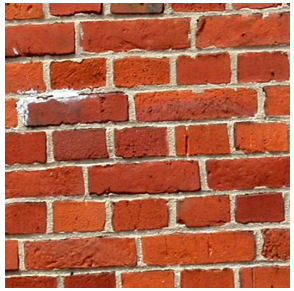
El tamaño de las imágenes del artista está comprendido entre 480 x 480 píxeles y 1400 x 1400 píxeles, pero el sistema, antes de iniciar el procesado, reescala a 480

Piedra Irregular
Abreviada como PI



Siluetas
Abreviada como SI

Piedra Texturada
Abreviada como PT



Piedra Geométrica
Abreviada como PG

Arquitectura Central
Abreviada como AC

Figura 3.5. Una imagen ejemplo de cada categoría que se establece en el conjunto de 150 imágenes del artista Planas.

Arquitectura Central
Abreviada como AC

Piedra Geométrica
Abreviada como PG

Piedra Irregular
Abreviada como PI

Piedra Texturada
Abreviada como PT

Siluetas
Abreviada como SI

Figura 3.6. 5 imágenes de cada categoría establecida en el conjunto de 150 imágenes de Miquel Planas.

píxeles las imágenes que superan este tamaño. El conjunto de partida del experimento lo constituyen 150 imágenes del artista previamente clasificadas y etiquetadas manualmente por miembros del equipo de investigación y expertos en arte en 5 categorías (Fig. 3.5) que se corresponden con 5 tipologías identificadas en el total de imágenes:

En la Fig. 3.6 se muestran 5 ejemplos de imágenes de cada una de las categorías asignadas a la muestra de 150.

Se divide el conjunto de 150 imágenes etiquetadas en dos grupos: 75 imágenes de entrenamiento y 75 imágenes de prueba (15 imágenes de cada categoría). El objetivo de esta clasificación es entrenar al sistema y generar el vocabulario para después predecir la categoría a la que pertenecen las 75 imágenes del grupo de prueba.

Como se detalla en la Fig. 3.1, con el conjunto de imágenes de entrenamiento se construye un vocabulario de 300 palabras visuales mediante el modelo *Bag-of-Words*. Después se computa la representación *PHOW* (Pyramid Histogram Of visual Words) de cada imagen. Por último, se calcula el mapa de características asociadas con la $\chi 2$-kernel y se estima el clasificador

	AC	PG	PI	SI	PT
AC	0.79	0	0.04	0.17	0.01
PG	0.14	0.41	0.31	0.11	0.03
PI	0.01	0.19	0.57	0	0.23
SI	0.11	0.01	0	0.85	0.03
PT	0	0.01	0.37	0	0.61

Figura 3.7 La verdadera categoría la indica la fila y la categoría pronosticada se encuentra en la columna. Las categorías son AC: Arquitectura Central, PG: Piedra Geométrica, PI: Piedra Irregular, SI: Siluetas, PT: Piedra Texturada. Las celdas de la tabla indican la media de la proporción de errores de predicción de cada categoría. El color verde corresponde a la media de acierto ara cada categoría.

	AC	PG	PI	SI	PT
AC	0.04	0	0.003	0.03	0.007
PG	0.02	0.04	0.05	0.011	0.011
PI	0.01	0.04	0.02	0	0.011
SI	0.01	0.01	0	0.014	0.011
PT	0	0.01	0.011	0	0.017

Figura 3.8 La verdadera categoría la indica la fila y la categoría pronosticada se encuentra en la columna. Las categorías son AC: Arquitectura Central, PG: Piedra Geométrica, PI: Piedra Irregular, SI: Siluetas, PT: Piedra Texturada. Las celdas de la tabla indican el error estándar de la proporción de imágenes clasificadas en categorías inadecuadas.

Figura 3.9. Se muestran dos ejemplos de imágenes que pertenecen a la clase Arquitectura Central y que el sistema ha clasificado erróneamente como Siluetas. Podemos comprobar con facilidad que, a pesar de que presentan construcciones situadas en el centro de la composición de la imagen, el grado de contraste acentuado de la escena haría posible también su pertenencia a la clase Siluetas, por lo que el error de clasificación del sistema podría calificarse de comprensible.

Figura 3.10. Se muestran un ejemplo de imagen que pertenece a la clase Arquitectura Central y que el sistema ha clasificado erróneamente como Piedra Irregular. La escena presenta una chimenea situada en el centro de la composición, pero el fondo corresponde a un conjunto de piedras del tipo gravilla que fácilmente podría corresponder a la categoría de Piedra Irregular. Por lo tanto también en este caso podemos pensar que el error es lógico ya que la figura pertenece a la clase Arquitectura Central pero el fondo perfectamente podría asignarse a Piedra Irregular.

Figura 3.11. Se muestran dos ejemplos de imágenes que pertenecen a la clase Siluetas y que el sistema ha clasificado erróneamente como Arquitectura Central. La imagen de la izquierda presenta una curiosa distribución de los elementos en luz y sombra que confiere a la parte central un protagonismo luminoso que bien se podría corresponder con una figura central, y la imagen de la derecha pertenece a la clase Siluetas por su elevado contraste, pero también contiene un edificio claramente en posición central. Estos dos errores son también comprensibles.

SVM multiclase.

Con el fin de evaluar el rendimiento de la metodología, clasificamos un conjunto de imágenes prueba (75 imágenes, 15 imágenes de cada categoría). El proceso de clasificación se repite 10 veces, cambiando de forma aleatoria las imágenes que componen los conjuntos de entrenamiento y de prueba.

La Fig. 3.7 muestra la media y la Fig. 3.8 el error estándar de la proporción de errores de clasificación de cada categoría.

A continuación comentaremos en detalle los resultados de la clasificación:

La clase Arquitectura Central (AC) presenta un porcentaje de acierto del 79 %. La mayoría de confusiones se producen con la clase Siluetas. En las Fig. 3.9 y 3.10 se muestran y comentan algunos ejemplos de estos errores con las clases SI y PI.

La categoría Siluetas (SI) es la que presenta una mayor proporción de clasificación correcta, el 85%. Curiosamente, el mayor porcentaje de confusión se produce con la clase arquitectura central. Se pueden ver ejemplos en la Fig. 3.11.

Posteriormente, encontramos las categorías Piedra Texturada (PT) y Piedra Irregular (PI) con el 61% y el 57% de clasificación correcta. La mayoría de los errores de clasificación en estas categorías se deben a confusiones producidos entre estas dos mismas categorías (Fig. 3.12).

La categoría Piedra Geométrica (PG) tiene la menor proporción de clasificación correcta, el 41%. La mayoría de los errores se producen con la categoría de Piedra Irregular (PI). Ver comentarios detallados en la Fig. 3.13.

Teniendo en cuenta que las metodologías utilizadas en este apartado han demostrado repetidamente en la literatura su buen rendimiento a la hora de clasificar imágenes y en vista de los resultados obtenidos en este experimento, podemos concluir que mantienen su buen comportamiento cuando se enfrentan a bases de datos en las que las categorías existentes vienen determinadas por contenidos semánticos involucrados en procesos de ideación y creación artística.

El porcentaje medio de clasificación correcta es de aproximadamente el 70 % y gran parte de los errores de predicción de categorías en el conjunto de pruebas se

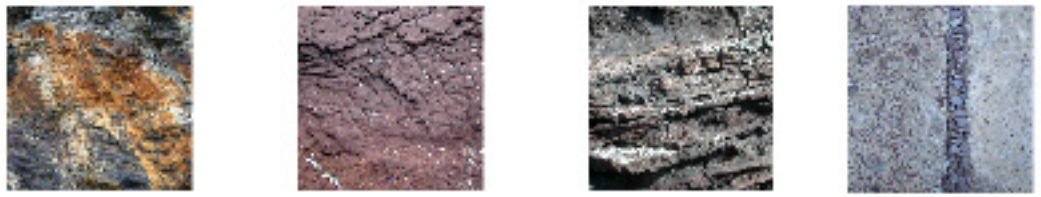


Figura 3.12. Las 3 primeras imágenes desde la izquierda, corresponden a tres ejemplos de imágenes que han sido clasificadas como Piedra Texturada por los expertos y que el sistema ha clasificado como Piedra Irregular. Finalmente la cuarta imagen, a la derecha, corresponde al caso inverso, una imagen perteneciente a la categoría de Piedra Irregular a la que el sistema le ha otorgado la categoría de Piedra Texturada Dada la tipología de las imágenes y los conceptos representados en estas categorías, no siempre resulta fácil para los expertos determinar el límite donde acaba una distribución irregular y en el que comienza una textura, así que, como en los casos comentados con anterioridad, estas confusiones del sistema de clasificación también resultan comprensibles.

Figura 3.13. Las 4 imágenes corresponden a ejemplos de imágenes que han sido clasificadas como Piedra Geométrica por los expertos y que el sistema ha clasificado como Piedra Irregular. Las imágenes pertenecen ciertamente a muros construidos con piezas geométricas de una forma más o menos regular, pero existe cierto grado de variabilidad en el contraste y algunos elementos que interfieren, por lo que también es de esperar el grado de confusión. Es difícil en ocasiones para el experto separar el conocimiento que tiene sobre que la construcción de las paredes se realiza con elementos geométricos de la verdadera percepción de la escena, que al final, debido a factores como la iluminación y el contraste, se asemejan más a un aspecto irregular o textura que a una distribución geométrica.

pueden justificar dado que la categorización de imágenes como las que trata el presente estudio no siempre resulta sencilla; se trata de un conjunto de imágenes tomadas, la mayoría, de exteriores y captadas desde diferentes ángulos y detalles (llegando a fragmentos y particularidades que se pueden captar como elementos abstractos y/o texturados). Esta complejidad de determinación afecta directamente a la fase de entrenamiento dado que no siempre es evidente la asignación de una imagen a una categoría concreta, ni es simple establecer los límites que excluyen a unas categorías de las otras.

Pero también precisamente por esta razón, el sistema se convierte en una herramienta de utilidad para el análisis del conjunto de obras, tanto por parte del artista creador como del estudioso del arte, al proporcionarle nuevos puntos de vista sobre lo observado, alejados de la experiencia conocida y condicionada por la propia percepción individual.

3.3 EXPERIMENTO DE APRENDIZAJE NO SUPERVISADO GENERATIVO

Los resultados obtenidos en el primer experimento de aprendizaje supervisado sobre la muestra de imágenes estudiada conducen al planteamiento de que, dada la dificultad de categorización previa de algunas tipologías de las imágenes, incluso por expertos en arte, cabría esperar que la clasificación totalmente automática fuese más efectiva, e incluso que fuese capaz de poner al descubierto relaciones novedosas e incluso sorprendentes. Al tratarse de imágenes que no se corresponden con temas clásicos tales como paisajes o bodegones, el conocimiento previo o la experiencia personal afecta notablemente a la hora de decidir si una imagen pertenece a una clase o a otra, lo cual conduce al establecimiento de unas fronteras difusas entre algunas de las clases utilizadas para el entrenamiento del sistema.

Se plantea así esta segunda experiencia sobre el total de la muestra inicial de Planas compuesta por un total de 2846 imágenes. También en esta ocasión el tamaño de las imágenes está comprendido entre 480 x 480 píxeles y 1400 x 1400 píxeles, y el sistema re-escala a

480 x 480 píxeles las imágenes que superan este tamaño.

Como se detalla en la Fig. 3.2, el proceso de esta segunda experiencia consiste también en construir el vocabulario de 300 palabras visuales mediante el modelo *Bag-of-Words* y descriptores *PHOW*, pero en esta ocasión a partir del total de las imágenes, para después computar la representación *PHOW* obtenida a través del modelo *pLSA* (Análisis Probabilístico de Aspectos Latentes*)* (Más detalles en el apartado 4 del Anexo A: Representación de aspectos latentes mediante *pLSA*). De esta forma finalmente se obtiene una distribución de aspectos latentes para cada imagen (Fig. 3.14). Por último, se pueden establecer un orden y clasificar las imágenes en función de la probabilidad de los aspectos que contienen. Con este sistema también es posible, dada una imagen problema, determinar a que categoría *pLSA* pertenece (Fig. 3.2).

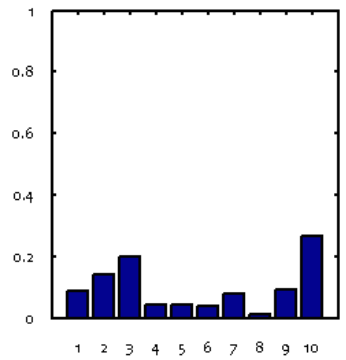


Figura 3.14. Representación de una imagen mediante el método *pLSA* basada en su distribución de aspectos, Se muestran los aspectos numerados del 1 al 10 en el eje de abscisas y en el eje de ordenadas se indica la probabilidad asociada a cada aspecto para esta imagen.

Con el total de imágenes se realizan diversas pruebas consistentes en aumentar o disminuir el tamaño del vocabulario, así como en modificar el número de aspectos latentes con el que trabajar en el *pLSA* (20,15,10). Finalmente se decide trabajar con 300 palabras visuales porque otros autores han puesto de manifiesto un buen rendimiento con este número de palabras visuales en la literatura (Bosch, 2007) y también se corresponden con resultados satisfactorios en nuestro caso. Las clasificaciones obtenidas para 5 aspectos mostraban que algunas tipologías no resultaban visibles, y con 20 y 15 aspectos las categorías quedaban poco representadas y algo dispersas. Por lo tanto, se decidió que la prueba más representativa para el total de imágenes analizadas era la que clasificaba la muestra en base a 10 aspectos latentes.

Para la evaluación del resultado de esta agrupación se consideran prioritariamente imágenes tipificadas en un determinado aspecto con una probabilidad igual o superior a 0.6. La forma en que se calcula dicha probabilidad está detallada en el apartado 4 del Anexo A. A pesar de esta consideración, en la discusión de cada aspecto se muestran las 16 imágenes con mayor probabilidad de estar asociadas a ese aspecto porque se considera que así el lector puede hacerse más fácilmente a la idea de la consistencia visual de dicho aspecto, aunque a veces haya imágenes por debajo del umbral de probabilidad de 0.6. Los casos que tienen alguna característica conflictiva o dudosa se comentan de forma más concreta

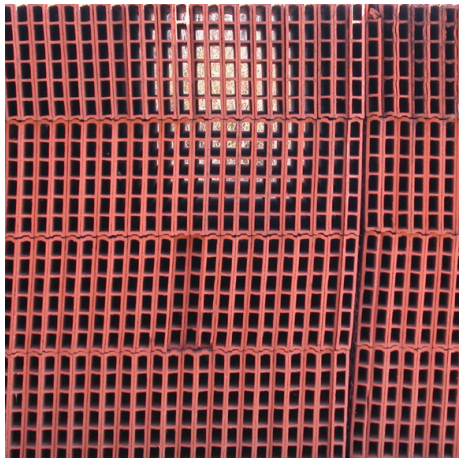


Figura 3.15. Imagen poco entrópica.

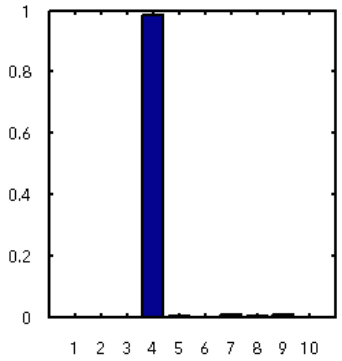


Figura 3.16. Histograma de imagen poco entrópica. En estas imágenes se observa que la probabilidad se concentra mayoritariamente en un aspecto.

para facilitar la comprensión.

Con la intención de que algunos de los resultados sean más comprensibles para el lector, recordamos que la evaluación computacional en este experimento se efectúa utilizando descriptores que trabajan con escala de grises.

Ha resultado complicado asignar un descriptor textual a los aspectos hallados por el sistema, ya que no siempre es fácil y directo asociar el contenido visual del conjunto de imágenes de un aspecto con una descripción literal. Los aspectos hallados por la máquina no sólo deben entenderse compositivamente sino que se debe tener presente que se basan en encontrar co-ocurrencias de palabras visuales. A modo de ejemplo aclaratorio, este método sería capaz de detectar y agrupar en el mismo aspecto las imágenes que contengan un rostro, por la co-ocurrencia de las palabras visuales ojos, nariz, boca, etc. De la misma manera, agruparía en una misma clase imágenes de caballos aunque fuesen de distinto color y estuviesen en distintas posiciones.

La metodología *pLSA* proporciona una representación de la imagen en base a la distribución de probabilidad de los aspectos que contiene. (Fig. 3.14) A la vista de los resultados obtenidos en un primer análisis considerando 10 aspectos, percibimos que el total de la muestra consta de dos tipologías de imágenes muy marcadas:

1- un tipo de fotografías que presenta un único aspecto muy destacado. Las llamaremos imágenes de *poca entropía*. (Fig 3.15 y 3.16)

2- y otro que presenta varios aspectos asociados simultáneamente. Las llamaremos imágenes de *más entropía*. (Fig 3.17 y 3.18)

Para poder distinguir y tratar separadamente estas dos tipologías de imágenes se ha utilizado el índice de entropía de Shannon (Cover & Thomas, 2006). (Para ampliar detalles sobre la manera de calcular este índice se puede consultar el apartado 8 del Anexo A: Índice de entropía de Shannon).

Para una imagen dada *d* tenemos un vector de probabilidades, tantas probabilidades como aspectos hayamos decidido. Podemos calcular el índice de Entropía de Shannon de la imagen *d*; por ejemplo, para una imagen que esté asociada a un único aspecto, es decir, una imagen con un vector de probabilidades que contenga valor 0 en todas las posiciones excepto en una posición que tenga un valor de 1, su índice de entropía será mínimo e igual a H(*d*)=0. Contrariamente, a una imagen que esté asociada por igual a todos los aspectos, es decir, con vector de probabilidades con valor 1/10 en cada componente, le corresponderá un índice de entropía máximo e igual a H(*d*)=2.3026. Los rangos de entropía teóricos respecto a 10 aspectos irían de 0 a 2.3026.

Calculamos los índices de Shannon para la muestra de nuestro estudio y observamos que los valores se encuentran entre 0 y 2,17. Las imágenes que tienen una entropía elevada son aquellas que el procedimiento ha asociado de manera equiprobable a cada uno de los aspectos.

De esta forma se decide seleccionar del total de la muestra las imágenes con un valor de entropía superior a 1,4 y repetir de nuevo la búsqueda de aspectos únicamente con en este nuevo conjunto formado por 1.482 imágenes. Se repite de nuevo todo el proceso generando los descriptores locales, el vocabulario visual y se intenta así que el sistema sea capaz de establecer nuevas relaciones entre imágenes visualmente más complejas dando lugar a nuevos aspectos latentes distintos de las 10 primeros. La prueba resulta un éxito y se generan otro conjunto distinto de 10 aspectos sobre la nueva muestra. En total el sistema es capaz de categorizar en 20 grupos el total de imágenes analizadas y estos son los resultados que pasaremos a discutir en el resto del capítulo.

A continuación se muestra una selección de las categorías de imágenes menos entrópicas i más entrópicas acompañadas por los respectivos histogramas de aspectos, numerados del 1 al 10 en el eje de abscisas y en el eje de ordenadas se indica la probabilidad asociada a cada aspecto (Fig. 3.19 y 3.20).

En el desarrollo del capítulo de resultados se utilizan histogramas como herramientas de descripción de las características formales de las imágenes analizadas.

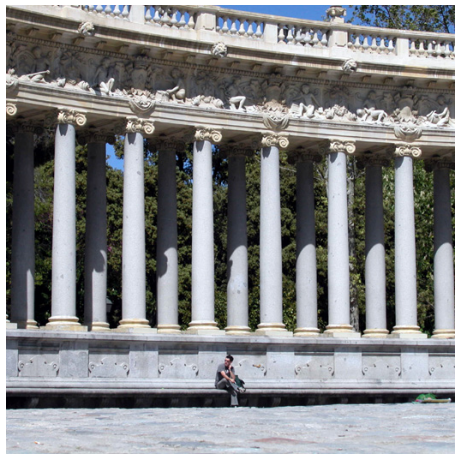


Figura 3.17. Imagen más entrópica.

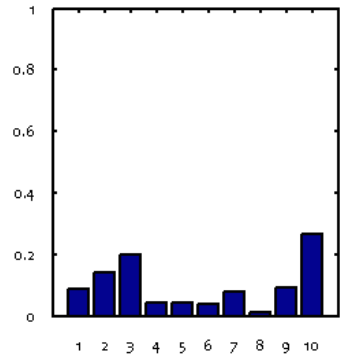


Figura 3.18. Histograma de imagen más entrópica. En estas imágenes la distribución de probabilidad se reparte entre diversos aspectos.

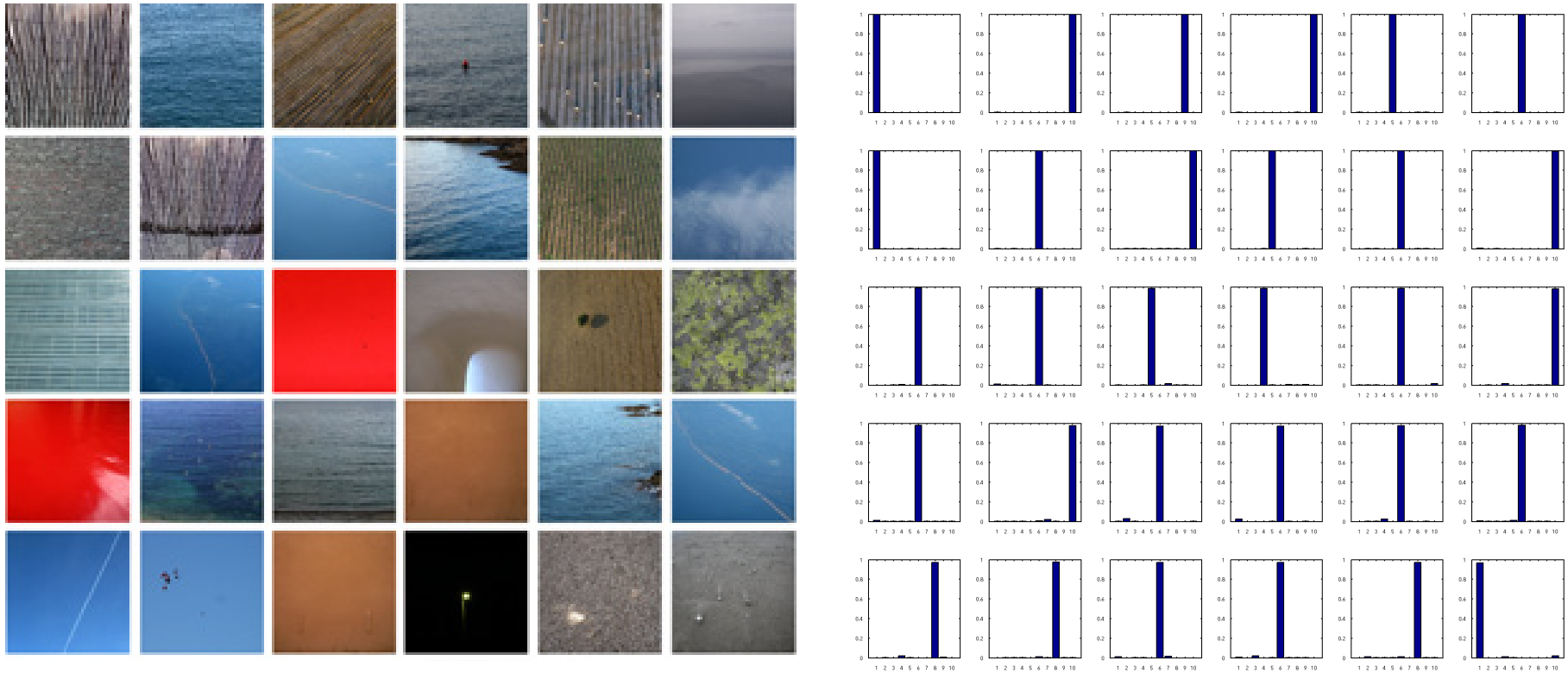


Figura 3.19. A la izquierda presentamos una muestra de las imágenes menos entrópicas según el índice de Shannon pertenecientes a la colección de Planas, y a su derecha una muestra de sus histogramas según la representación por distribución de aspectos resultante de la aplicación del modelo *pLSA*.

Figura 3.20. A la derecha presentamos una muestra de las imágenes más entrópicas según el índice de Shannon pertenecientes a la colección de Planas, y a su izquierda una muestra de sus histogramas según la representación por distribución de aspectos resultante de la aplicación del modelo *pLSA*.

A partir de este momento vamos a describir en detalle;

1- Los 10 aspectos resultantes del análisis con *pLSA* de la muestra de 1.482 imágenes de la colección de Planas, seleccionada así por presentar estas fotografías un índice de Shannon inferior a 1.4. A estos aspectos les asignaremos una descripción textual precedida por las siglas PE que se refieren a aspectos *Poco Entrópicos*.

2- Los 10 aspectos que son el resultado de analizar el segundo grupo de 1364 imágenes de la misma colección que presentan un índice se Shannon superior a 1.4. A cada aspecto de este segundo grupo de le antepondremos las siglas ME referidas a *Más Entrópicos* para que en todo momento resulte fácil diferenciar unos aspectos de otros.

En los conjunto de 16 imágenes que se muestran para explicar las características de cada aspecto latente, las fotografías están ordenadas de mayor a menor probabilidad de tener el aspecto comentado (de izquierda a derecha y de arriba a abajo). Así, la imagen con mayor probabilidad de tener el aspecto concreto es la primera de arriba a la izquierda y la menos probable es la número 16 que quedará abajo a la derecha.

3.3.1 Aspectos Latentes del conjunto de imágenes poco entrópicas de Planas

Figura 3.21. Conjunto de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 1: Lineas Finas Definidas.

3.3.1.1 Aspecto PE1: Lineas Finas Definidas

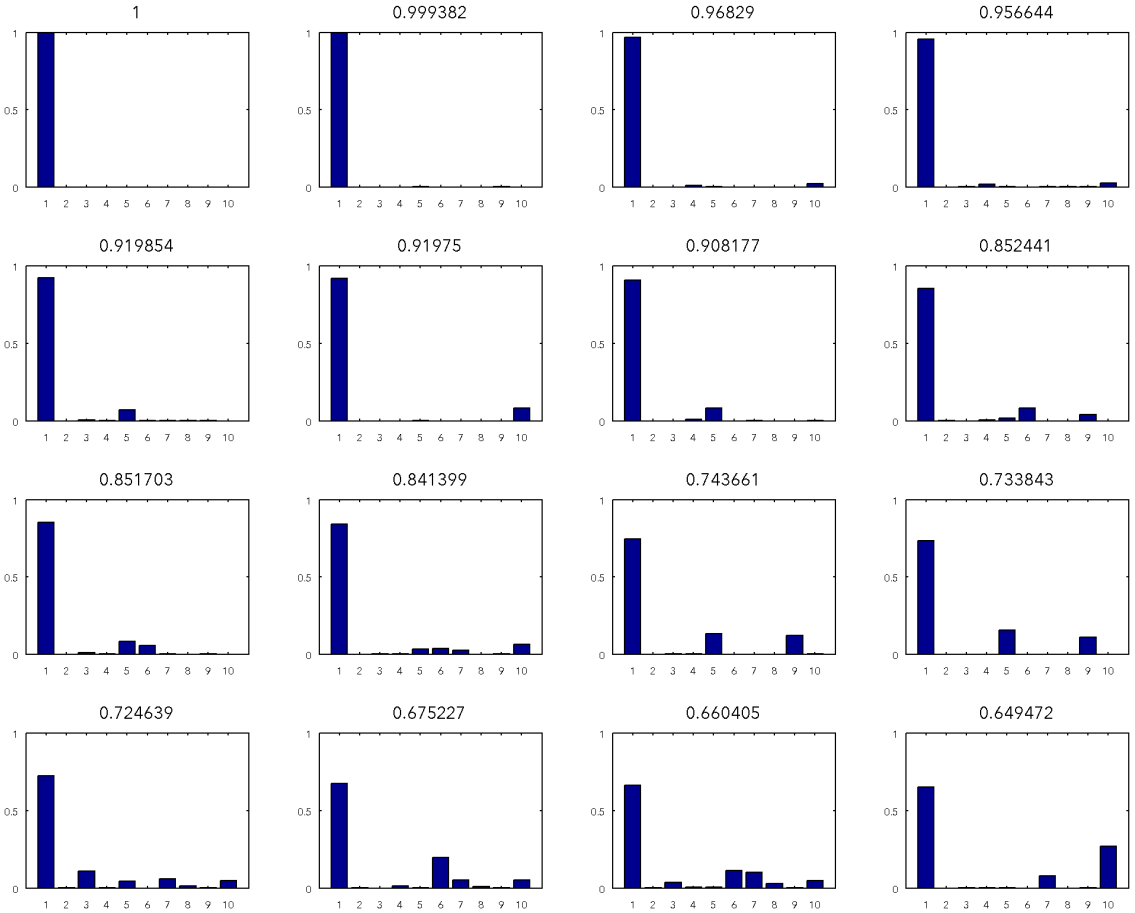


Figura 3.22. Conjunto de histogramas de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 1: Lineas Finas Definidas.

El aspecto Líneas Finas Definidas agrupa un conjunto de imágenes que contienen trazados lineales en las que predomina una dirección, que puede ser horizontal, vertical o ambos (Fig. 3.21 y 3.22).

Para clarificar un poco este aspecto se ha girado la tercera imagen como sigue y se ha vuelto a comprobar el aspecto mayoritario. El resultado se muestra en la Fig. 3.23:

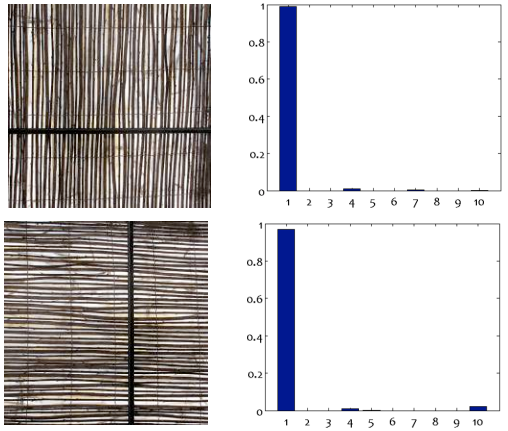


Figura 3.23. Si giramos la imagen y calculamos de nuevo el aspecto mayoritario observamos que sigue siendo el 1, por lo tanto se pude concluir que la verticalidad no es determinante para este aspecto.

Figura 3.24. Conjunto de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 2: Diagonal Descendente.

3.3.1.2 Aspecto PE2: Diagonal Descendente

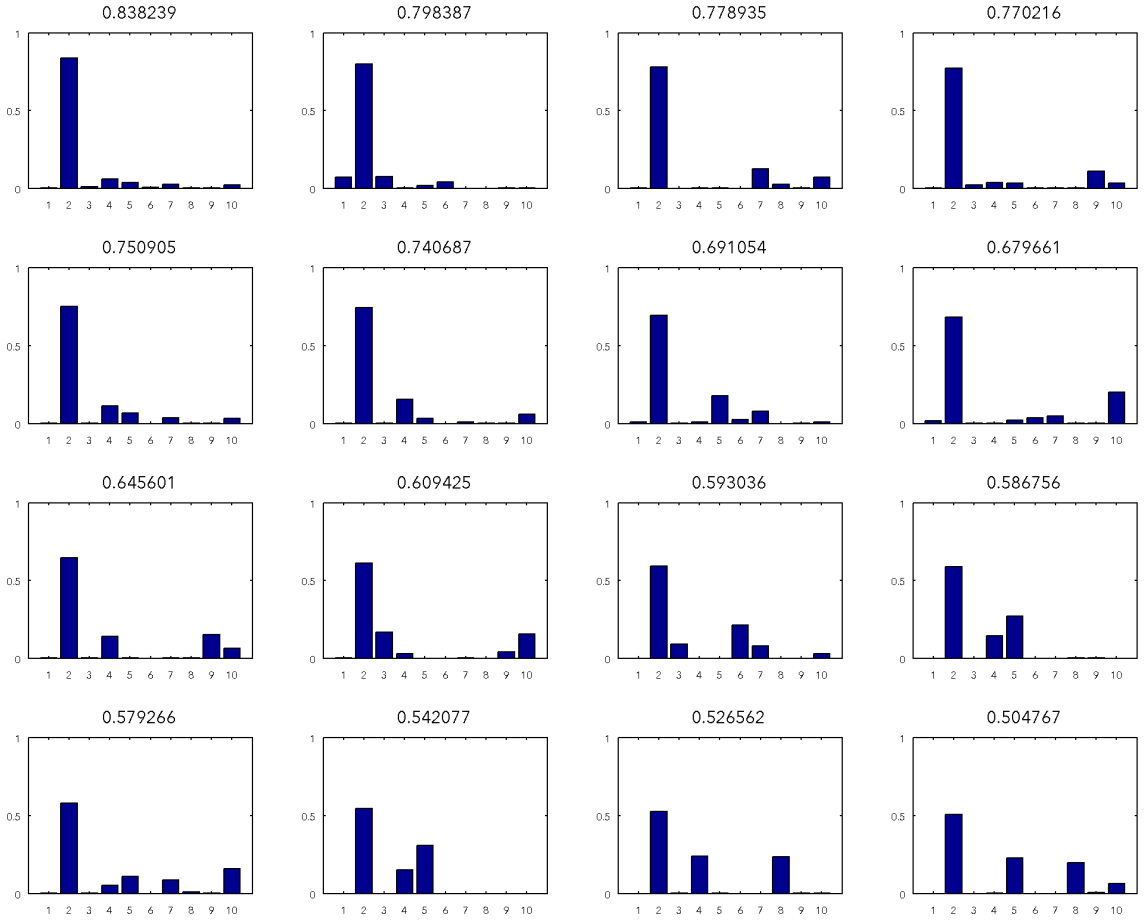


Figura 3.25. Conjunto de histogramas de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 2: Diagonal Descendente.

El aspecto Diagonal Descendente agrupa las imágenes en las que predomina básicamente una direccionalidad de bajada desde el ángulo superior izquierdo al ángulo inferior derecho (Fig. 3.24 y 3.25).

A partir de la probabilidad 0.6 se observa que la característica definitoria de este aspecto sólo se presenta en un fragmento de la imagen.

También observamos en la primera imagen componentes del aspecto- PE4 Textura Heterogénea, que se corresponderían con la zona superior de la imagen. En la segunda imagen vemos una componente perteneciente al aspecto PE1 Líneas Finas Definidas y en la tercera imagen componentes de los aspectos PE7 Horizontal Estrecha y aspecto PE10 Horizontal Vibrante.

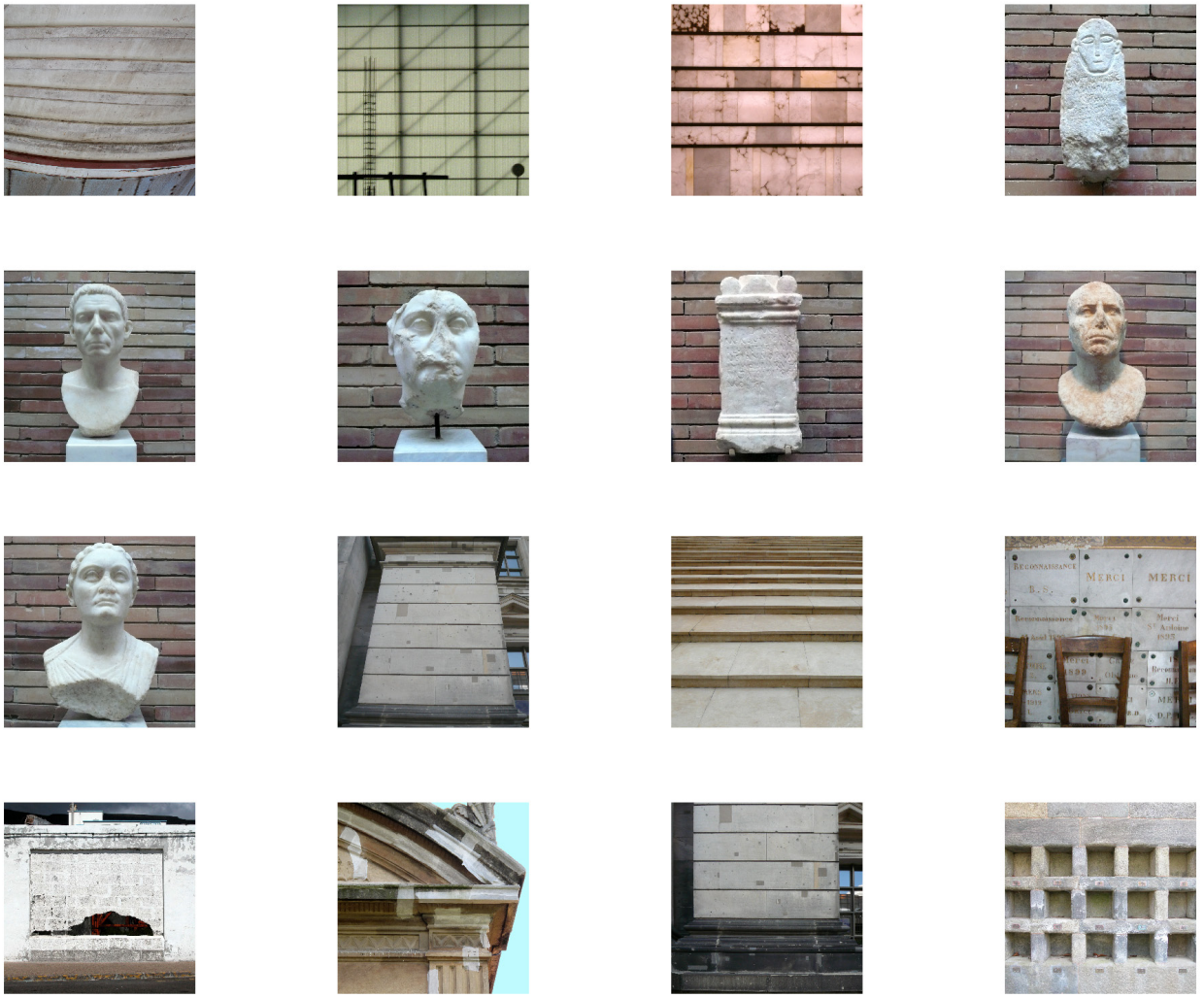


Figura 3.26. Conjunto de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 3: Horizontal Amplia.

3.3.1.3 Aspecto PE3: Horizontal Amplia

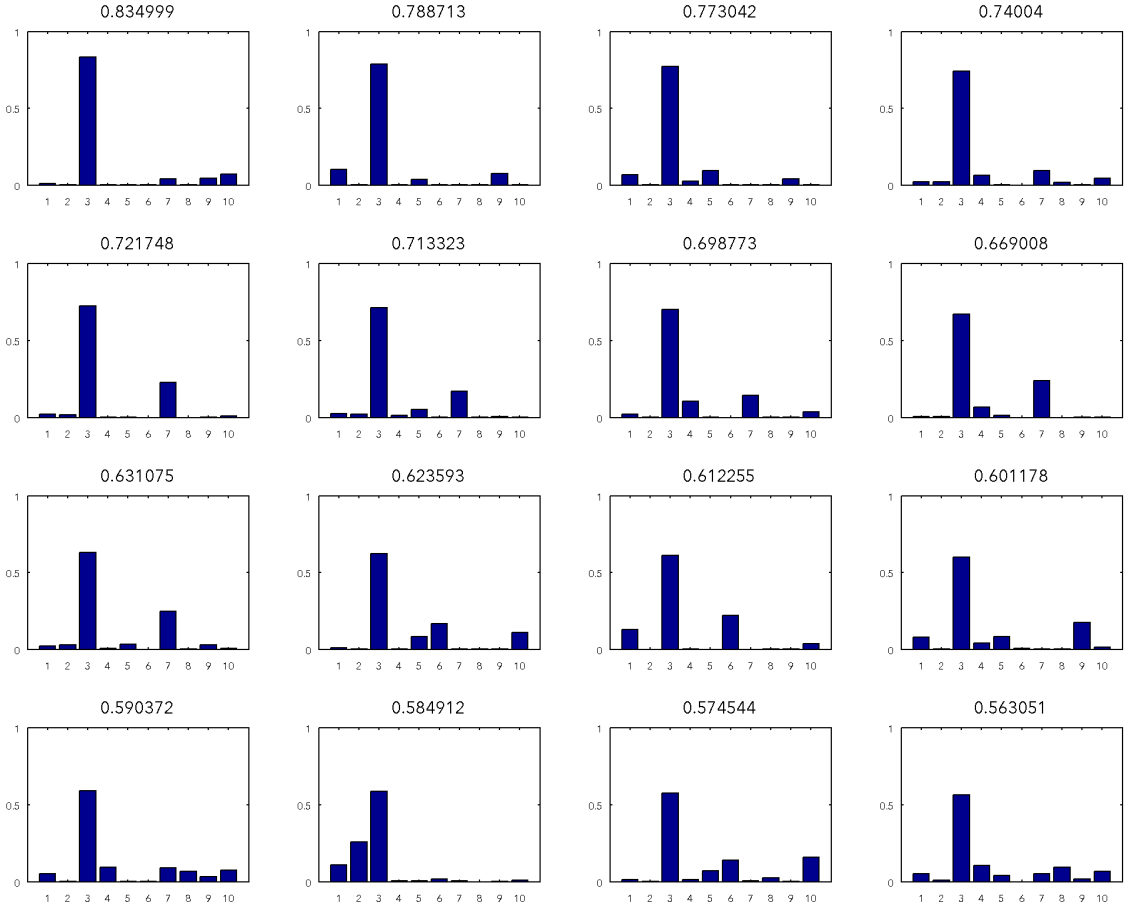


Figura 3.27. Conjunto de histogramas de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 3: Horizontal Amplia.

El aspecto Horizontal Amplia reúne un conjunto de imágenes caracterizadas por presentar una bandas horizontales anchas muy marcadas (Fig. 3.26 y 3.27).

En la Fig. 3.28 se puede destacar el componente de aspecto PE10 Horizontal Vibrante de la primera imagen y el aspecto PE1 Líneas Definidas de la tercera imagen. Detallamos en concreto las probabilidades de la segunda imagen en la Fig. 3.28: un 0.1 del aspecto PE1 Líneas Finas Definidas, un 0.07 del aspecto PE9 Diagonal Ascendente y un 0.036 del aspecto PE5 Vertical Irregular Texturada.

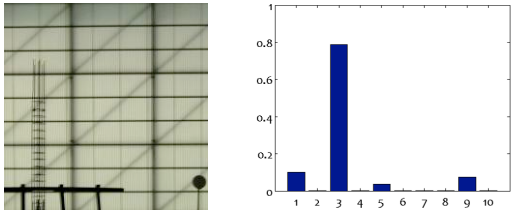


Figura 3.28. Segunda imagen e histograma.

De la cuarta a la sexta imagen observamos una escultura central que no debemos tener en cuenta para valorar el aspecto, ya que el proceso, en este caso, nos muestra el aspecto más probable que es el fondo. A partir de la imagen 12 el aspecto se diluye.

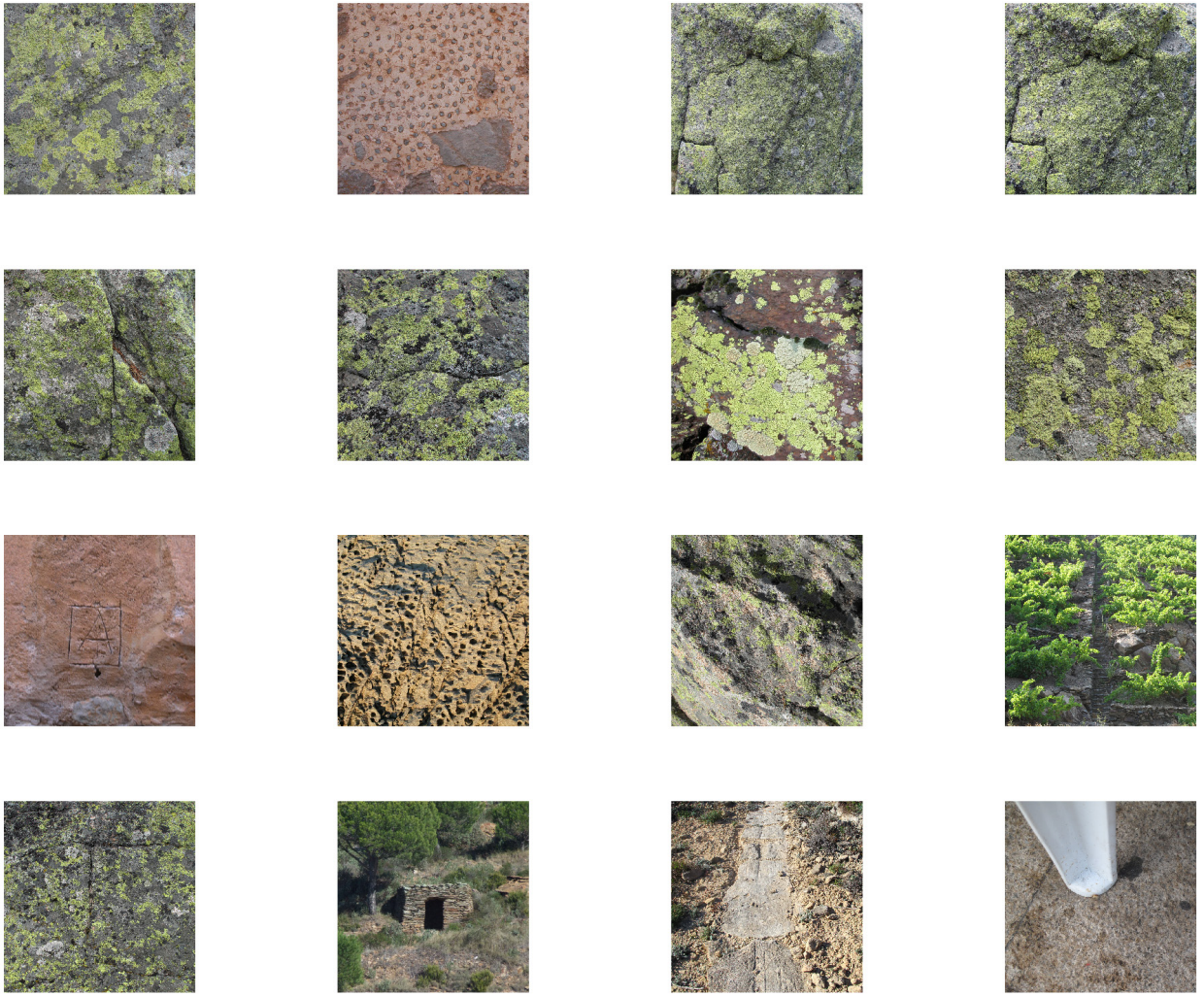


Figura 3.29. Conjunto de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 4: Textura Heterogénea.

3.3.1.4 Aspecto PE4: Textura Heterogénea

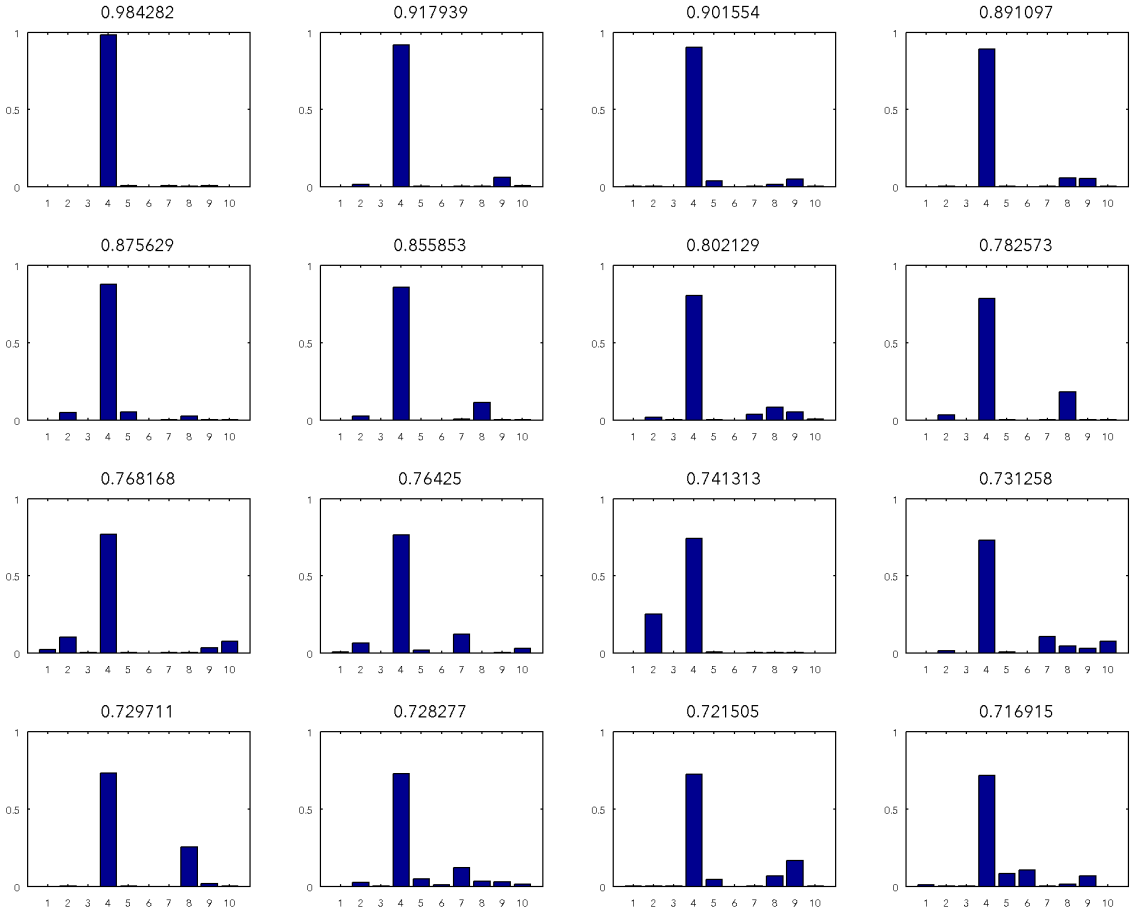


Figura 3.30. Conjunto de histogramas de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 4: Textura Heterogénea.

El aspecto Textura Heterogénea presenta una agrupación bastante sólida, sobre todo teniendo en cuenta la percepción en escala de grises de la computadora (Fig. 3.29 y 3.30).

Todas las imágenes mostradas tienen una probabilidad superior al 0,7.

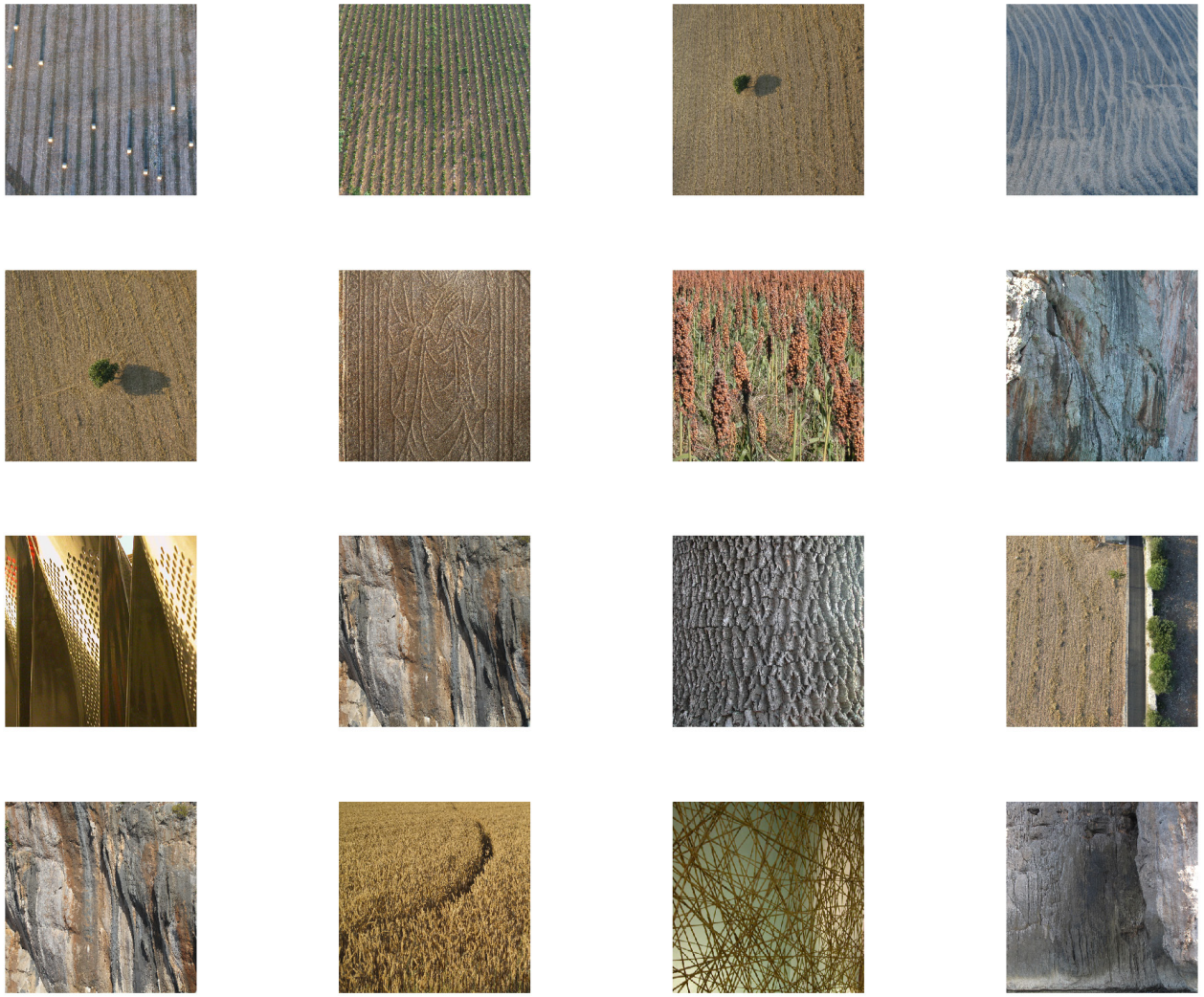
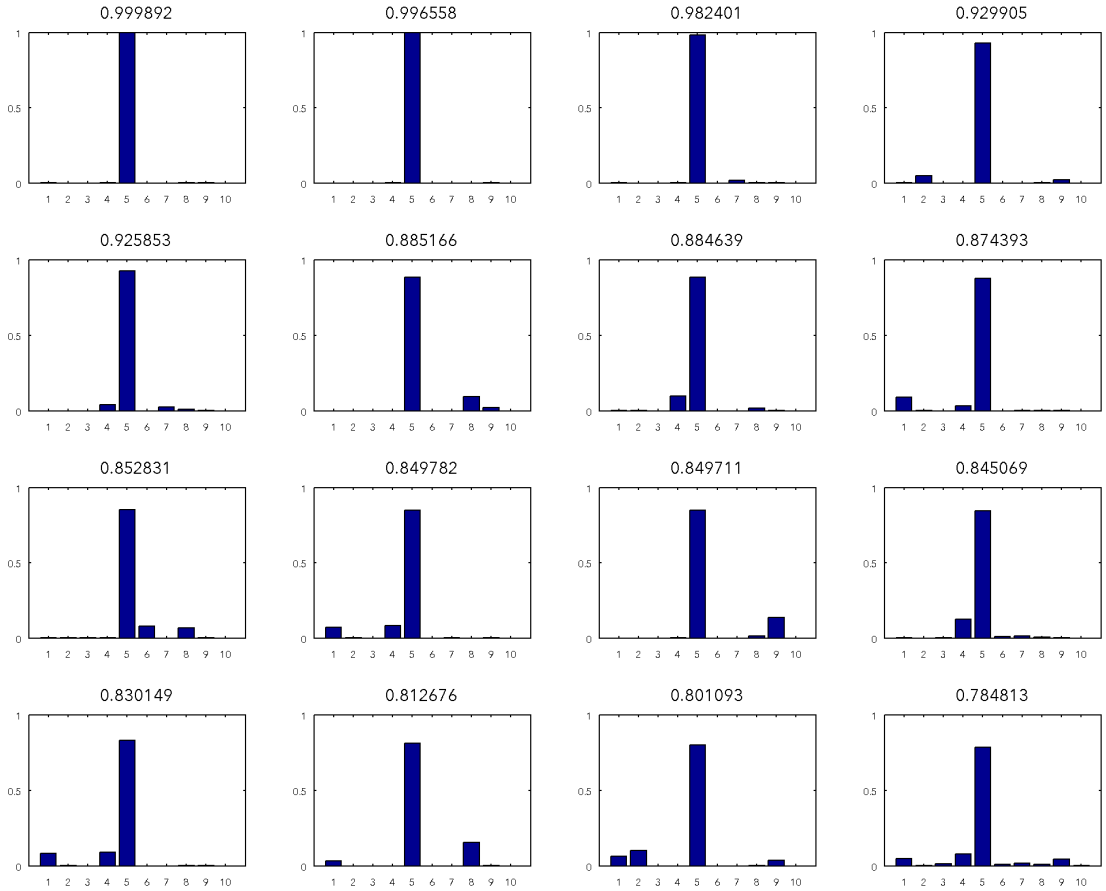


Figura 3.31. Conjunto de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 5: Vertical Irregular Texturada.

3.3.1.5 Aspecto PE5: Vertical Irregular Texturada



El aspecto Vertical Irregular Texturada es un aspecto muy compacto y fácilmente identificable. Las 16 imágenes tienen un índice de probabilidad superior al 0.7. (Fig. 3.31 y 3.32)

Figura 3.32. Conjunto de histogramas de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 5: Vertical Irregular Texturada.

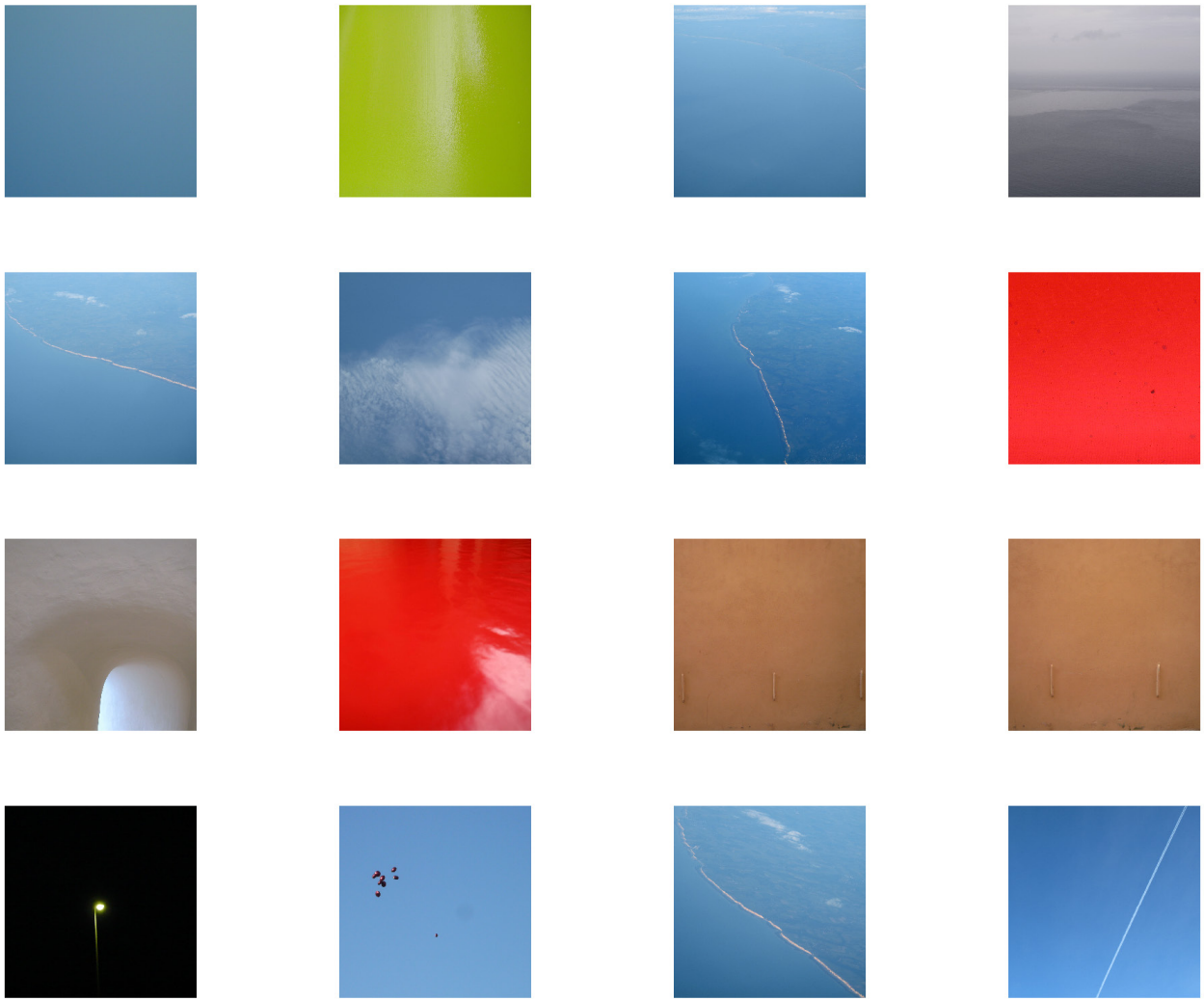


Figura 3.33. Conjunto de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 6: Liso.

3.3.1.6 Aspecto PE6: Liso

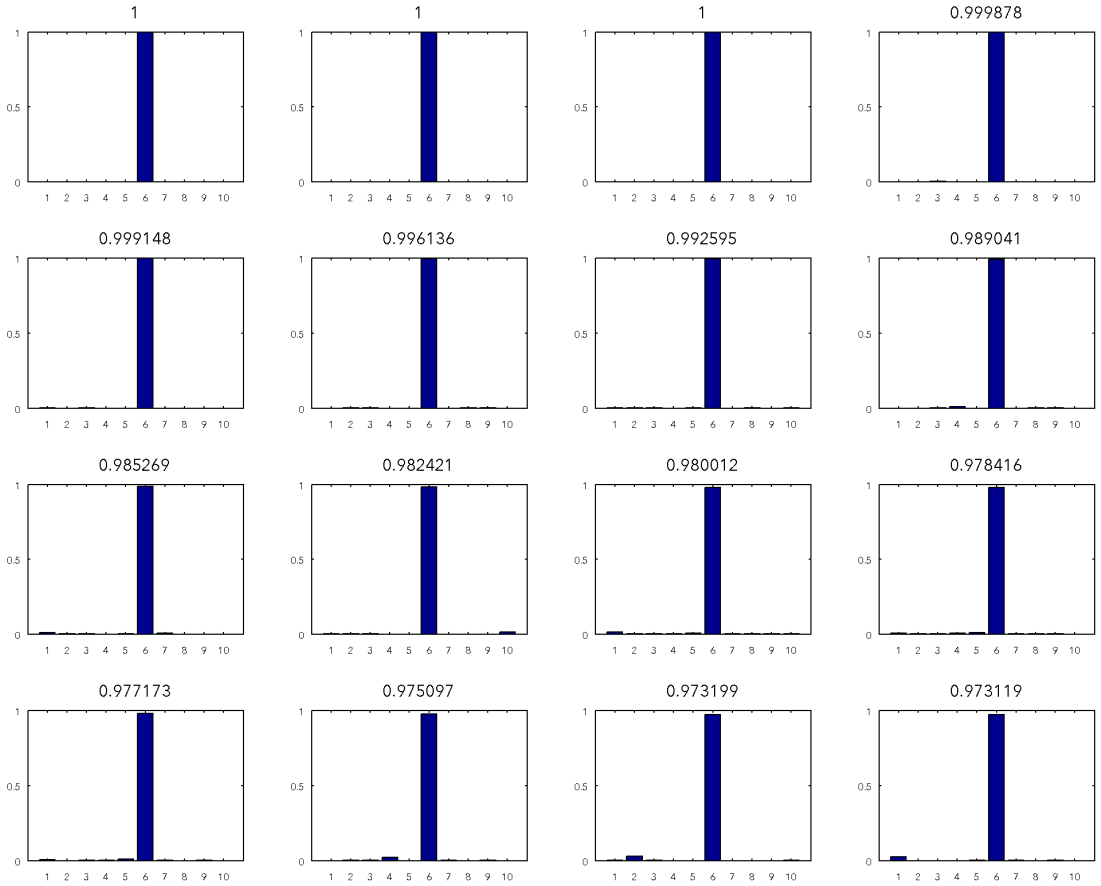


Figura 3.34. Conjunto de histogramas de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 6: Liso.

Aspecto Liso, carente de textura. Para valorar este aspecto en su justa medida es necesario recordar que la máquina sólo considera la escala de grises. (Fig. 3.33. y 3.34)

Aspecto indiscutible en el que las 3 primeras imágenes presentan una probabilidad de 1, o sea, que sólo tienen este aspecto, y el resto hasta las 16 tienen probabilidades superiores a 0.97.

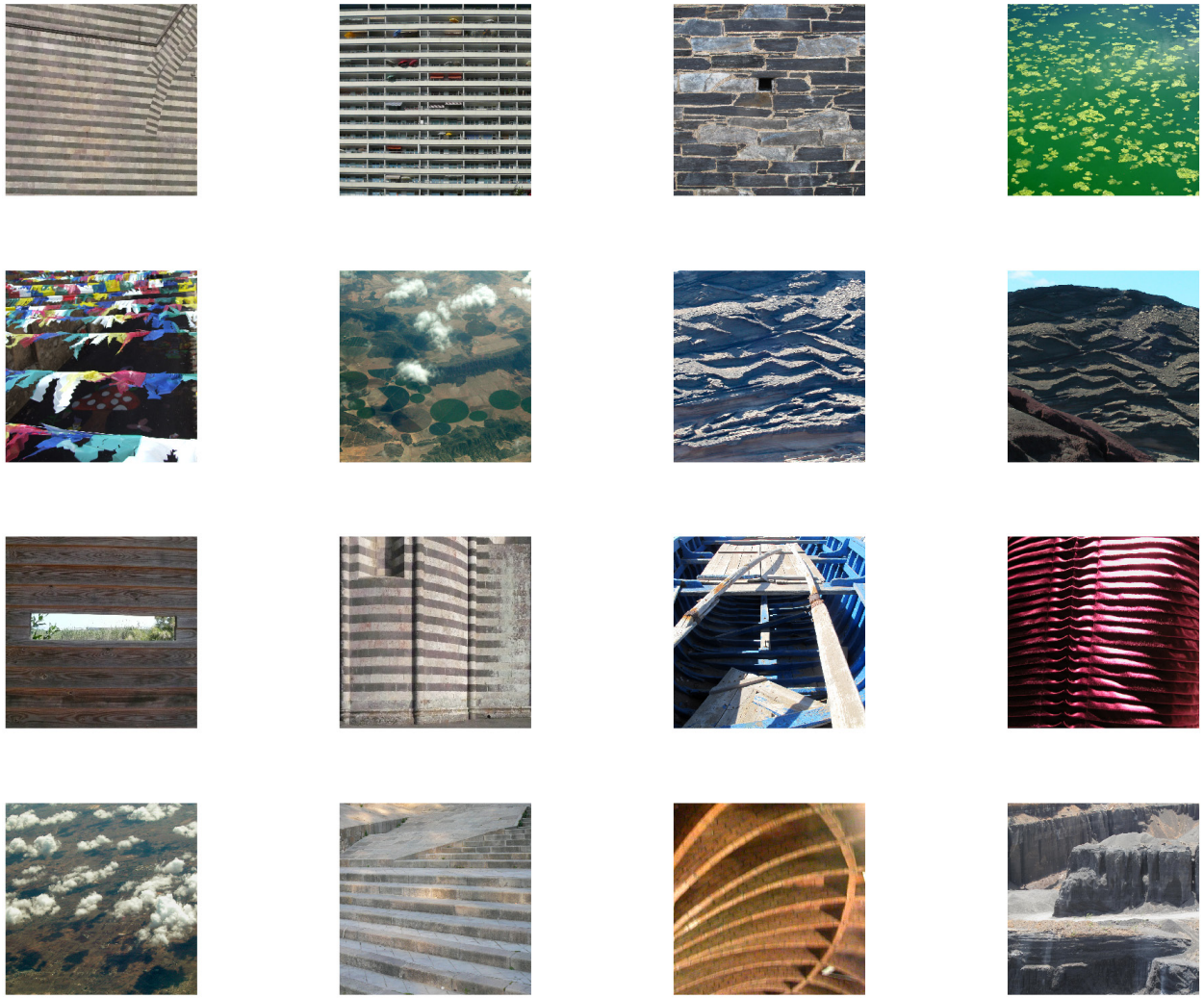


Figura 3.35. Conjunto de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 7: Horizontal Estrecha.

3.3.1.7 Aspecto PE7: Horizontal Estrecha

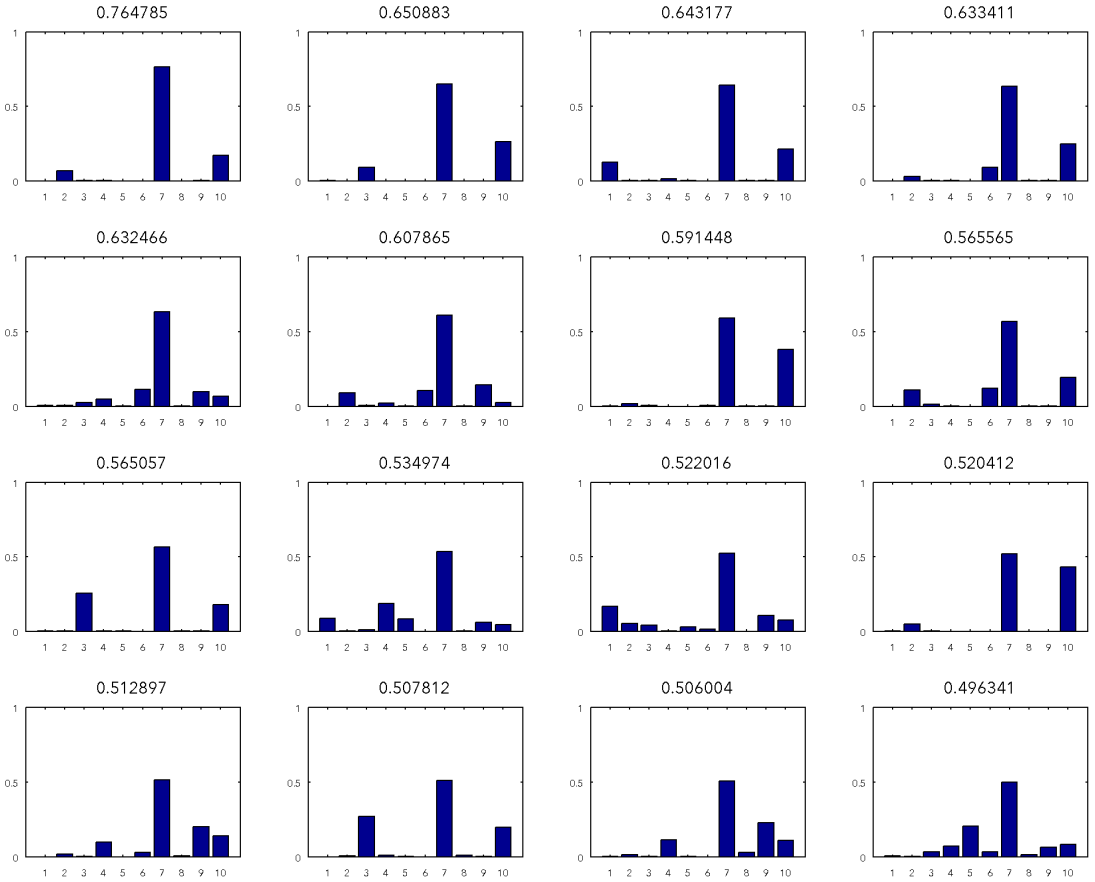


Figura 3.36. Conjunto de histogramas de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 7: Horizontal Estrecha.

El aspecto Horizontal Estrecha presenta unas bandas horizontales más comprimidas que en el aspecto anterior. (Fig. 3.35 y 3.36)

Disponemos de la función HOG (*histogram of oriented gradients*) que descompone una imagen en pequeñas celdas cuadradas y calcula un histograma de orientación en cada celda. La utilizamos para generar una versión pictórica de las características de las imágenes y así poder tener una mejor comprensión visual de los gradientes que son base del proceso computacional. Así podemos clarificar la tercera imagen, comprobando un marcado aspecto horizontal en los gradientes. (Fig. 3.37)

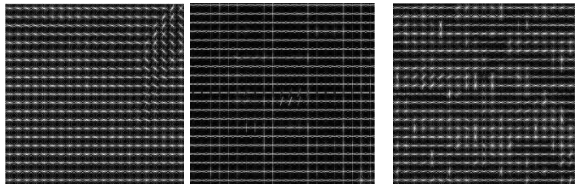


Figura 3.37. HOG de las imágenes 1, 2 y 3.

Es destacable la importante componente de aspecto PE10 Horizontal Vibrante que presentan muchas de las imágenes y que no se produce en las imágenes del aspecto PE3 Horizontal Amplia.

Figura 3.38. Conjunto de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 8: Textura Homogénea.

3.3.1.8 Aspecto PE8: Textura Homogénea

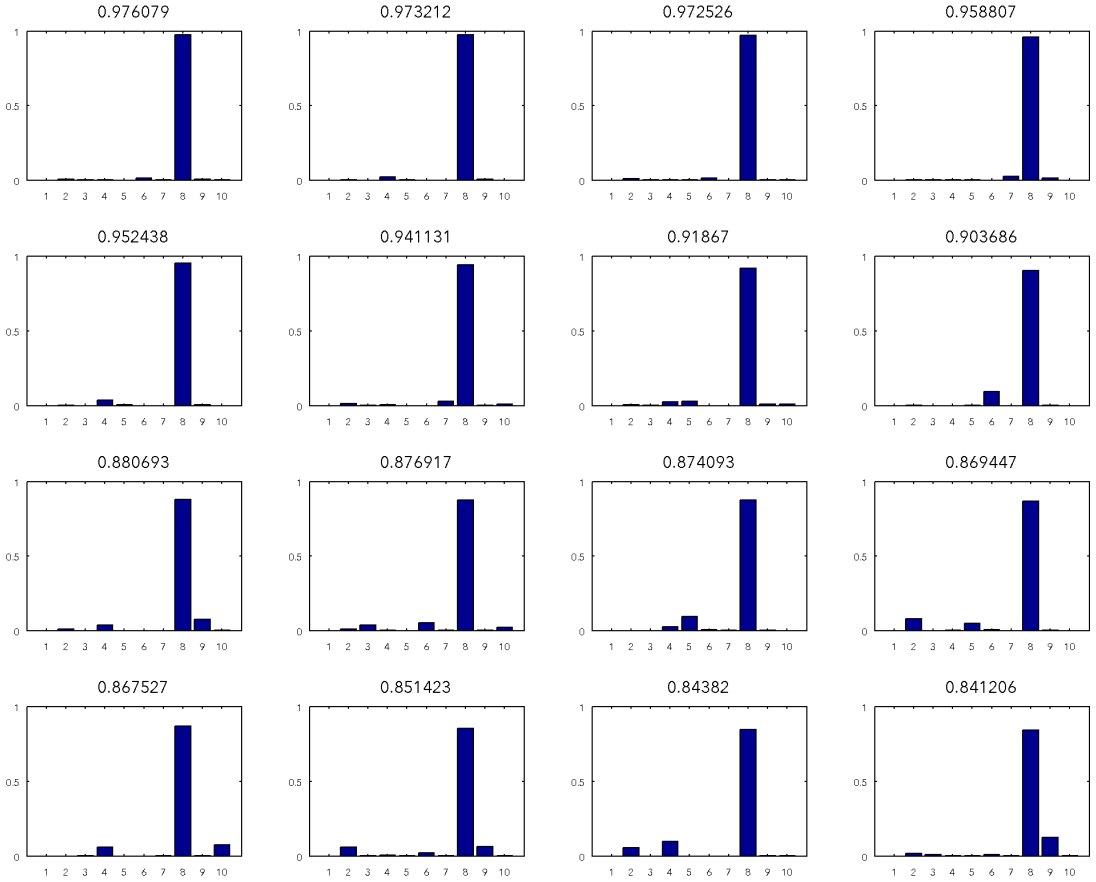


Figura 3.39. Conjunto de histogramas de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 8: Textura Homogénea.

El aspecto Textura Homogénea es muy similar al anterior en cuanto a que es compacto y muy bien definido, aunque es ligeramente rugoso. (Fig. 3.38 y 3.39)

Observamos pequeños elementos que son los responsables de la disminución de la probabilidad.

Figura 3.40. Conjunto de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 9: Diagonal Ascendente.

3.3.1.9 Aspecto PE9: Diagonal Ascendente

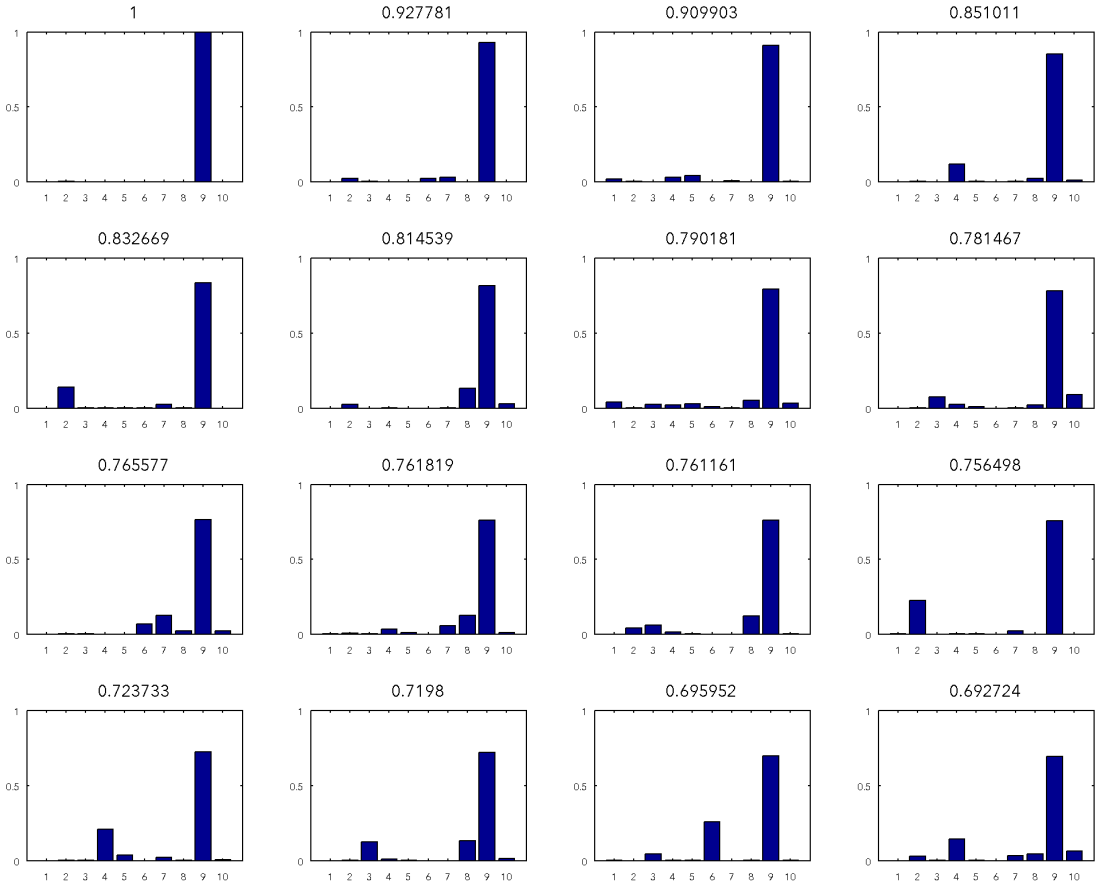


Figura 3.41. Conjunto de histogramas de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 9: Diagonal Ascendente.

El aspecto Diagonal Ascendente agrupa las imágenes en las que predomina básicamente una direccionalidad de subida desde el ángulo inferior izquierdo al ángulo superior derecho (Fig. 3.40 y 3.41).

Es interesante destacar el modo en que la probabilidad disminuye a la vez que la diagonal es menos perfecta y cómo aumentan en la tercera imagen el aspecto PE4 Textura Heterogénea que correspondería a las zonas superior e inferior de la imagen, y el aumento también de la probabilidad del aspecto PE5 Vertical Irregular Texturada que pertenece a la parte derecha de la imagen en donde la diagonal ya es prácticamente una vertical.

Figura 3.42. Conjunto de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 10: Horizontal Vibrante.

3.3.1.10 Aspecto PE10: Horizontal Vibrante

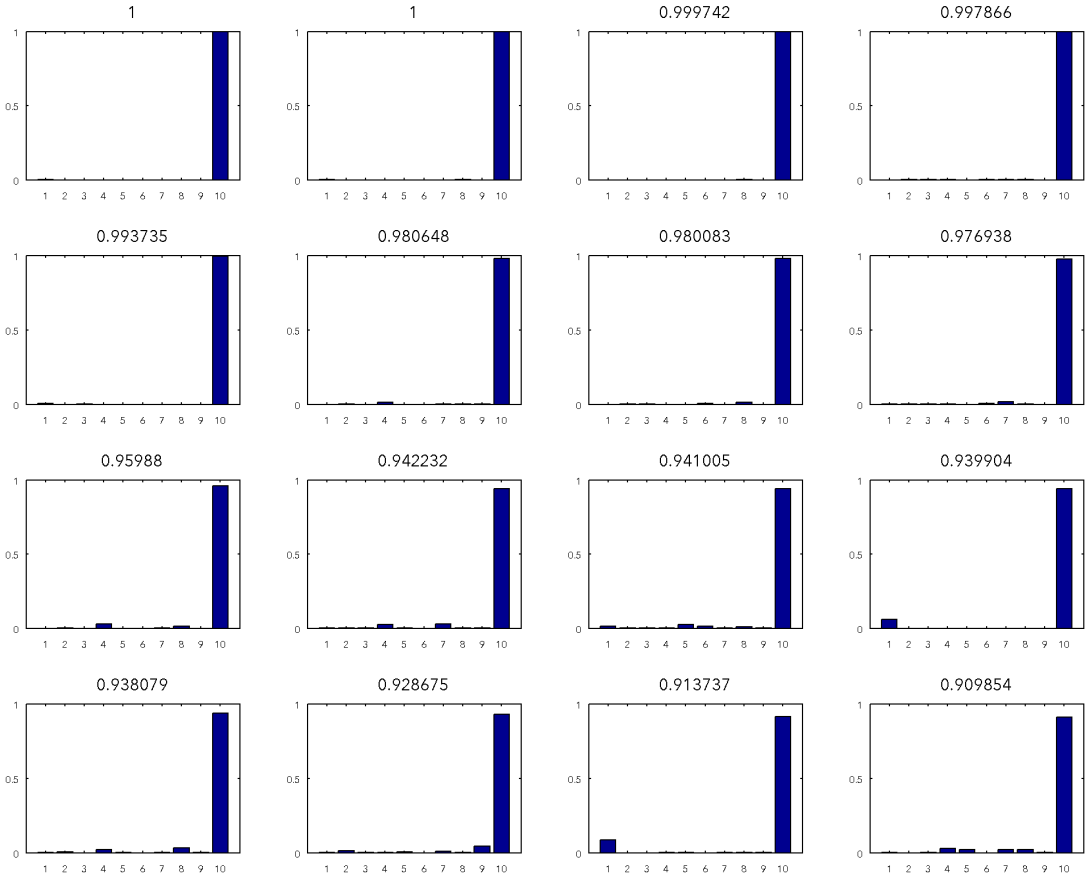


Figura 3.43. Conjunto de histogramas de las 16 imágenes poco entrópicas que presentan con mayor probabilidad el aspecto 10: Horizontal Vibrante.

Este aspecto Horizontal Vibrante presenta una textura de marcada horizontalidad de aspecto más bien rugoso y vibrante. Resulta bastante claro el conjunto de las imágenes de este aspecto, que además presentan todas probabilidades superiores a 0.9, concretamente las 2 primeras imágenes se corresponden de manera exclusiva con este aspecto, de ahí su probabilidad de 1. (Fig. 3.42 y 3.43)

3.3.2 Aspectos Latentes del conjunto de imágenes más entrópicas de Planas

La complejidad de esta categoría de imágenes es mucho mayor que la de imágenes poco entrópicas y el análisis de los resultados es complicado dado que ya no predomina de forma muy marcada un solo aspecto, sino que las imágenes están constituidas por multitud de ellos. A partir de ahora los aspectos pertenecientes a esta categoría de imágenes más entrópicas los nombraremos abreviados como ME (Más Entrópicas).

3.3.2.1 Aspecto ME1: Figura-Fondo

Denominamos aspecto Figura-Fondo al conjunto de imágenes en las que predomina una figura central en primer plano sobre un fondo bastante homogéneo. A pesar de tener probabilidades moderadas de 0,6, es un aspecto difícil de interpretar. (Fig. 3.44)

Figura 3.44. Conjunto de las 16 imágenes más entrópicas que presentan con mayor probabilidad el aspecto 1: Figura-Fondo.

3.3.2.2 Aspecto ME2: Textura Homogénea Bipolar

Figura 3.45. Conjunto de las 16 imágenes más entrópicas que presentan con mayor probabilidad el aspecto 2: Textura Homogénea Bipolar.

Este aspecto pasado a escala de grises presenta dos tonalidades marcadas que dividen la imagen en dos secciones o polos tonales y/o estructurales de composición. (Fig. 3.45)

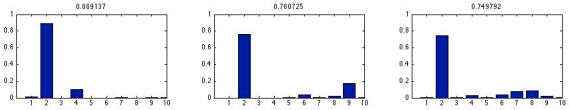


Figura 3.46. Conjunto de los 3 histogramas de las 3 primeras imágenes del aspecto 2: Textura Homogénea Bipolar.

En la Fig. 3.46 se muestran los histogramas de las 3 primeras imágenes y es informativo percibir, además del aspecto ME2 que se está tratando, en la primera imagen la componente del aspecto ME4 Imagen Bipartita, en la segunda imagen el aspecto ME9 Estructura Heterogénea Contrastada y en la tercera imagen la componente del aspecto ME8 Planos Interseccionados.

3.3.2.3 Aspecto ME3: Estructuras Verticales

En el aspecto Estructuras Verticales se notan claramente unas franjas o columnas verticales en algún caso con una pequeña inclinación. Especialmente en las primeras imágenes se percibe un cierto predominio central en la composición. (Fig. 3.47)

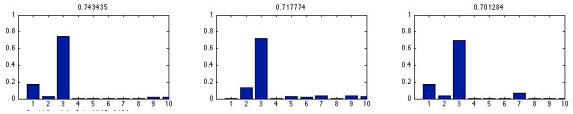


Figura 3.48. Conjunto de los 3 histogramas de las 3 primeras imágenes del aspecto 3: Estructuras Verticales.

Podemos observar en la Fig. 3.48, en la primera y tercera fotografía aparece una componente importante del aspecto ME1 Figura-Fondo posiblemente debido a las columnas que aparecen en el primer término.

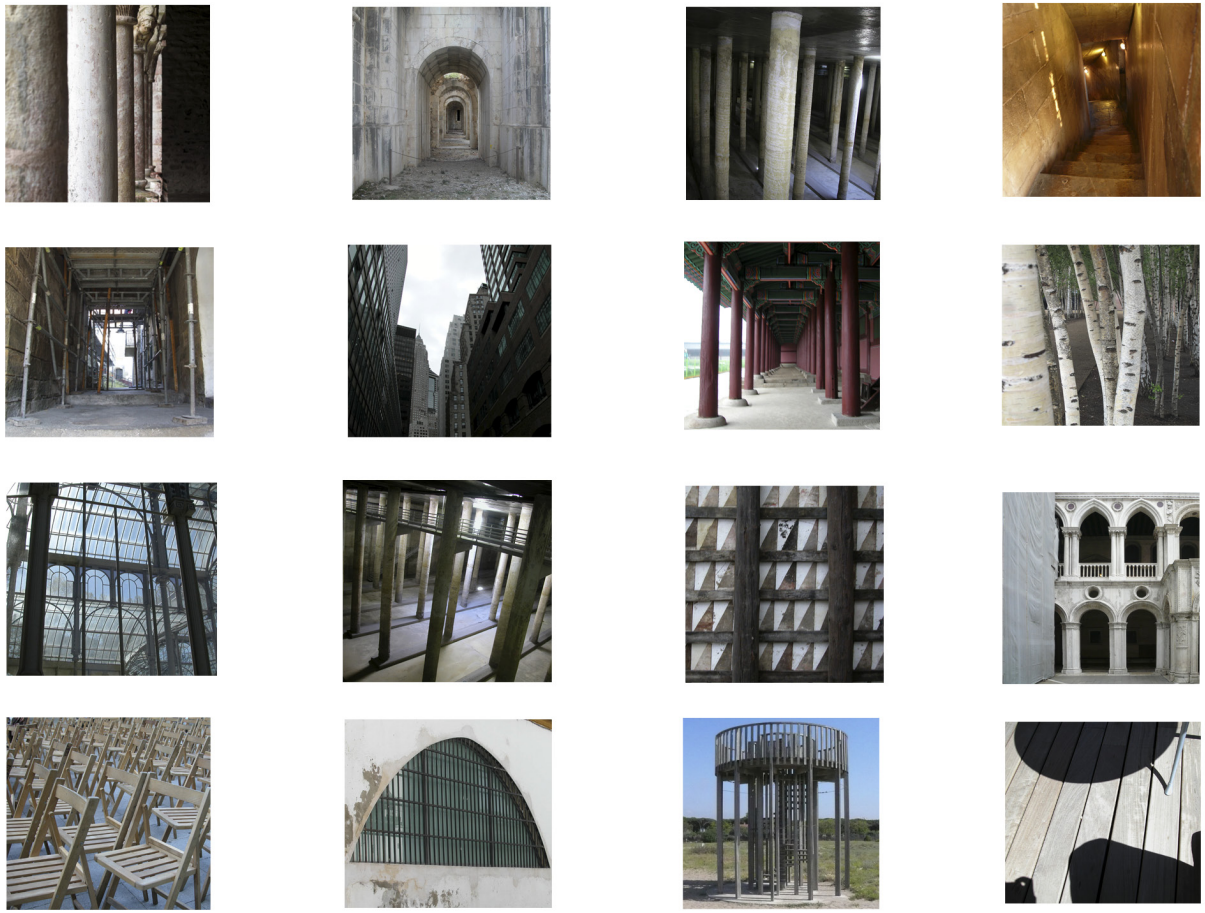


Figura 3.47. Conjunto de las 16 imágenes más entrópicas que presentan con mayor probabilidad el aspecto 3: Estructuras Verticales.

3.3.2.4 Aspecto ME4: Imagen Bipartita

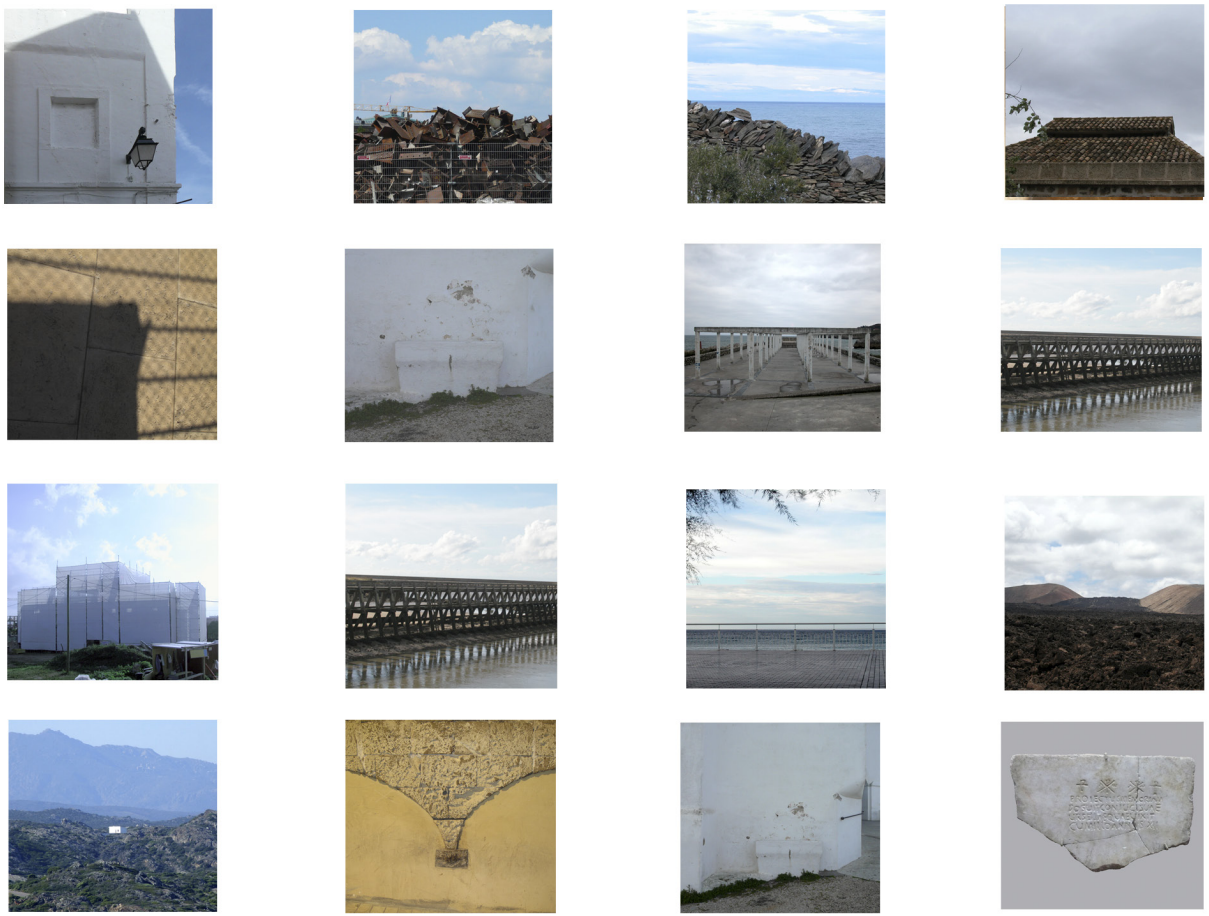


Figura 3.49. Conjunto de las 16 imágenes más entrópicas que presentan con mayor probabilidad el aspecto 4: Imagen Bipartita.

El aspecto Imagen Bipartita presenta claramente dos tipologías de textura dentro de la misma composición que suele coincidir con un paisaje de horizonte marcado, que en algunos casos se refiere a un elemento no paisajístico, pero en las probabilidades menores de 0.56. (Fig. 3.49)

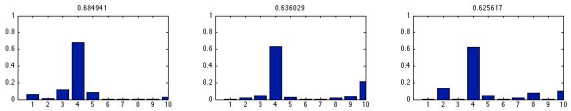


Figura 3.50. Conjunto de los 3 histogramas de las 3 primeras imágenes del aspecto 4: Imagen Bipartita.

El aspecto ME4 Imagen Bipartita presenta claramente dos tipologías de textura dentro de la misma composición que suele coincidir con un paisaje de horizonte marcado. Nótese en la primera imagen las componentes de aspecto ME3 Estructuras Verticales y aspecto ME5 cuadrícula. (Fig. 3.50).

3.3.2.5 Aspecto ME5: Cuadrícula

Aspecto claramente perceptible, sobre todo en la primera fila donde se pueden visualizar perfectamente la estructura cuadricular de la composición. (Fig. 3.51)

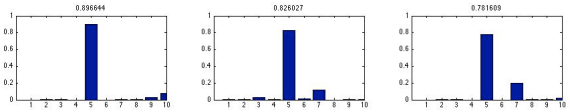


Figura 3.52. Conjunto de los 3 histogramas de las 3 primeras imágenes del aspecto 5: Imagen Cuadrícula.

Es curioso observar (Fig. 3.52) cómo a medida que disminuye la probabilidad del aspecto ME5 Cuadrícula, la retícula es menos perfecta y a la vez aumenta la probabilidad del aspecto ME7 Agrupaciones Rocosas.

Figura 3.51. Conjunto de las 16 imágenes más entrópicas que presentan con mayor probabilidad el aspecto 5: Cuadrícula.

3.3.2.6 Aspecto ME6: Estructuras Horizontales

Figura 3.53. Conjunto de las 16 imágenes más entrópicas que presentan con mayor probabilidad el aspecto 6: Estructuras Horizontales.

Este aspecto agrupa imágenes con una marcada tendencia a la horizontalidad en las estructuras predominantes. A diferencia del aspecto horizontal definido en las imágenes de baja entropía, éste no presenta únicamente líneas, sino más bien estructuras en forma de bandas. (Fig. 3.53)

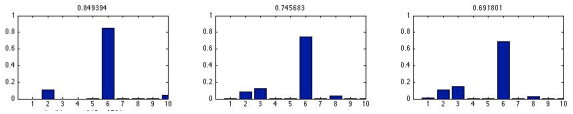


Figura 3.54. Conjunto de los 3 histogramas de las 3 primeras imágenes del aspecto Estructuras Horizontales.

A diferencia del aspecto PE3 Horizontal Amplia definido en las imágenes de poca entropía, éste no presenta únicamente líneas, sino más bien estructuras en forma de bandas. Es destacable la presencia en las 3 primeras imágenes del aspecto ME2 Textura Homogénea Bipolar (las 3 imágenes están compuestas básicamente por 2 texturas distintas) y la aparición en la segunda y tercera imágenes del aspecto ME3 Estructuras Verticales debido a la componente vertical de los ladrillos. (Fig. 3.54)

3.3.2.7 Aspecto ME7: Agrupaciones Rocosas

Bajo este aspecto percibimos como elemento común, en la mayoría de las imágenes, la piedra. Algunas presentan guijarros y otras restos de muros. Muchas de ellas están dispuestas por la mano del hombre en forma de muro o presentan un cierto orden natural que las caracteriza. (Fig. 3.55)

Es interesante prestar atención a la sutileza de clasificación que el sistema demuestra al categorizar en dos clases diferentes las dos construcciones de piedras que muestra la Fig. 3.56.

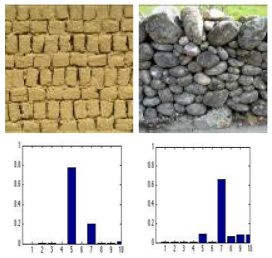


Figura 3.56. Primera imagen tipificada en el aspecto ME5 Cuadrícula y segunda imagen tipificada en el aspecto ME7 Agrupaciones Rocosas y sus histogramas de aspectos correspondientes.

Figura 3.55. Conjunto de las 16 imágenes más entrópicas que presentan con mayor probabilidad el aspecto 7: Agrupaciones Rocosas.

3.3.2.8 Aspecto ME8: Líneas y Planos Interseccionados

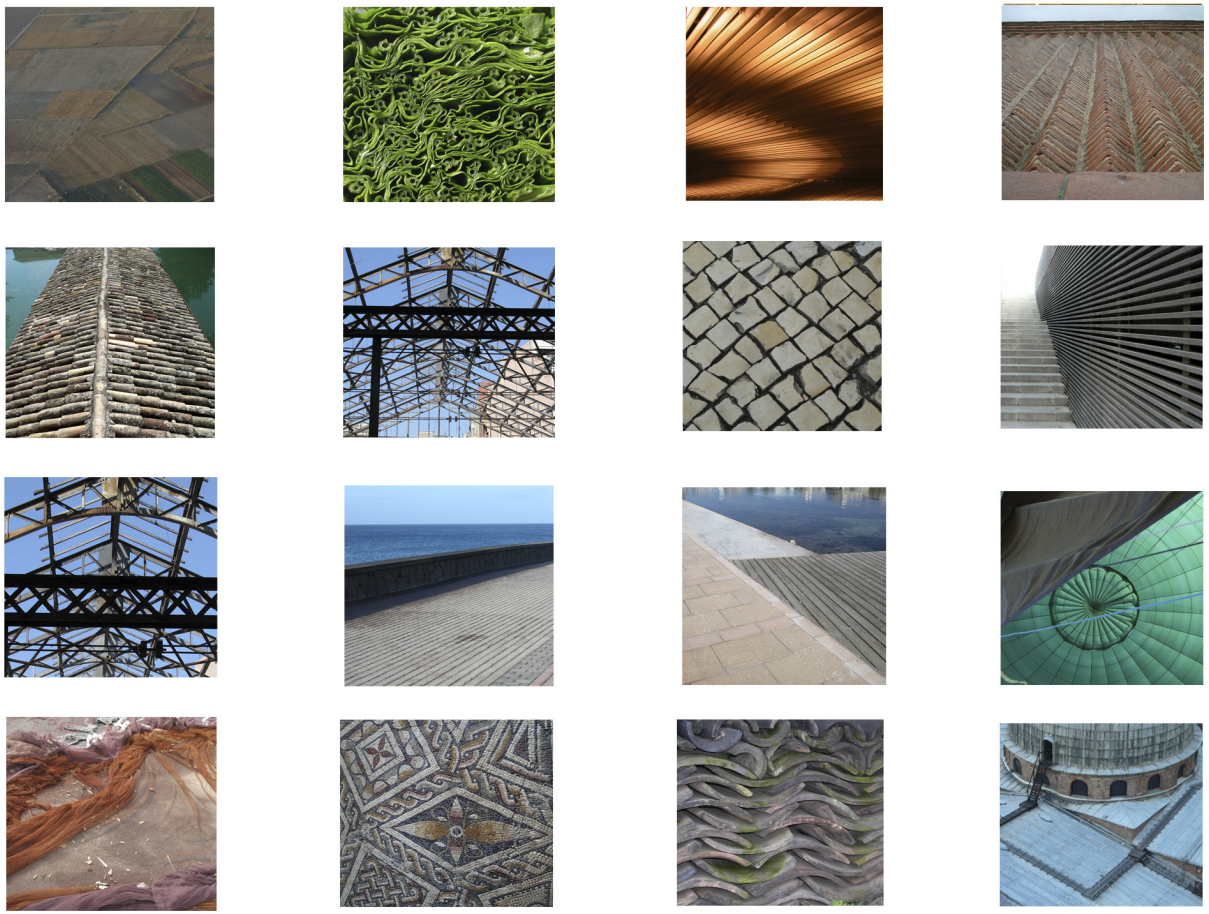


Figura 3.57. Conjunto de las 16 imágenes más entrópicas que presentan con mayor probabilidad el aspecto 8: Líneas y Planos Interseccionados.

En el aspecto Líneas y Planos Interseccionados se recogen un conjunto de imágenes en las que predominan diferentes direccionalidades superpuestas en un mismo plano o que confluyen y que generalmente llenan todo el espacio. (Fig. 3.57)

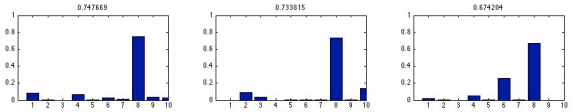


Figura 3.58. Conjunto de los 3 histogramas de las 3 primeras imágenes del aspecto Estructuras Líneas y Planos Interseccionados.

Es destacable la componente de aspecto ME6 Estructuras Horizontales de la imagen tercera, que puede explicarse si nos fijamos en las zonas sombreadas que crean unas bandas con tendencia horizontal. (Fig. 3.58)

3.3.2.9 Aspecto ME9: Estructura Heterogénea Contrastada

Este aspecto es muy compacto y aglutina un conjunto de imágenes en las que predominan estructuras volumétricas conseguidas por un claroscuro muy contrastado. (Fig. 3.59 y 3.60)

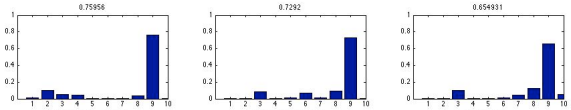


Figura 3.60. Conjunto de los 3 histogramas de las 3 primeras imágenes del aspecto Estructuras Heterogénea Contrastada.

Figura 3.59. Conjunto de las 16 imágenes más entrópicas que presentan con mayor probabilidad el aspecto 9: Estructura Heterogénea Contrastada.

3.3.2.10 Aspecto ME10: Figura con Fondo Texturado

Figura 3.61. Conjunto de las 16 imágenes más entrópicas que presentan con mayor probabilidad el aspecto 10: Figura con Fondo Texturado.

Este conjunto de imágenes presenta recuerda mucho al aspecto ME1 Figura-Fondo, con la variante de que el fondo no es liso, sino texturado. Destacamos el componente de aspecto ME4 Imagen Bipartita en la segunda imagen. Este aspecto tampoco resulta es fácil de interpretar. (Fig. 3.61 y 3.62)

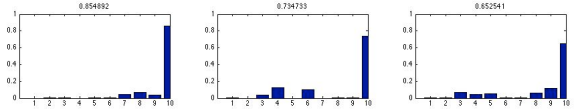


Figura 3.62. Conjunto de los 3 histogramas de las 3 primeras imágenes del aspecto figura con Fondo Texturado.

3.3.3 Aspectos Latentes del conjunto de imágenes de la colección Tàpies

A la vista de los resultados comentados en los apartados anteriores se decidió dar un paso más en la investigación poniendo a prueba la metodología de aprendizaje no supervisado del *pLSA* en un nuevo conjunto de datos: la colección de 434 imágenes digitalizadas de pintura y obra gráfica del artista Antoni Tàpies que posee su Fundación en Barcelona (Tàpies, 2001). En este nuevo conjunto se detectaron algunas imágenes de obra escultórica del artista y fueron descartadas debido a que las representaciones en dos dimensiones de una obra tridimensional dependen del punto de vista desde el que se ha tomado la instantánea y su inclusión en el análisis podía distorsionar la construcción de los descriptores del conjunto.

El artista Antoni Tàpies fundó en los años 40 en Barcelona el grupo *Dau al Set* junto a Modest Cuixart, Juan José Tharrats, Joan Brossa y Arnau Puig. Apostó en un principio por un tipo de pintura a caballo entre el surrealismo tardío y ciertas concepciones dadaístas, pero en los años 50 descubrió el papel expresivo de la materia (materiales heterogéneos, encontrados...) y entró abiertamente en el lenguaje informalista que imponían desde París Fautrier, Dubuffet y Wols. Su proceso creativo se centra en las posibilidades de la expresión matérica, próxima al grattage, amplios gestos, manchas y chorreados, y en cuadros de tonalidades neutras en los que el óleo se combina con técnicas mixtas (Guasch, 1997,p. 36) En palabras literales de Ruhrberg:

> Tàpies es tal vez el representante más significativo del arte informal y la pintura matérica. Sus obras están conformadas con una gran inteligencia artística organizativa. Las emociones que penetran en ellas tienden a tener una naturaleza tranquila y meditativa, y no el carácter volcánico tan característico de otros abstractos gestuales. En lugar de expresar sus impulsos íntimos en el campo de la pintura, ordena conscientemente el plano en forma de paisajes expansivos, a menudo vacíos, y formas simples y generosas, que han recibido una estructura ordenada...
> Tàpies crea su propia realidad: paisajes pictóricos cuya tensión formal procede del contraste entre los espacios vacíos y silenciosos y las configuraciones dotadas de forma, entre positivo y negativo, entre protuberancias y depresiones, entre azar y orden,

> libertad y control...
> A menudo oscuros, pardos, aparentemente monocromos pero en realidad atravesados por una gama de colores extraordinariamente sutil, sus cuadros parecen apartados del contexto del tiempo presente. (Ruhrberg, Schneckenburger, Fricke & Honnef, 2001, p. 260)

La colección que ha sido facilitada por la Fundació Antoni Tàpies (Tàpies, 2001) para llevar a cabo el presente estudio contiene obras del artista pertenecientes a distintas épocas y realizadas con múltiples técnicas sobre soportes variados; incluye grabados, acuarelas, dibujos, pinturas y collages con diferentes tipologías de materiales incorporados.

Las espectativas del experimento serían que el sistema informático fuese capaz de capturar en su análisis estadístico algún tipo de patrón entre los elementos que componen las obras. Serían deseables resultados que ayudasen a desvelar ciertas relaciones que el artista establece en las configuraciones de las formas de sus obras para construir su lenguaje. Al contener el conjunto obras del artista pertenecientes a distintas épocas, el estudio también ayudaría a detectar las conformaciones que se mantienen en el tiempo y las que desaparecen. De alguna manera las conclusiones nos podrían acercar al proceso creativo del artista.

En la colección de imágenes anteriormente estudiada del artista Planas, las palabras visuales estaban constituidas por pequeñas regiones de imágenes que se correspondían con elementos naturales tales como agua, piedras, cielo etc. En el presente experimento las palabras se corresponderán con elementos totalmente abstractos creados por el artista que será necesario interpretar.

Se decide continuar trabajando con un vocabulario de 300 palabras visuales dado el buen rendimiento obtenido en el experimento anterior. Se realiza un tanteo de los resultados respecto al número de aspectos latentes y se comprueba que la muestra de 434 imágenes queda bien representada con 13 aspectos latentes.

El valor del índice de entropía de Shannon (ver apartado 8 del Anexo A) para una imagen *d*

que estuviese asociada por igual a los 13 aspectos, es decir, con vector de probabilidades con 1/13 en cada componente, sería máximo e igual a H(d)=2.5649.

Así, para el caso de estudio que nos ocupa del artista Tàpies, los rangos de entropía teóricos respecto a 13 aspectos irían de 0 a 2.5649. Los observados en nuestra muestra van de prácticamente 0 a 2.2226. Las imágenes que tienen una entropía elevada son aquellas que el procedimiento ha asociado de manera equiprobable a cada uno de los aspectos.

De esta forma se decide seleccionar del total de la muestra las imágenes con un valor de entropía superior a 1,65 y repetir de nuevo la búsqueda de aspectos en este nuevo conjunto formado por 228 imágenes. Se calcula de nuevo todo el proceso generando los descriptores locales, el vocabulario visual y los aspectos latentes intentando que el sistema sea capaz de establecer nuevas relaciones entre imágenes visualmente más complejas dando lugar a nuevos aspectos latentes distintos de las 13 primeros.

Los resultados de esta prueba, a diferencia del caso anteriormente tratado de Planas, son difíciles de interpretar. Es bastante comprensible dado que el conjunto anterior constaba de 2846 imágenes y el actual de Tàpies únicamente tiene 434.

Pasaremos a comentar los 13 aspectos latentes resultantes de computar el total de imágenes de la colección Tàpies sin hacer segregación por el índice de entropía de Shannon, aunque si presentamos las figuras correspondientes a las tipologías de imágenes poco entrópicas y más entrópicas del artista por considerarlas ilustrativas (Fig. 3.63 y Fig. 3.64).

En este caso no utilizaremos las siglas PE ni ME delante del nombre del aspecto ya que, para calcularlos se ha utilizado toda la muestra sin tener en cuenta el índice de entropía.

En los conjunto de 16 imágenes que se muestran para explicar las características de cada aspecto latente, las fotografías están ordenadas de mayor a menor probabilidad de tener el aspecto comentado (de izquierda a derecha y de arriba a abajo). Así, la imagen con mayor probabilidad de tener el aspecto concreto es la primera de arriba a la izquierda y la menos probable es la número 16 que quedará abajo a la derecha.

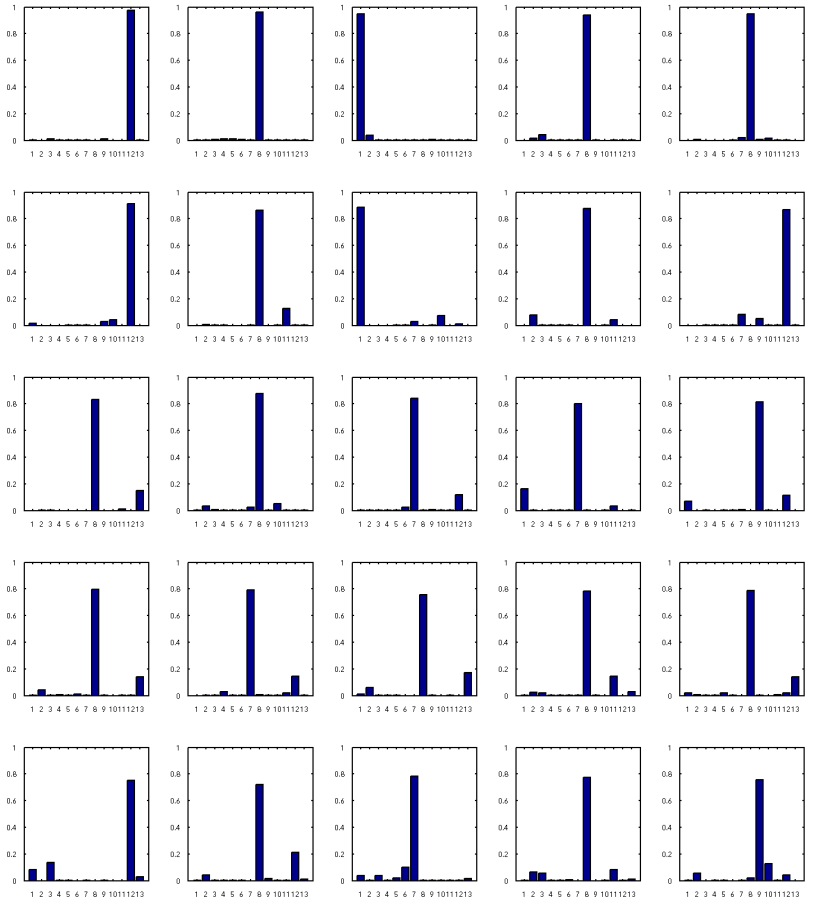


Figura 3.63. A la izquierda presentamos una muestra de las imágenes menos entrópicas según el índice de Shannon pertenecientes a la colección de Tàpies, y a su derecha una muestra de sus histogramas según la representación por distribución de aspectos resultante de la aplicación del modelo *pLSA*.

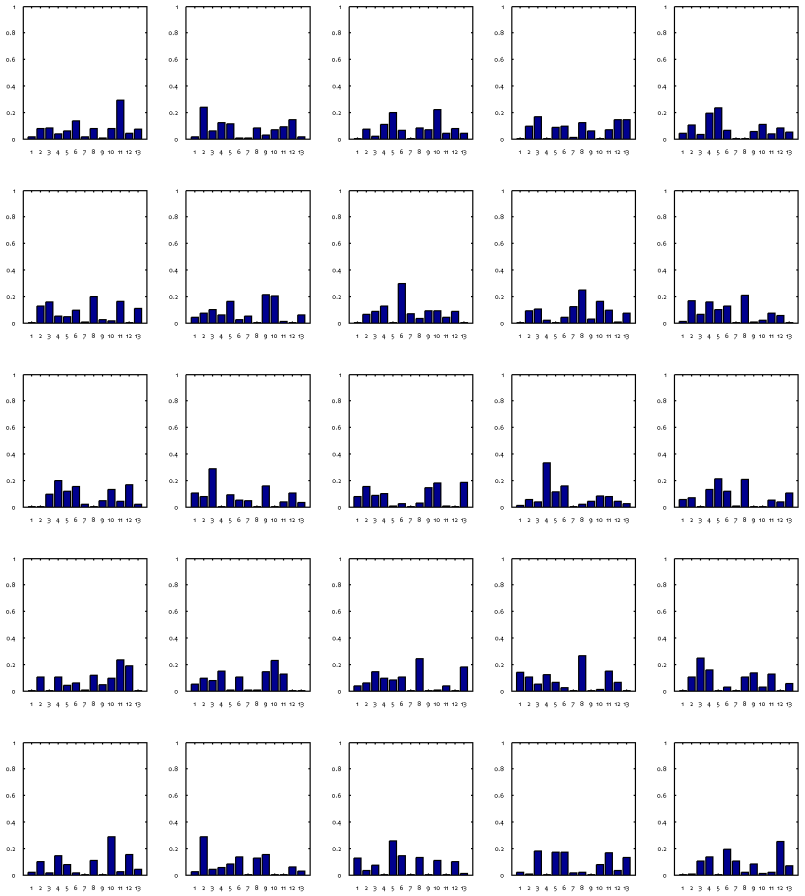


Figura 3.64. A la derecha presentamos una muestra de las imágenes más entrópicas según el índice de Shannon pertenecientes a la colección de Tàpies, y a su izquierda una muestra de sus histogramas según la representación por distribución de aspectos resultante de la aplicación del modelo *pLSA*.

Figura 3.65. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 1: Trazo Vibrante Paralelo.

3.3.3.1 Aspecto 1: Trazo Vibrante Paralelo

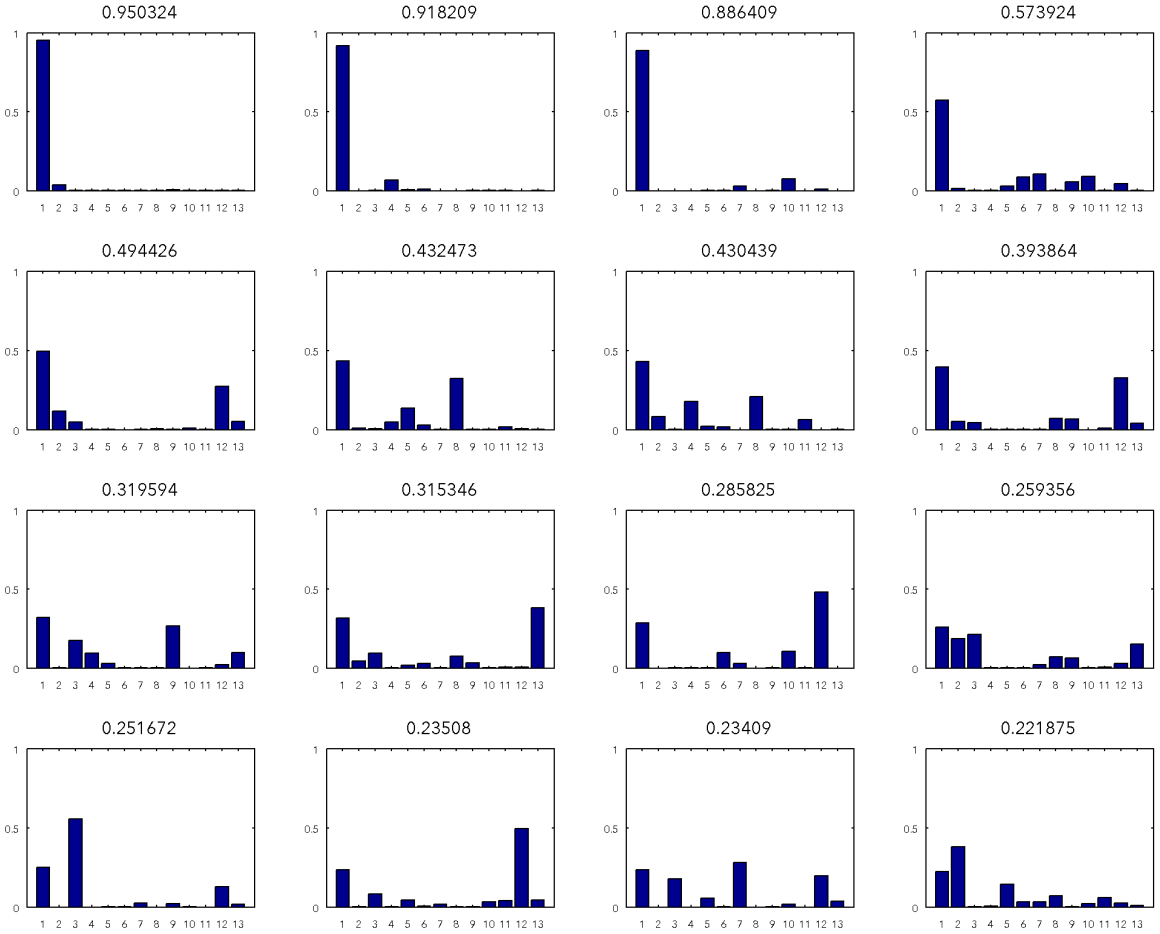


Figura 3.66. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 1: Trazo Vibrante Paralelo.

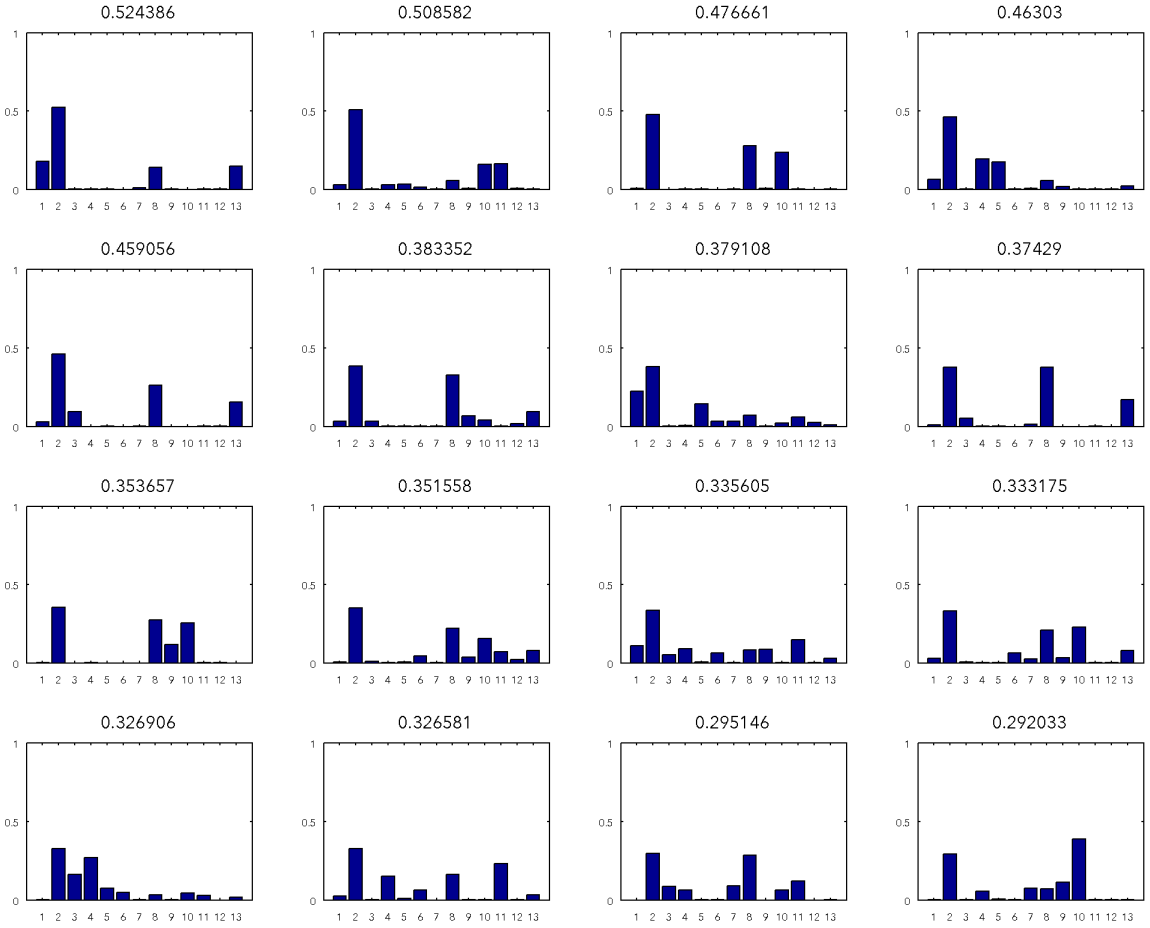
Este aspecto es muy claro en las 3 primeras imágenes, pero la décima tiene más probabilidad del aspecto 13 que del 1, por lo que centraremos nuestro análisis en las 9 primeras (Fig. 3.65 y 3.66).

El aspecto detecta una distribución de líneas sinuosas bastante cercanas unas de otras que se percibe muy claramente en las imágenes número 2, 3, 5, 6 y 7. Muchas veces corresponde con texto, y por ejemplo, en las imágenes 6 y 7, en las que el texto sólo se encuentra en una de las dos páginas, el aspecto 1 tiene la mitad de probabilidad que en la imagen 2 en la que e l texto corresponde prácticamente a todo el plano.

Aparentemente es importante la horizontalidad, pero si consideramos las imágenes 1 y 4, podemos concluir que no es determinante del aspecto ya que estas contienen los grupos de lineas también inclinados.

Figura 3.67. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 2: Figura Contrastada.

3.3.3.2 Aspecto 2: Figura Contrastada

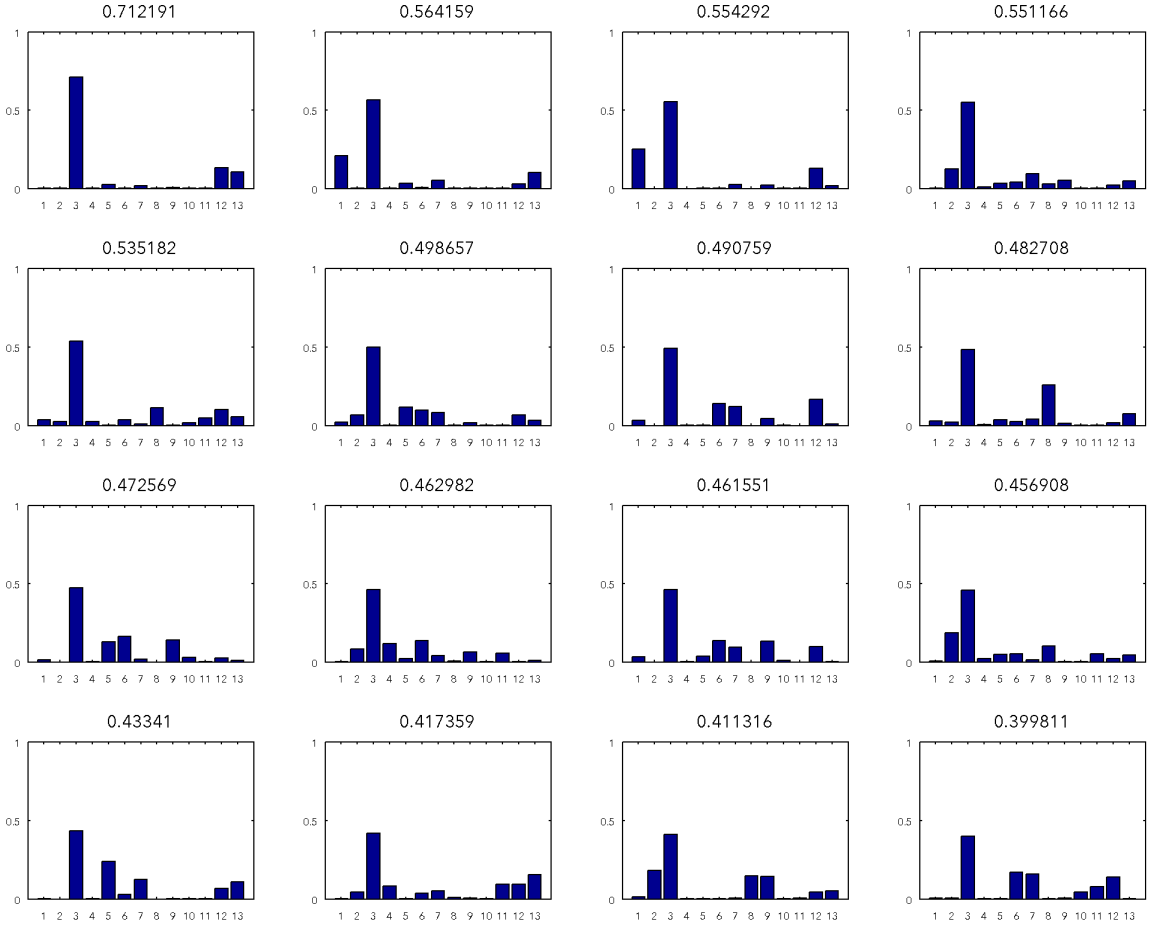


Presencia de objetos o formas descatadas de formas contundentes contrastadas sobre un fondo más o menos homogéneo.

Figura 3.68. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 2: Figura Contrastada.

Figura 3.69. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 3: Línea Narrativa - Figurativa.

3.3.3.3 Aspecto 3: Línea Narrativa - Figurativa

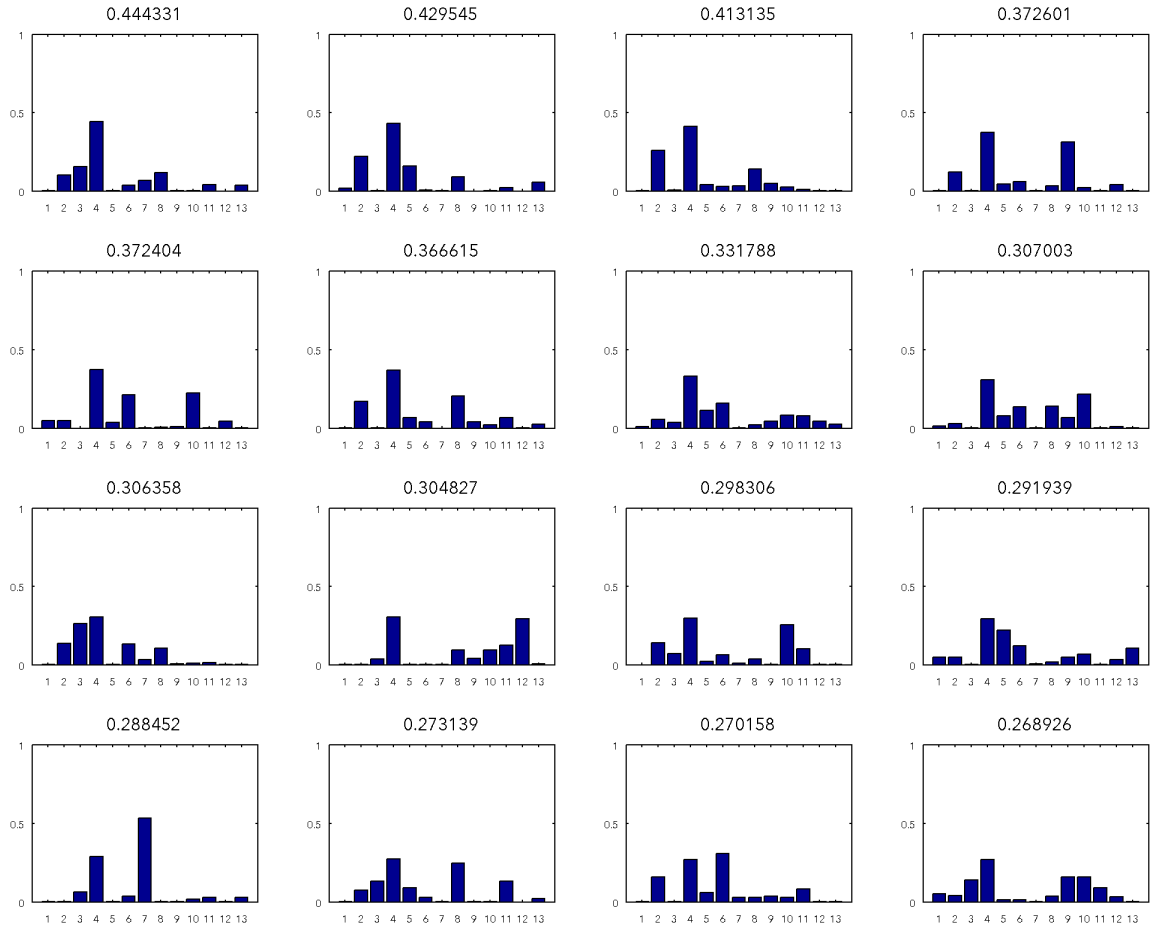


Muchas de las obras que agrupa este aspecto son narrativas, aunque no todas. Si vemos lo que comparten las imágenes figurativas con las abstractas podría ser un tipo de línea sensible que ocupa prácticamente toda la superficie y de diferentes gruesos. La imagen 3 pertenece a un gofrado, pero la incidencia de la luz hace que la máquina la trate como una línea. Es un aspecto bastante homogéneo. (Fig. 3.69 y 3.70)

Figura 3.70. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 3: Línea Narrativa - Figurativa.

Figura 3.71. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 4: Trazo Grueso Denso.

3.3.3.4 Aspecto 4: Trazo Grueso Denso

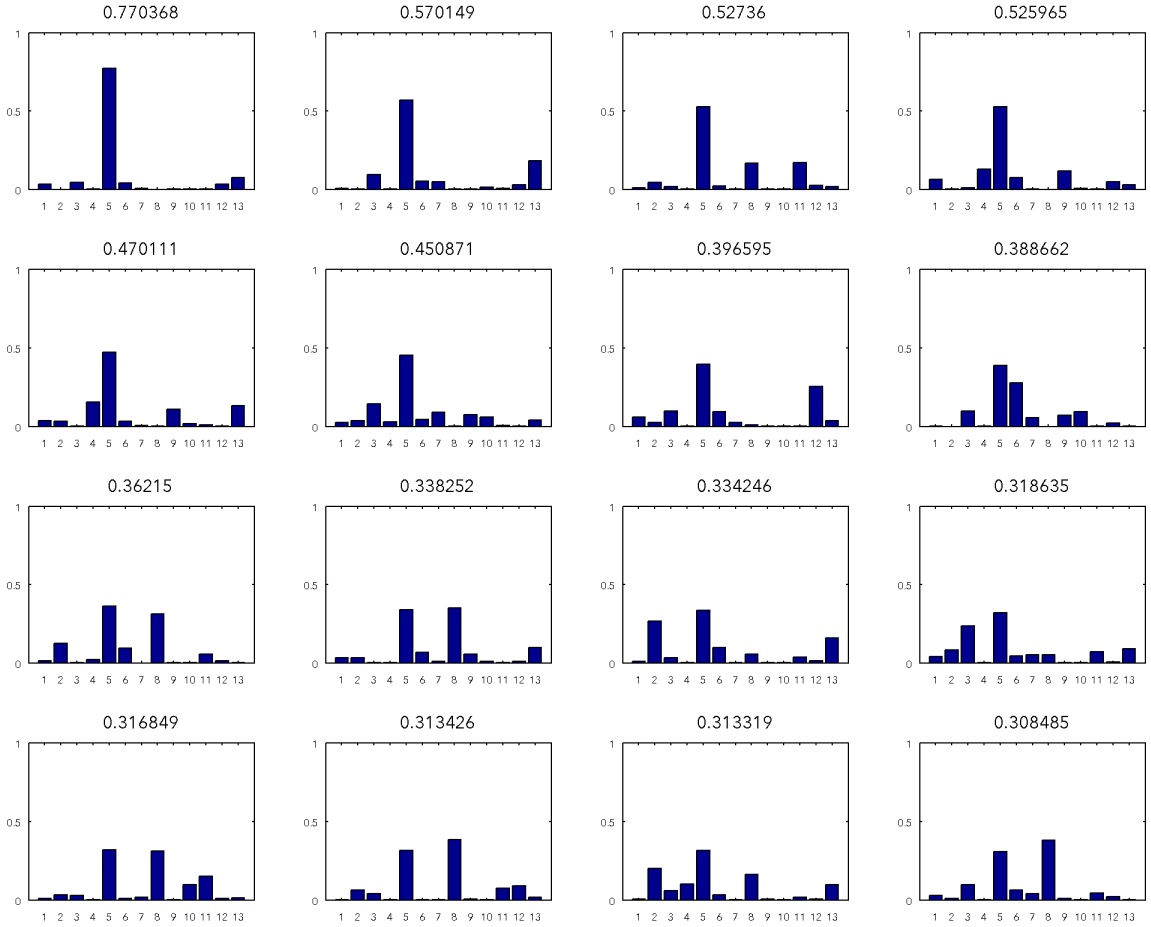


Aspecto muy compacto. Pocos trazos, amplios y gestuales. La mayoría son pinceladas oscuras sobre fondo más claro, pero también hay ejemplos del caso contrario. Tendencia a que el motivo principal se sitúe en el centro abarcando toda la imagen. (Fig. 3.71 y 3.72)

Figura 3.72. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 4: Trazo Grueso Denso.

Figura 3.73. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 5: Trazo Texturado.

3.3.3.5 Aspecto 5: Trazo texturado

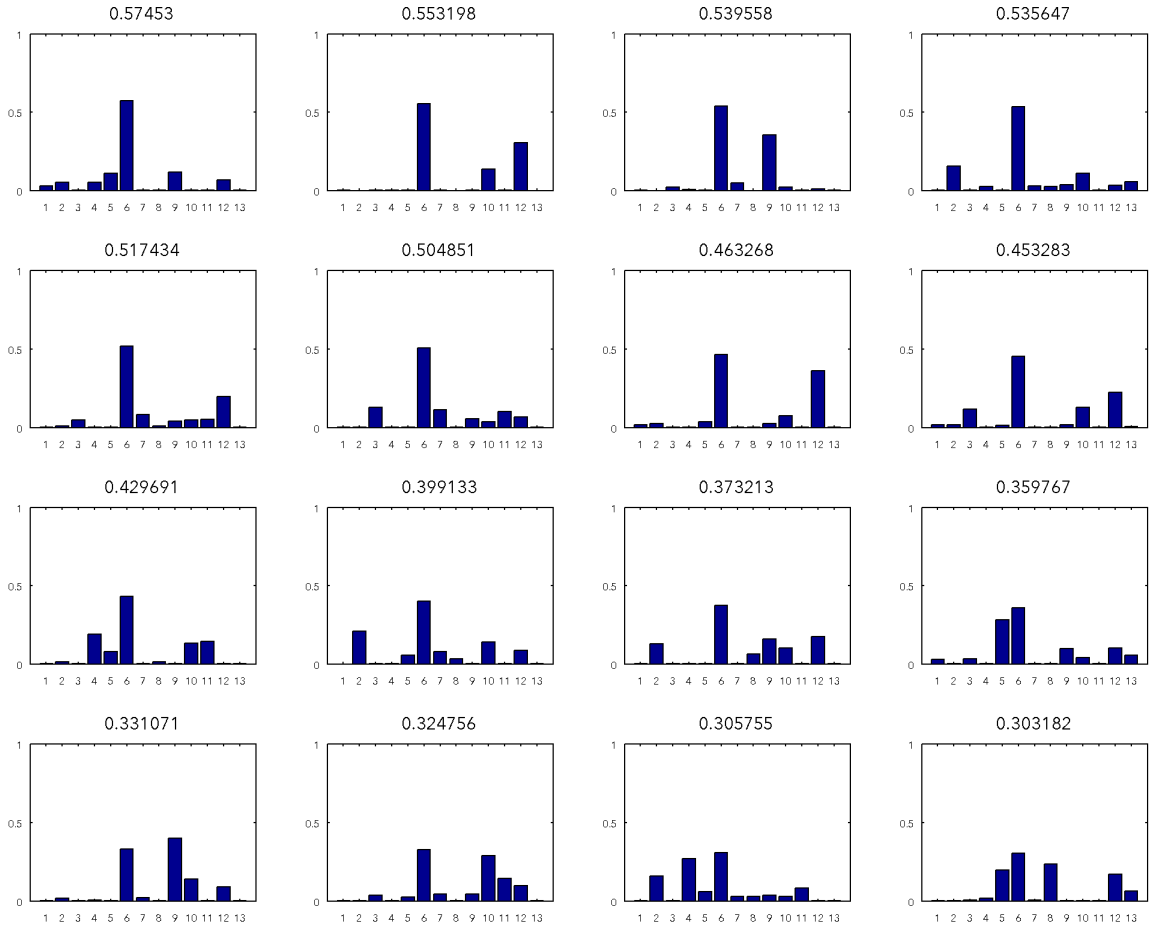


Las imágenes más representativas son las 4 primeras. El aspecto agrupa obras que presentan una distribución dispersa de zonas claras de tinta más plana, aunque no totalmente lisa, con zonas más grisáceas en las que la pincelada no es totalmente densa. (Fig. 3.73 y 3.74)

Figura 3.74. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 5: Trazo Texturado.

Figura 3.75. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 6: Atmósfera Difusa.

3.3.3.6 Aspecto 6: Atmósfera Difusa



Aspecto que presenta imágenes con un tratamiento principalmente atmosférico y difuso. (Fig. 3.75 y 3.76)

Figura 3.76. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 6: Atmósfera Difusa.

Figura 3.77. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 7: Fondo Tramado.

3.3.3.7 Aspecto 7: Fondo Tramado

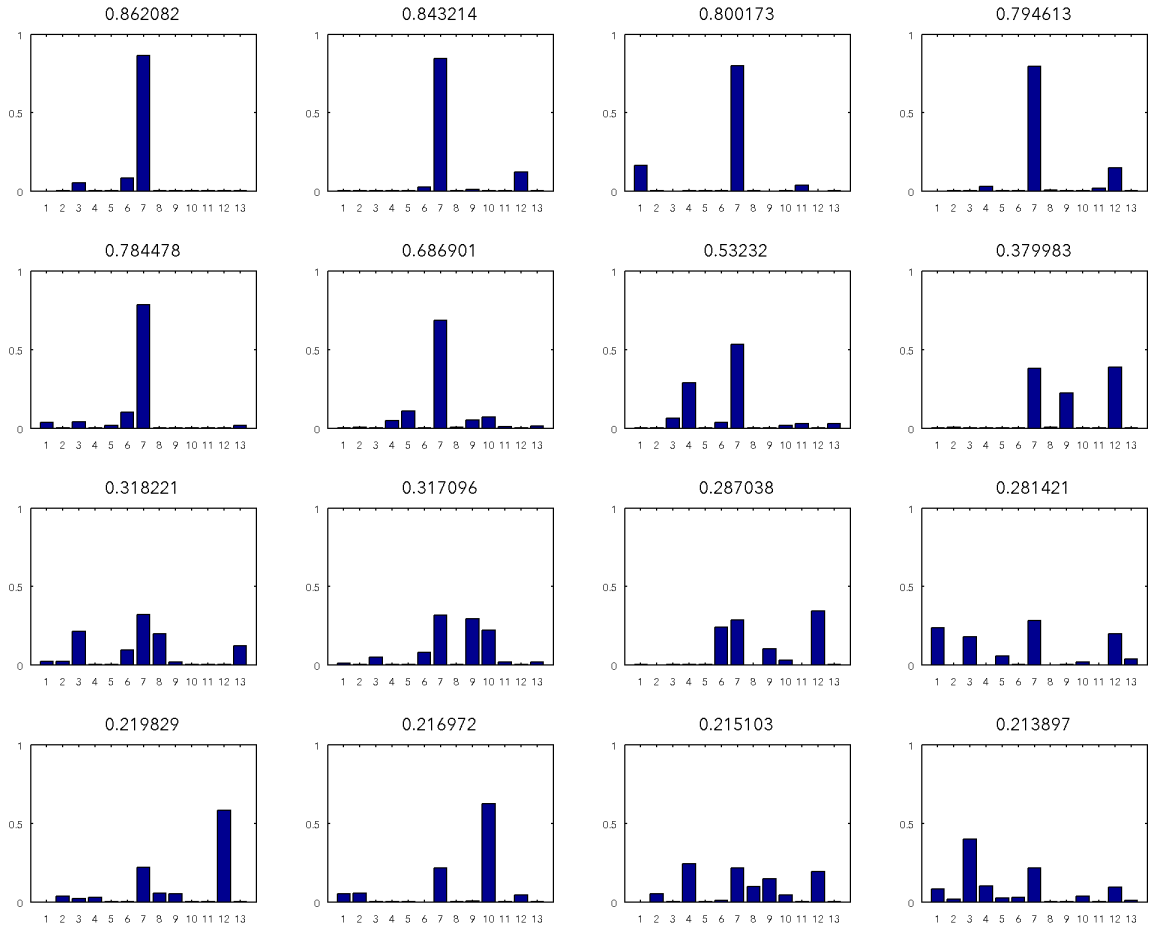
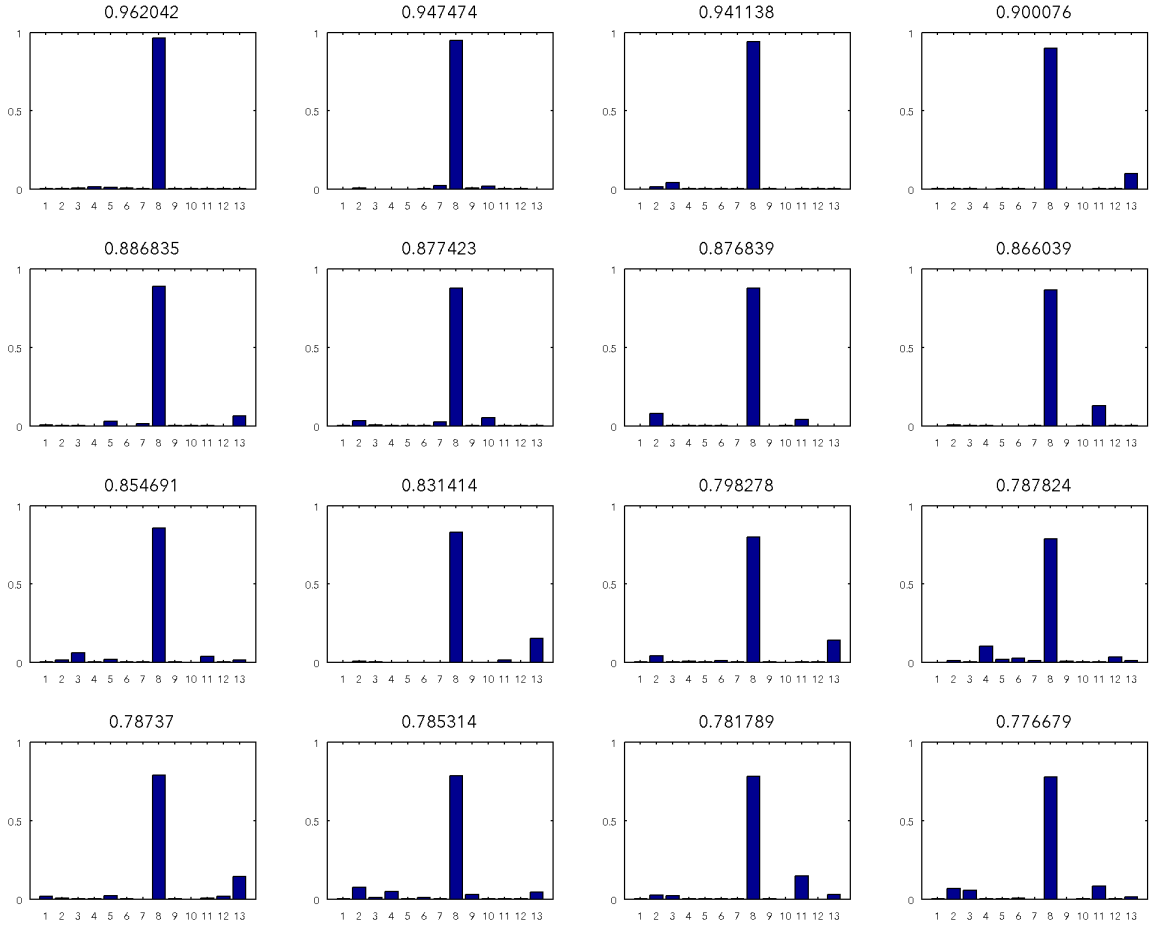


Figura 3.78. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 7: Fondo Tramado.

Este aspecto es bastante claro. Si analizamos las cuatro primeras imágenes vemos que comparten un fondo con textura tramada. El aspecto se refiere al fondo y no a los diferentes motivos que presentan las obras, y lo podemos comprobar prestando atención a la imagen 7 que tiene mayoritariamente este aspecto y el motivo oscuro que presenta en posición más centrada pertenece a la cualidad del aspecto 4 gestual. (Fig. 3.77 y 3.78)

Figura 3.79. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 8: Detalle sobre Fondo Plano.

3.3.3.8 Aspecto 8: Detalle sobre Fondo Plano



Uno de los aspectos más claros de toda la colección. Gran cantidad de espacio vacío, limpio, medido, con uno o varios motivos muy pensados situados en posición más o menos central o equilibrada. (Fig. 3.79 y 3.80)

Figura 3.80. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 8: Detalle sobre Fondo Plano.

Figura 3.81. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 9: Trazo Oscuro sobre Fondo Claro.

3.3.3.9 Aspecto 9: Trazo Oscuro sobre Fondo Claro

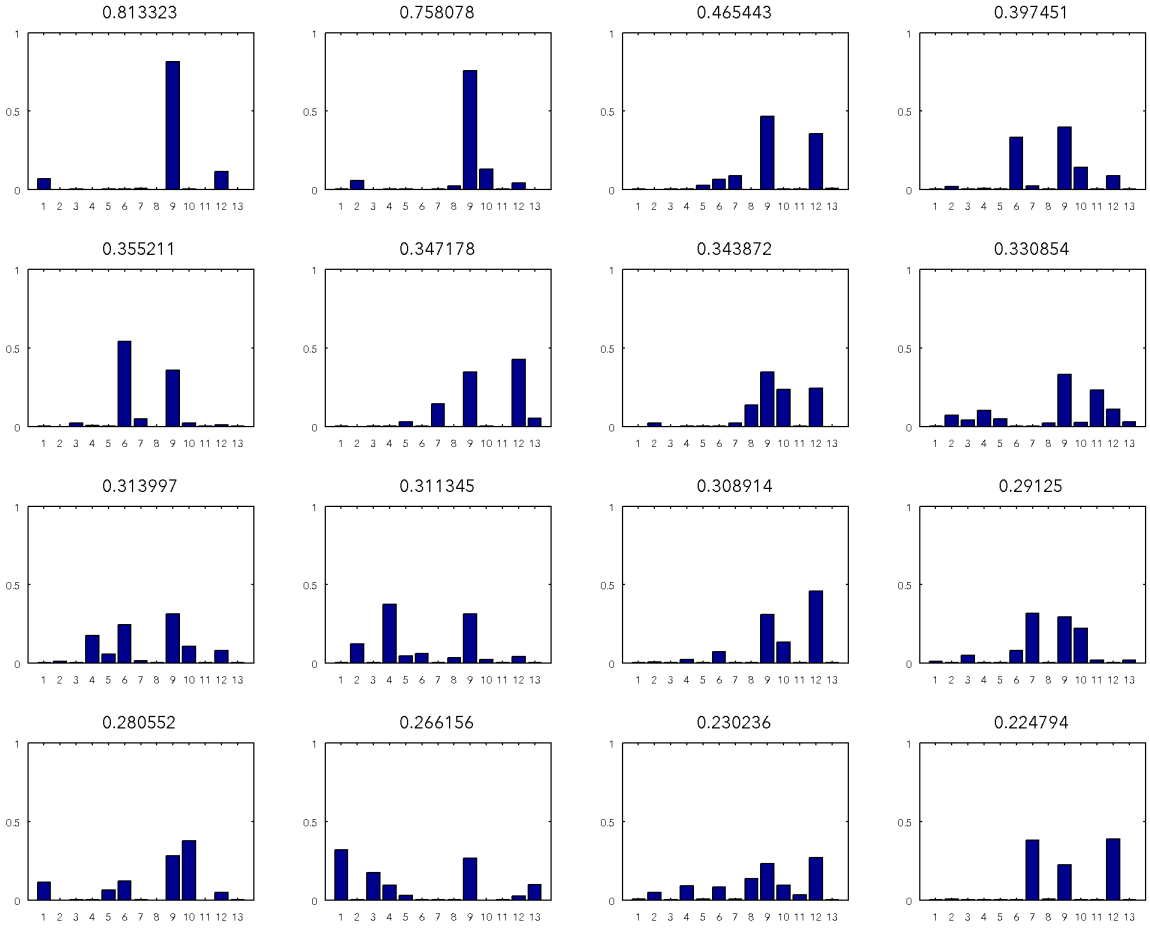


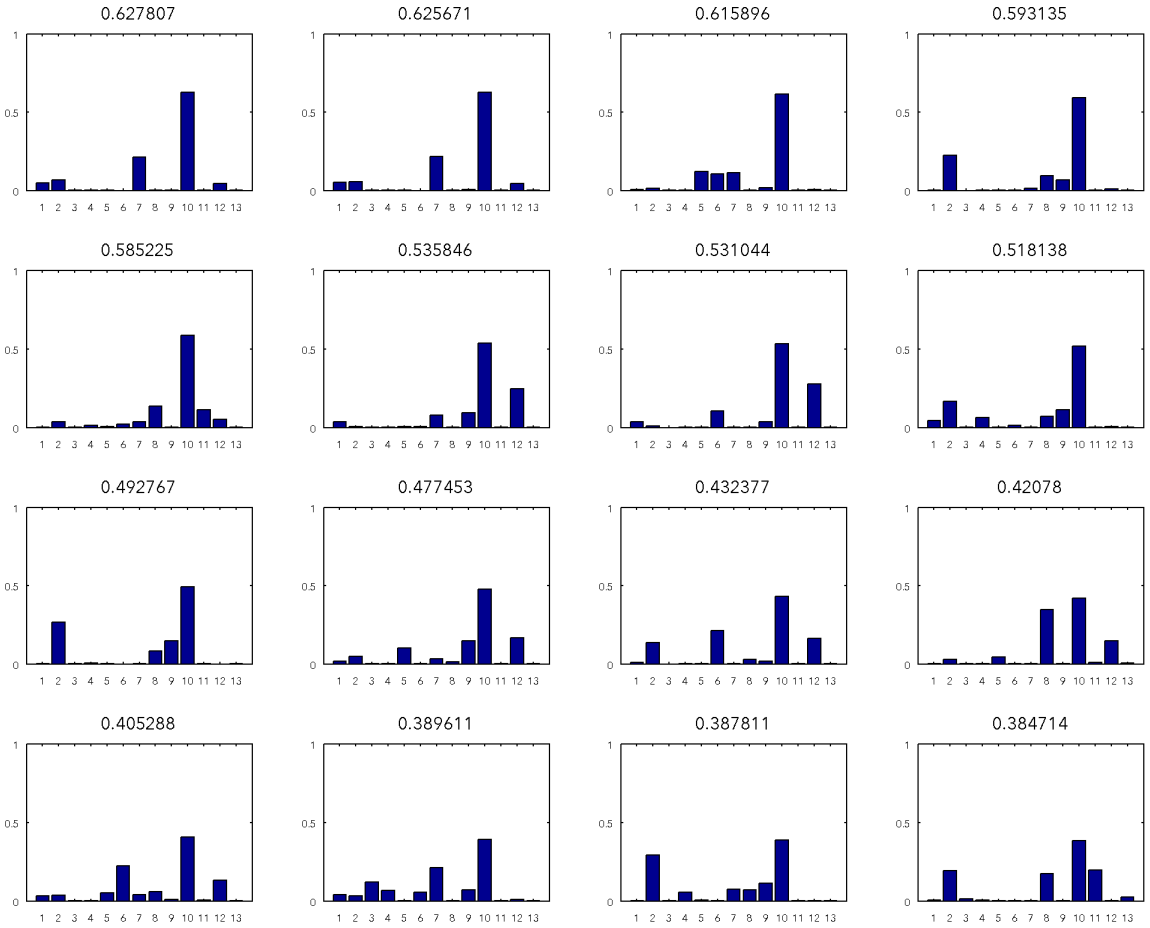
Figura 3.82. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 9: Trazo Oscuro sobre Fondo Claro.

Aspecto complicado que se presenta más claro en las 2 primeras imágenes y que en las otras comparte protagonismo con varios aspectos más. Se trata por lo tanto de un aspecto con bastante entropía.

Las dos primeras imágenes presentan un trazo vigoroso de color oscuro sobre un fondo más claro. (Fig. 3.81 y 3.82)

Figura 3.83. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 10: Trazo Claro sobre Fondo Oscuro.

3.3.3.10 Aspecto 10: Trazo Claro sobre Fondo Oscuro

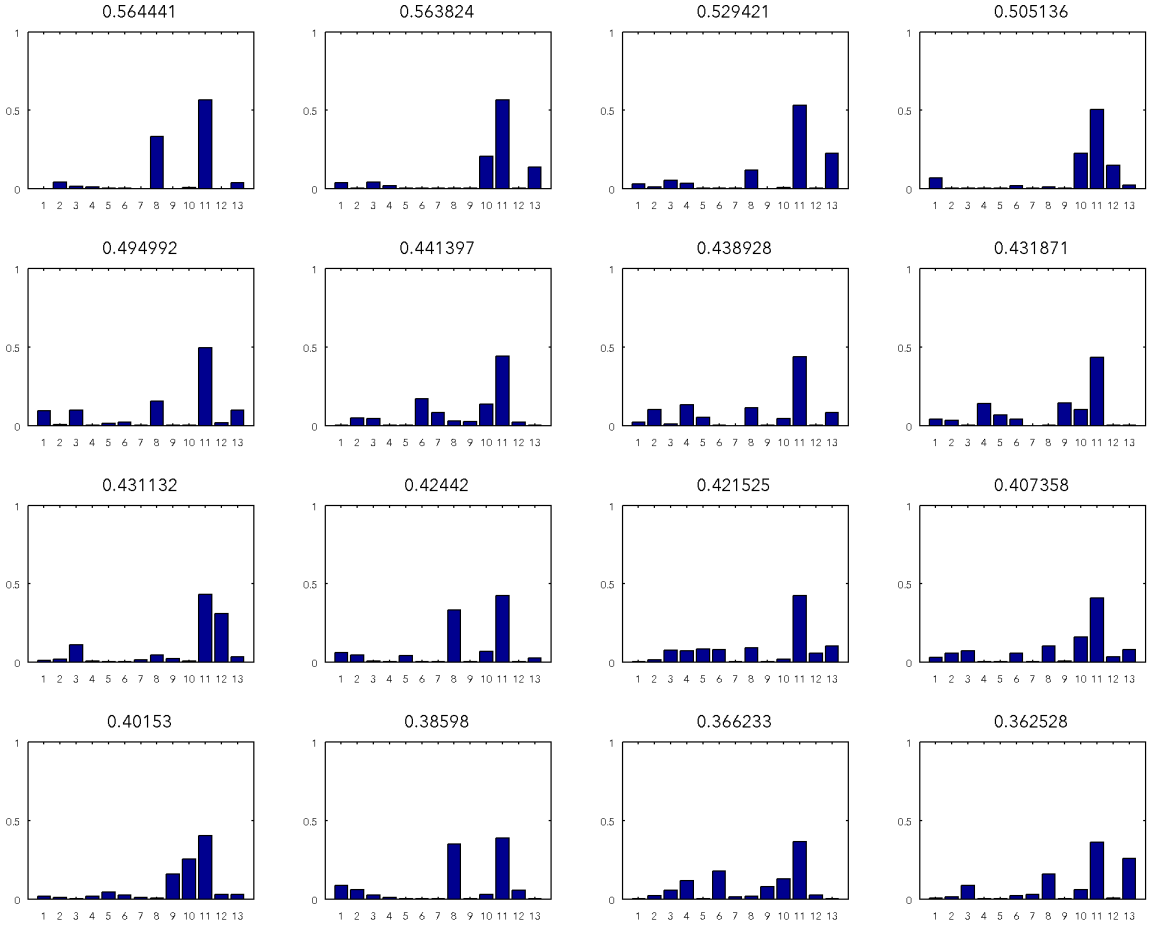


Aspecto en el que lineas claras se dibujan sobre un fondo oscuro. (Fig. 3.83 y 3.84)

Figura 3.84. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 10: Trazo Claro sobre Fondo Oscuro.

Figura 3.85. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 11: Equilibrio Compositivo.

3.3.3.11 Aspecto 11: Equilibrio Compositivo

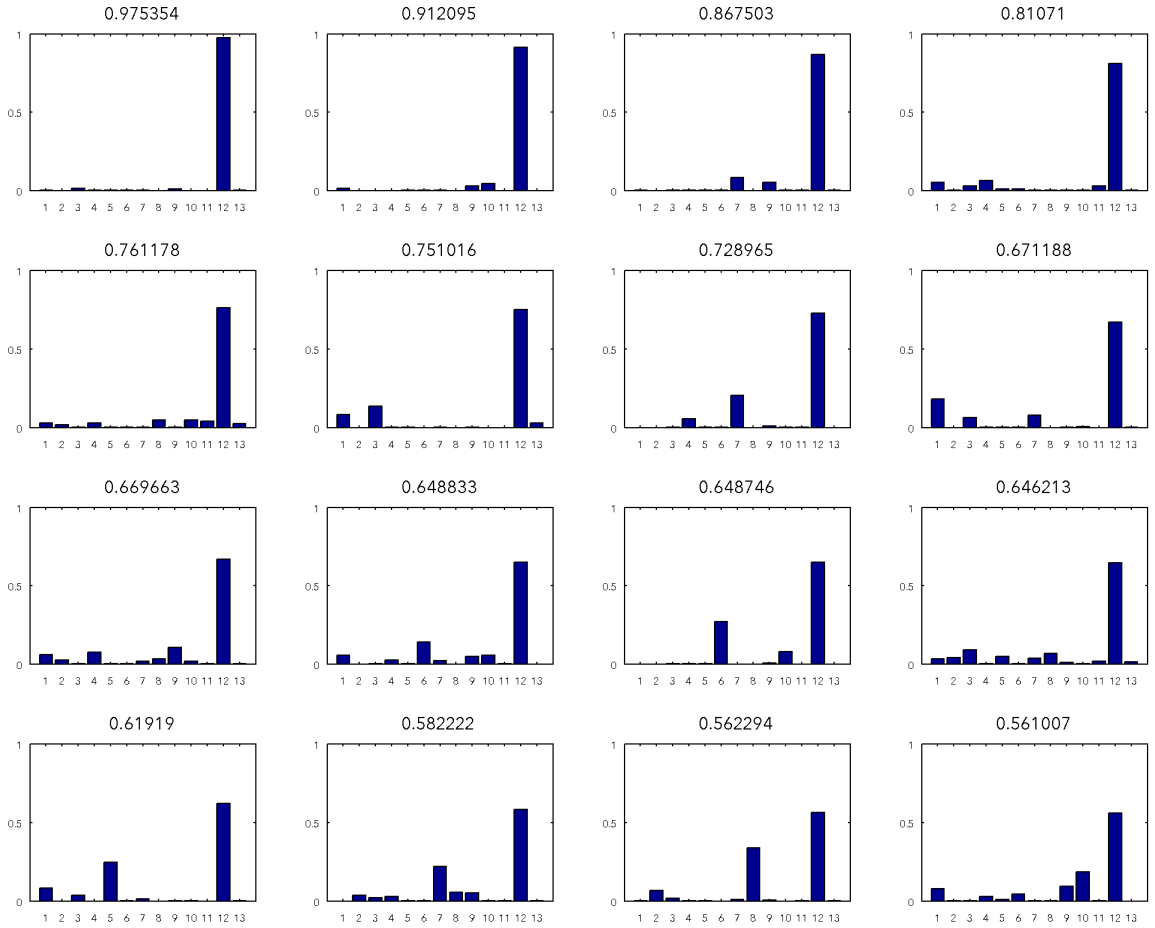


A pesar de tratarse de un aspecto bastante entrópico, agrupa imágenes de obras que transmiten idea de orden, composición y equilibrio. En bastantes de ellas aparecen figuras rectangulares o cuadradas bien definidas. Transmite también la sensación de simetría. (fig. 3.85 y 3.86)

Figura 3.86. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 11: Equilibrio Compositivo.

Figura 3.87. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 12: Textura Granulada.

3.3.3.12 Aspecto 12: Textura Granulada

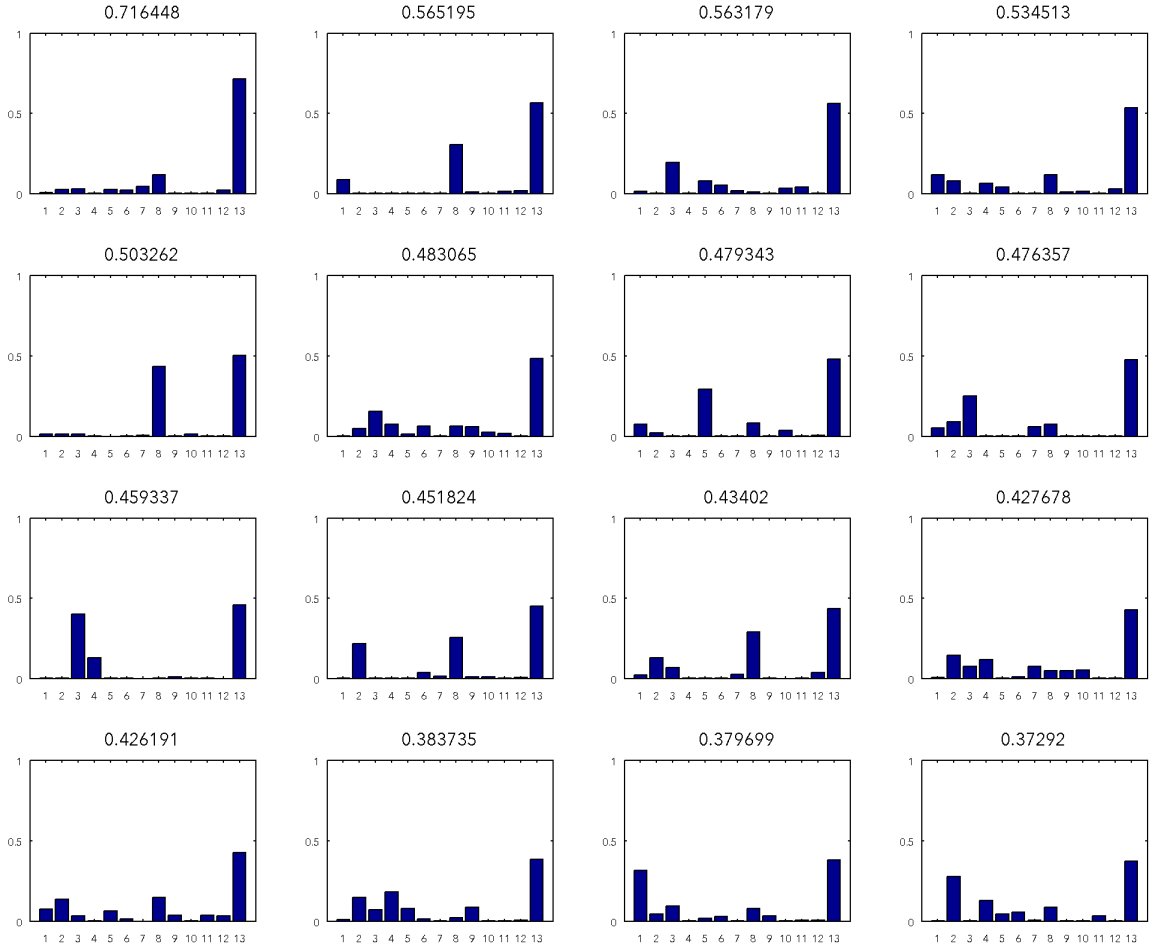


Aspecto claro referido a un fondo granulado, algunas de las obras contienen arena adherida y otras presentan una superficie punteada o en la que la pintura parece haber sido vaporizada. (Fig. 3.87 y 3.88)

Figura 3.88. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 12: Textura Granulada.

Figura 3.89. Conjunto de las 16 imágenes que presentan con mayor probabilidad el aspecto 13: Líneas Sencillas.

3.3.3.13 Aspecto 13: Líneas Sencillas



Línea fina y ordenada que se distribuye por toda la imagen y que resalta limpiamente contra el fondo. Si miramos detenidamente la imagen 8 veremos que presenta aproximadamente un 0.3 de probabilidad del aspecto 3, que se refería a un dibujo figurativo, y el resto de la probabilidad pertenece a las líneas del cuerpo que se extienden por el plano y que presentan un tratamiento más sencillo y esquemático que el rostro. (Fig. 3.89 y 3.90)

Figura 3.90. Conjunto de histogramas de las 16 imágenes que presentan con mayor probabilidad el aspecto 13: Líneas Sencillas.

Figura 3.91. Conjunto de patches de diferentes imágenes del Aspecto 5 (Trazo Texturado), Están asociados a la palabra más probable de este aspecto latente.

3.3.4 Vocabulario Visual de Tàpies

En este apartado adjuntamos una serie de figuras que muestran conjuntos de pequeñas regiones o patches que se corresponden con la zona de 16 x 16 píxeles alrededor del nodo que ha sido utilizado para calcular el descriptor *SIFT* de la región. Estas pequeñas regiones de las imágenes, al ser visualizadas, contribuyen a la comprensión de las características formales del aspecto al que corresponden. Al mirarlas agrupadas se perciben las constantes que han motivado el agrupamiento en la misma palabra visual. Pasamos a comentar los ejemplos con más detenimiento.

La Fig. 3.91 presenta patches pertenecientes a la palabra visual más probable del aspecto 5 (Trazo Texturado) de la colección Tàpies. Los fragmentos asociados a esta palabra visual pertenecen a las imágenes 2 (arriba a la izquierda), 6 (abajo a la izquierda) y 7 (fragmentos color ocre) de la Fig. 3.73. Es especialmente fácil distinguir los que pertenecen a la imagen 7 dado que presentan el color ocre característico de la obra. Se percibe perfectamente el rastro de la pincelada poco densa en todos los fragmentos. El número de fracciones que se muestran de cada imagen corresponde al número de veces que la palabra visual está presente en esa imagen concreta. Puede notarse que la imagen en la que está más presente esta palabra visual es la 7.

En la Fig. 3.92 se muestran fragmentos correspondientes a la palabra más probable del aspecto 3 (Línea Narrativa - Figurativa). Están representadas 12 de las 16 imágenes de la Fig. 3.69. Los que se identifican a simple vista pertenecen a la imagen 4 y se debe a su color anaranjado, así como los pertenecientes a la imagen 3 por tratarse de un gofrado que el ordenador procesa como si se fuesen manchas de gris. Se trata de fragmentos más complejos y elaborados que los de la Fig. 3.91. Poseen líneas mezcladas con tintas planas y se percibe una cierta tendencia diagonal en dirección ascendente de izquierda a derecha. Resulta clara la correspondencia de esta palabra visual con un aspecto más figurativo, con más adornos y ornamentos que el comentado anteriormente.

La Fig. 3.93 muestra fragmentos de palabras visuales pertenecientes a imágenes del as-

Figura 3.92. Conjunto de patches de diferentes imágenes del Aspecto 3 (Línea Narrativa - Figurativa), Están asociados a la palabra más probable de este aspecto latente.

Figura 3.93. Conjunto de patches de diferentes imágenes del Aspecto 10 (Trazo Claro sobre Fondo Oscuro), Están asociados a la palabra más probable de este aspecto.

Figura 3.94. Palabras visuales 39, 81 y 289 de izquierda a derecha constituidas por conjuntos de patches de diferentes imágenes de la colección Tàpies.

pecto 10 (Trazo Claro sobre Fondo Oscuro). Se pueden ver estas imágenes en la Fig. 3.83: los que pertenecen a la segunda imagen tienen algunos rastros de color ocre, así como los que pertenecen a la imagen número 8 tienen restos de color rojo. La palabra visual está constituida por líneas con direccionalidad horizontal, más claras que el fondo y dispuestas muy cerca unas de otras. En este caso también queda patente la forma en que la palabra visual más probable explica las principales cualidades del aspecto 10.

En las Fig. 3.94, 3.95, 3.96 y 3.97 se exponen algunas de las 300 palabras visuales que conforman el vocabulario de la colección Tàpies y que han sido utilizadas para realizar las clasificaciones. En este caso los fragmentos de las distintas imágenes se han situado totalmente enganchados los unos con los otros formando un cuadrado compacto para obtener una versión más descriptiva de la configuración que ha producido su agrupación en la misma clase. El tamaño del cuadrado está directamente relacionado con la presencia de la palabra en el vocabulario. Por ejemplo la palabra xx está mucho más presente que la xx.

La posibilidad de visualizar el vocabulario particular que utiliza una artista plástico al ejecutar sus obras y a la vez de medir la frecuencia del uso de unas palabras sobre otras, resulta muy significativo y de utilidad para la comprensión y el estudio de su producción.

A su vez, la posibilidad de configurar un vocabulario visual complejo más amplio, compuesto por palabras de diversos artistas, es muy sugerente y sería también de gran utilidad como fondo para la creación digital de nuevas posibilidades estéticas.

Figura 3.95. Palabras visuales 49, y 227 de izquierda a derecha constituidas por conjuntos de patches de diferentes imágenes de la colección Tàpies.

Figura 3.96. Palabras visuales 19, 22, 47, 48, 62 y 77 de izquierda a derecha constituidas por conjuntos de patches de diferentes imágenes de la colección Tàpies.

Figura 3.97. Palabras visuales 1, 15 y 2 de izquierda a derecha constituidas por conjuntos de patches de diferentes imágenes de la colección Tàpies. La diferencia de tamaño se debe a que unas palabras están más presentes en la colección que otras.

3.4 DISTANCIA DE BHATTACHARYYA ENTRE DISTRIBUCIONES DE ASPECTOS

En estadística, la distancia de Bhattacharyya (Bhattacharyya, 1943) mide la similitud de dos distribuciones de probabilidad discretas o continuas. Para ampliar detalles sobre el cálculo de esta distancia se puede consultar el apartado 7 del Anexo A.

La ordenación de las 16 imágenes que se ha comentado en el apartado anterior está hecha en base a un rango de mayor a menor probabilidad de contener un aspecto concreto. En este tipo de imágenes abstractas que no se corresponden con categorías que podamos asociar de manera natural como paisajes, bodegones etc, se observó que algunas imágenes, a pesar de no tener en común un aspecto mayoritario, presentaban cierto parecido que se veía correspondido con una distribución de aspectos similar. La proximidad de unas imágenes con otras puede venir propiciada, no solo por compartir el aspecto mayoritario, sino por tener una combinación de distintos aspectos similar.

En nuestro caso estamos tratando con distribuciones de probabilidad discretas y la distancia de Bhattacharyya existente entre los histogramas de frecuencia resultantes del *pLSA* contribuye a encontrar similitudes entre las representaciones de las imágenes como aspectos latentes, a pesar de que los histogramas no tengan el mismo aspecto mayoritario. Esta nueva vista de los resultados contribuye a establecer nuevas relaciones existentes entre las imágenes, porque imágenes que pertenecen a distintas aspectos, pueden finalmente ser muy parecidas.

También en base a esta distancia es posible dibujar un dendograma o esquema gráfico en forma de árbol en el que los datos se van dividiendo hasta el nivel deseado. Como método de aglomeración se ha utilizado el de Ward (Gordon, 1999) y la distancia de Bhattacharyya.

Este tipo de representación permite distinguir de forma clara las relaciones de similitud y de agrupación que se van estableciendo entre las imágenes. (Fig. 3.98). El gráfico es muy informativo pues genera una especie de filogenia formal entre las imágenes que

Figura 3.98. A la derecha, dendograma entre imágenes del conjunto de obras de Tàpies en el eje de abscisas y la distancia de Bhattacharyya entre pares de distribuciones de probabilidad de sus aspectos latentes en el eje de ordenadas.

Figura 3.99. De izquierda a derecha y de arriba a abajo, la imagen 3 pertenece al aspecto 9 y el resto al aspecto 12. La distancia de Bhattacharyya indica que su distribución de frecuencias es cercana y las coloca todas juntas en el grupo 5.

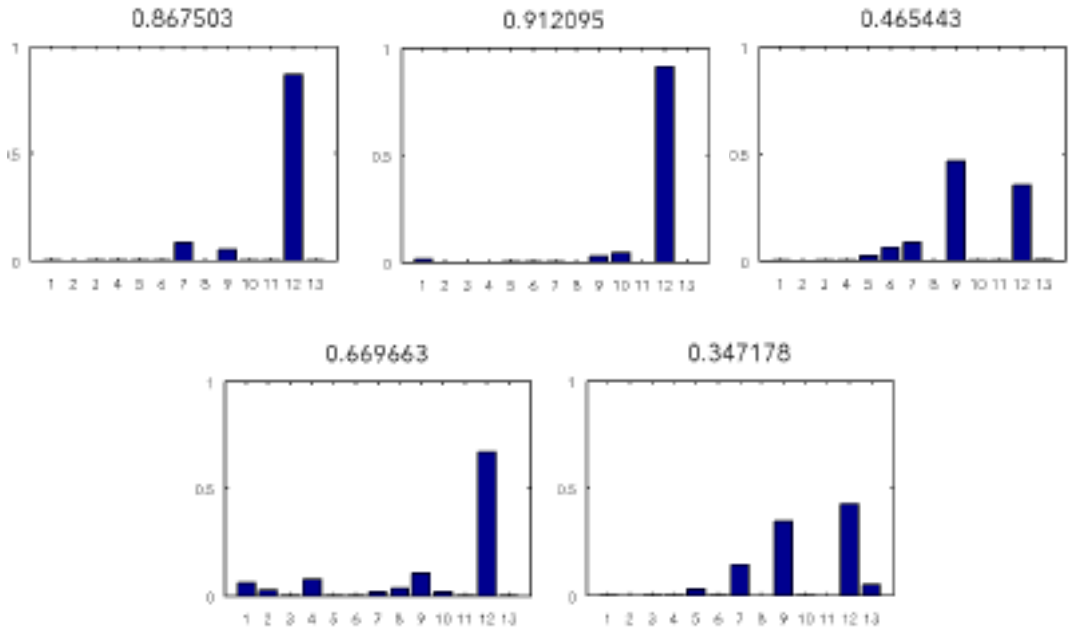


Figura 3.100. Histogramas correspondientes a las imágenes de la Fig. 3.99. De izquierda a derecha y de arriba a abajo, el histograma 3 tiene aomo mayoritario al aspecto 9 y el resto al aspecto 12. La distancia de Bhattacharyya indica que su distribución de frecuencias es cercana y coloca todas las imágenes juntas en el grupo 5.

se puede seguir para entender las posibles relaciones entre categorías.

El dendograma que se muestra se ha cortado en el nivel de distancia marcado en rojo porque a esta altura genera 13 grupos, que equivaldría a un rango parecido a los 13 aspectos seleccionados para trabajar con el *pLSA*. Se podría hacer el corte más arriba o más abajo para realizar un estudio más exhaustivo o más general de la muestra, según se desee.

Para ilustrar mejor esta idea en la figura 3.99 mostramos 5 imágenes y en la figura 3.100 sus 5 histogramas de aspectos correspondientes. La distancia de Bhattacharyya las sitúa dentro del mismo grupo 5, sin embargo la imagen 3 tiene el aspecto 9 como mayoritario.

Todas las imágenes presentan gran similitud, con un fondo suavemente texturado surcado por líneas sutiles y esporádicamente alguna mancha más densa.

En la Fig. 3.101 se muestra otro ejemplo del concepto que se intenta ilustrar: cuatro imágenes con un fondo claro, más homogéneo que el ejemplo anterior, con unos trazos fuertes más oscuros a modo de figura que ejerce el protagonismo de la com-

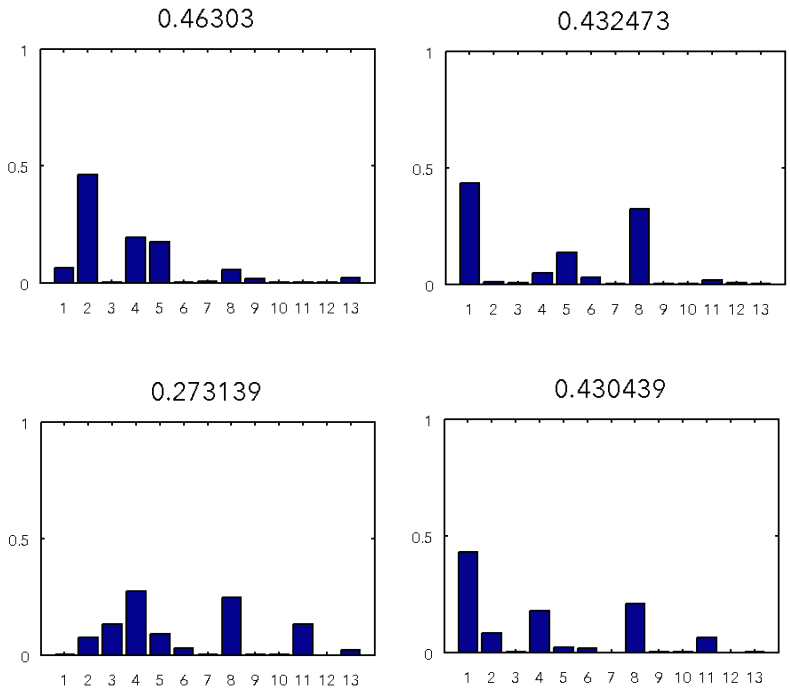


Figura 3.101. De izquierda a derecha y de arriba a abajo, la imagen 1 tiene al aspecto 2 como mayoritario, la imagen 2 y la 4 tienen el aspecto 1 como mayoritario y la imagen 3 tiene el aspecto 4 como mayoritario, sin embargo, la distancia de Bhattacharyya indica que su distribución de frecuencias es cercana y las coloca todas juntas en el grupo 7. En la parte derecha mostramos los respectivos histogramas en el mismo orden.

posición. La similitud de la configuración es evidente a pesar de que cada una presenta mayoritariamente aspectos diferentes.

A continuación mostramos las imágenes de los 13 grupos generados. En las figuras se muestran todas las imágenes que pertenecen al grupo.

Esta nueva aproximación a la colección Tàpies nos indica que hay dos grandes clases: la primera formada por las imágenes que pertenecen a los grupos del 1 al 5, y la segunda formada por el resto. Realizando una descripción muy "a vista de pájaro" y general, la percepción es que la primera gran clase está formada por obras trabajadas a sangre, en las que hay múltiples líneas que se entrecruzan con texturas, manchas y/o granulados ocupando todo el dominio de la imagen. Del grupo 7 en adelante las imágenes se estructuran compositivamente en torno a la relación figura-fondo y se hace más patente la presencia de un elemento predominante que se equilibra con el resto. Presentan superficies diferenciadas que dialogan entre sí, ponderándose, estabilizándose.

Si bajamos un nivel más, observamos que el grupo 1 se compone de obras en las que las líneas y trazos tienen preponderancia sobre el resto, mientras que en los grupos 2,3,4 y 5 todos los elementos están más integrados entre sí. Los grupos 6 y 7 presentan trazos gruesos y gestuales con cierta predilección central, mientras que los grupos 8,9,10,11,12 y 13 tratan conformas geométricas o líneas y elementos más sutiles sobre fondos de tintas planas.

Así se podría continuar bajando de nivel para ir discriminando las dicotomías que ha establecido el dendograma. Este sistema constituye una herramienta visual que el artista o el experto puede utilizar para explorar y analizar la complejidad del conjunto de imágenes objeto de estudio. Facilita la posibilidad de profundizar en su lectura buscando patrones, relaciones entre unas imágenes y otras, o entre diferentes grupos de ellas.

Figura 3.102. Imágenes del Grupo 1 atendiendo a la distancia de Bhattacharyya.

3.4.1 Distancia de Bhattacharyya: Grupo 1

Figura 3.103. Imágenes del Grupo 2 atendiendo a la distancia de Bhattacharyya.

3.4.2 Distancia de Bhattacharyya: Grupo 2

3.4.3 Distancia de Bhattacharyya: Grupo 3

Figura 3.104. Imágenes del Grupo 3 atendiendo a la distancia de Bhattacharyya.

3.4.4 Distancia de Bhattacharyya: Grupo 4

Figura 3.105. Imágenes del Grupo 4 atendiendo a la distancia de Bhattacharyya.

3.4.5 Distancia de Bhattacharyya: Grupo 5

Figura 3.106. Imágenes del Grupo 5 atendiendo a la distancia de Bhattacharyya.

3.4.6 Distancia de Bhattacharyya: Grupo 6

Figura 3.107. Imágenes del Grupo 6 atendiendo a la distancia de Bhattacharyya.

Figura 3.108. Imágenes del Grupo 7 atendiendo a la distancia de Bhattacharyya.

3.4.7 Distancia de Bhattacharyya: Grupo 7

3.4.8 Distancia de Bhattacharyya: Grupo 8

Figura 3.109. Imágenes del Grupo 8 atendiendo a la distancia de Bhattacharyya.

3.4.9 Distancia de Bhattacharyya: Grupo 9

Figura 3.110. Imágenes del Grupo 9 atendiendo a la distancia de Bhattacharyya.

3.4.10 Distancia de Bhattacharyya: Grupo 10

Figura 3.111. Imágenes del Grupo 10 atendiendo a la distancia de Bhattacharyya.

3.4.11 Distancia de Bhattacharyya: Grupo 11

Figura 3.112. Imágenes del Grupo 11 atendiendo a la distancia de Bhattacharyya.

Figura 3.114. Imágenes del Grupo 13 atendiendo a la distancia de Bhattacharyya.

3.4.13 Distancia de Bhattacharyya: Grupo 13

3.4.12 Distancia de Bhattacharyya: Grupo 12

Figura 3.113. Imágenes del Grupo 12 atendiendo a la distancia de Bhattacharyya.

3.5 CLASIFICACIÓN MEDIANTE DESCRIPTORES DE TEXTURA DE HARALICK

En la obra de Tàpies la textura de la propia materia es un elemento fundamental. En sus obras cobra un papel predominante la pasta, el material espeso que él conseguía mezclando pintura al óleo con blanco de España. Sus obras muchas veces se convierten en auténticas arquitecturas que exaltan el trabajo físico de la pintura. En su tarea de pintor, fuerza a la pintura a sus límites para así profundizar en su lenguaje. Centra su atención en los ritmos plásticos, en los gestos, crea tensiones entre las formas.

De sus telas bidimensonales sobresale la materia y los elementos simbólicos que incorpora. Lamentablemente, las imágenes digitales que utilizamos en nuestro análisis no contienen todo este caudal de grafías, rascados, superposiciones que constituyen el alfabeto formal del artista, pero estas texturas tan marcadas generan sombras en la superficie de la obra que si son susceptibles de ser analizadas.

El lenguaje de Tàpies toma gran parte de su significación de la manera esquemática y depurada con la que construye sus obras. En su manera de trabajar, de aparentemente ir acumulando de forma azarosa elementos sobre la tela, hay una organización extremadamente calculada. Las formas dialogan en sus cuadros.

Esta forma de componer, junto con las sombras que generan las texturas en las obras al ser capturada su imagen por una cámara digital, conducen a la elección de descriptores de texturas como candidatos óptimos para realizar un nuevo análisis de la colección.

Para contrastar los resultados obtenidos con los descriptores *SIFT* y para tener un contrapunto, se decide realizar una prueba representando el conjunto de 434 imágenes de Tàpies mediante descriptores de textura de Haralick y realizando la posterior agrupación a través del algoritmo *K-means*. La metodología de cálculo de este tipo de descriptores aparece detallada en el apartado 2.5 de la metodología y en el apartado 2 del Anexo A.

Después de realizar diversas pruebas de agrupación de toda la muestra con 10, 13 y 21 grupos, se decidió que la agrupación en 21 clases era la más significativa y adecuada para contrastar los resultados obtenidos a partir de los descriptores de textura de Haralick con los 13 aspectos latentes previos elaborados a partir del modelo *pLSA* y descriptores *SIFT*.

Los resultados de textura proporcionan menos información que los aspectos latentes ya que el resultado se limita a la obtención de una agrupación de imágenes.

No proporcionan a nivel particular de cada obra ninguna indicación más que la del clúster al que pertenece, por lo que el análisis es mucho menos versátil que el de los aspectos latentes que, además del aspecto más probable, proporcionan la composición relativa del resto de aspectos de cada imagen y permiten así poder profundizar en el estudio de sus particularidades.

No obstante, se ha encontrado correlación con algunas de las agrupaciones de textura y de aspectos latentes. Para discutir estos resultados, junto al panel de las imágenes resultantes de cada grupo de texturas se muestra un histograma que indica; en el eje de ordenadas el número de imágenes del grupo que presentan el aspecto latente que indica el eje de abscisas. Aclaramos que no se tiene en cuenta la probabilidad con que tiene asociado el aspecto, simplemente se contabiliza si la imagen contiene el aspecto. De esta manera se puede comprobar si existe una relación entre el grupo de textura y algún aspecto o combinación determinada de aspectos latentes.

Cada panel de grupos de textura que se muestra contiene todas las imágenes que el clasificador ha considerado que pertenecían a la clase concreta. Por este motivo, en unos paneles hay gran cantidad de imágenes y en otros muy pocas, a diferencia del tratamiento anterior de los aspectos latentes en el que siempre se mostraban las 16 imágenes, por orden de mayor a menor probabilidad, que contenían el aspecto en cuestión.

Este dato si que es informativo, en el sentido de que vemos la distribución global de las obras: los grupos de textura que contienen muchas obras son más representativos de las características formales del conjunto que los grupos con pocas imágenes.

A continuación pasamos a describir los grupos obtenidos que se pueden asociar de manera clara con los aspectos latentes discutidos previamente. Se ha obviado el comentario del resto de grupos por estar constituidos por un gran número de imágenes bastante entrópicas y de difícil examen.

El modelo de descriptores de textura, al ofrecer como resultado únicamente la agrupación, proporciona menos herramientas para el análisis y la comprensión del agrupamiento que los histogramas de distribución de aspectos latentes.

Figura 3.115. Imágenes del Grupo 1 atendiendo a la textura de Haralick.

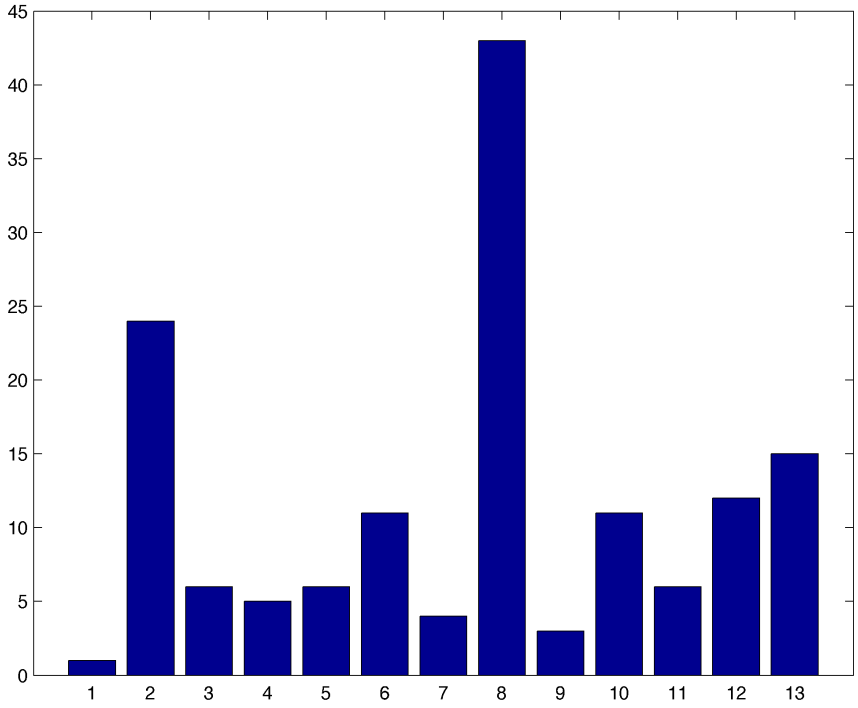


Figura 3.116. Histograma que presenta en el eje de abscisas los aspectos latentes obtenidos con *pLSA* en la colección de imágenes de Tàpies y en el eje de ordenadas el número de imágenes que lo presentan en el grupo 1 de textura de Haralick.

3.5.1 Textura de Haralick: Grupo 1

Este grupo, como vemos en la figura x, lo componen 49 imágenes, de las cuales 43 presentan el aspecto latente 8 (Detalle sobre Fondo Plano) y 24 también presentan el aspecto latente 2 (Figura - Fondo). Es una clase bastante clara.

Son generales las tintas planas con algún motivo;pueden ser detalles sutiles de línea o también pequeños motivos de tinta plana o pincelada densa.

Figura 3.117. Imágenes del Grupo 8 atendiendo a la textura de Haralick.

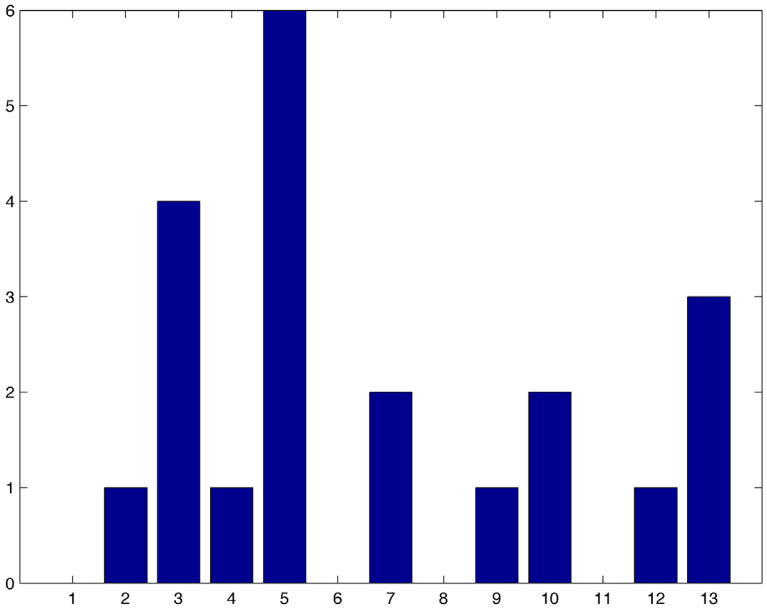


Figura 3.118. Histograma que presenta en el eje de abscisas los aspectos latentes obtenidos con *pLSA* en la colección de imágenes de Tàpies y en el eje de ordenadas el número de imágenes que lo presentan en el grupo 8 de textura de Haralick.

3.5.2 Textura de Haralick: Grupo 8

Esta clase de textura consta de 6 imágenes, todas las cuales poseen el aspecto latente 5 (Trazo Texturado) y 4 de ellas presentan también el aspecto 3 (Linea Narrativa - Figurativa).

Es un grupo muy compacto cuyas obras presentan líneas y trazos texturados mezclados con trazos densos (en algunos casos) que llenan totalmente el plano compositivo. De factura vigorosa, las obras transmiten movimiento, gesto y energía.

Podríamos decir que en este caso la agrupación realizada en función de los descriptores de textura de Haralick presenta una clase más evidente que la que sería equivalente en el aspecto latente 5 Trazo Texturado.

Figura 3.119. Imágenes del Grupo 11 atendiendo a la textura de Haralick.

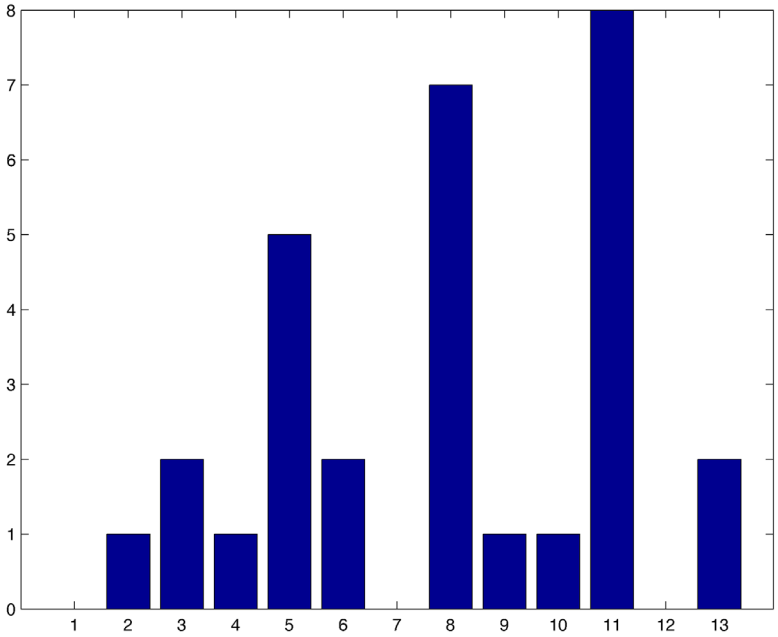


Figura 3.120. Histograma que presenta en el eje de abscisas los aspectos latentes obtenidos con *pLSA* en la colección de imágenes de Tàpies y en el eje de ordenadas el número de imágenes que lo presentan en el grupo 11 de textura de Haralick.

3.5.2 Textura de Haralick: Grupo 11

El grupo 11 de textura está formado por 10 imágenes, 8 de las cuales presentan el aspecto 11 (Equilibrio Compositivo) y 7 el aspecto 8 (Detalle sobre Fondo Plano).

Constituye un conjunto en el que se aprecia una marcada horizontalidad en las composiciones, formada por una línea fina en algún caso o por bandas amplias de tintas planas en la mayoría de los casos.

Presenta gran coherencia visual y aporta nueva información al estudio de la colección Tàpies ya que en los aspectos latentes no ha aparecido ninguna configuración formal con marcada horizontalidad compositiva.

Figura 3.121. Imágenes del Grupo 17 atendiendo a la textura de Haralick.

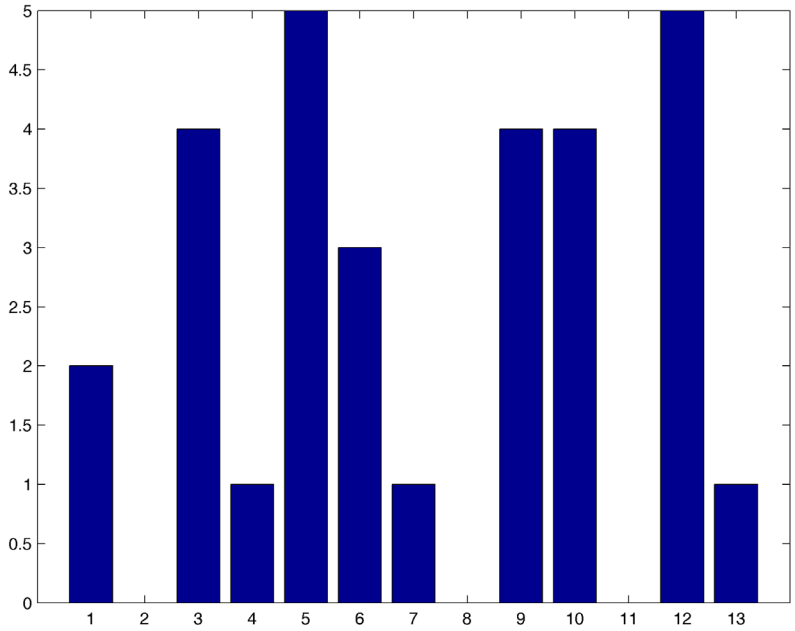


Figura 3.122 Histograma que presenta en el eje de abscisas los aspectos latentes obtenidos con *pLSA* en la colección de imágenes de Tàpies y en el eje de ordenadas el número de imágenes que lo presentan en el grupo 17 de textura de Haralick.

3.5.3 Textura de Haralick: Grupo 17

Grupo con 10 imágenes que presentan los aspectos 5 (Trazo Texturado), 12 (Textura Granulada), 9 (Trazo Oscuro sobre Fondo Claro) y 3 (Línea Narrativa - Figurativa).

Constituye un grupo de gran interés plástico con multitud de líneas de diversas calidades ocupando prácticamente todo el espacio e intercaladas de salpicaduras y texturas. Es curioso constatar que estas obras no contienen los aspectos 2 (Figura - Fondo), ni 8 (Detalle sobre Fondo Plano), ni 11 (Equilibrio Compositivo).

Presenta gran coherencia visual esta categoría que podríamos calificar de “horror vacui”.

Figura 3.123. Imágenes del Grupo 19 atendiendo a la textura de Haralick.

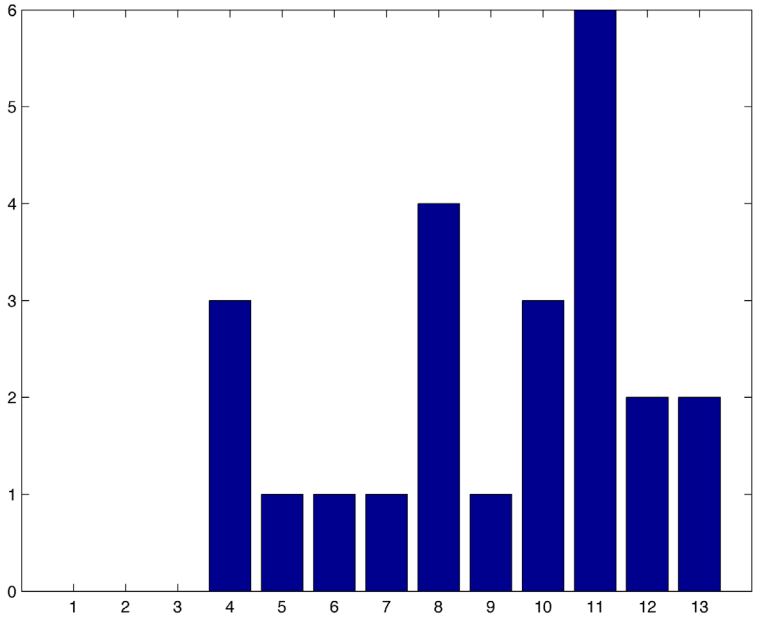


Figura 3.124. Histograma que presenta en el eje de abscisas los aspectos latentes obtenidos con *pLSA* en la colección de imágenes de Tàpies y en el eje de ordenadas el número de imágenes que lo presentan en el grupo 19 de textura de Haralick.

3.5.4 Textura de Haralick: Grupo 19

En este grupo 19 de textura encontramos 8 imágenes de obras que transmiten sensación de medida, equilibrio, composición simétrica.

Esta clase se ajusta bastante al aspecto latente 11 denominado Equilibrio Compositivo, y como se puede comprobar en la figura x, 6 de las 8 imágenes lo contienen.

Otra información que podemos obtener de este gráfico es que ninguna de ellas contienen los aspectos latentes 1, 2 ni 3. Este dato nos puede ayudar a comprender de forma paralela, las cualidades de estos aspectos.

Figura 3.125. Imágenes del Grupo 21 atendiendo a la textura de Haralick.

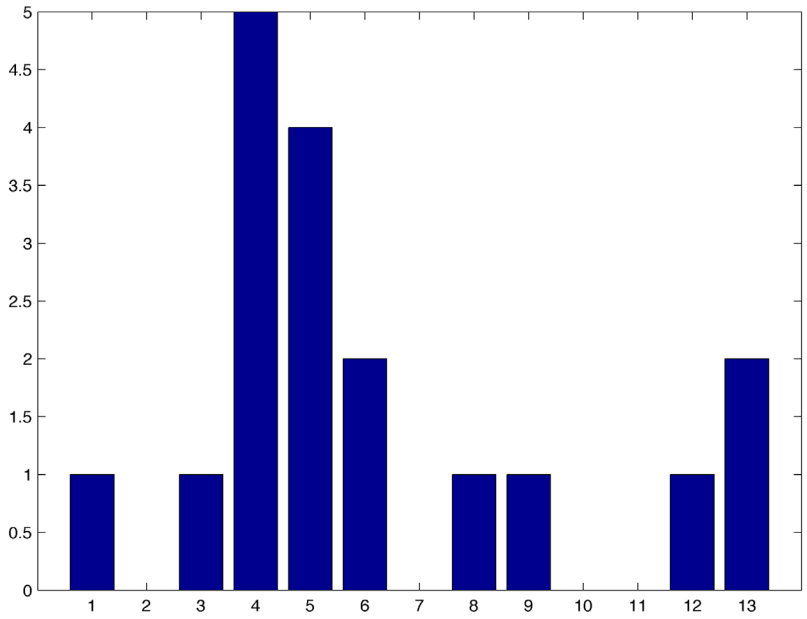


Figura 3.126. Histograma que presenta en el eje de abscisas los aspectos latentes obtenidos con *pLSA* en la colección de imágenes de Tàpies y en el eje de ordenadas el número de imágenes que lo presentan en el grupo 21 de textura de Haralick.

3.5.5 Textura de Haralick: Grupo 21

Este grupo de textura presenta 6 imágenes con unos pocos trazos gestuales densos muy marcados.

En esta clase, 5 de las imágenes poseen el aspecto latente 4 (Trazo Grueso Denso) y 4 de ellas presentan también el aspecto 5 (Trazo Texturado). Se trata así de obras con una pincelada de gesto importante que se impone contra un fondo más neutro.

3.6 DISCUSIÓN DE LOS RESULTADOS

Establecer el criterio de bondad a la hora de calificar los resultados obtenidos que se comentan en el presente capítulo no es tarea fácil, pero es determinante para decidir si el sistema informático programado consigue los objetivos que se persiguen. El motivo, como ya se ha discutido ampliamente a lo largo de la tesis, es que las obras analizadas no contienen objetos o configuraciones a las que se asocie un significado unívoco, se trata de imágenes de contenido no figurativo.

La evaluación estética de las agrupaciones obtenidas ha sido validada por expertos del mundo del arte: catedráticos de pintura y escultura de la facultad de Bellas Artes de la Universidad de Barcelona, expertos de la Fundació Tàpies, artistas visuales y teóricos del arte. Todos ellos han juzgado como significativas las relaciones establecidas por el sistema computacional descrito.

Han considerado una contribución especialmente interesante el hecho de que es posible repetir, bajo el mismo criterio, los análisis en diferentes colecciones de artistas, así como el hecho de que, el análisis es independiente del conocimiento previo que se tiene de las obras, tanto a nivel de la época en la que han sido realizadas como emotivo.

Ha sido especialmente valioso el testimonio del Dr. Miquel Planas, director de la investigación y a la vez artista implicado. Además de facilitar el acceso a su colección de imágenes, en todo momento se ha mostrado abierto y generoso en sus comentarios sobre los estudios realizados sobre sus obras, aportando sugerencias que han ayudado al desarrollo de la investigación.

En vista de los resultados descritos en los apartados anteriores, concluimos que la representación de la imagen mediante aspectos latentes resulta de gran utilidad para el estudio y la comprensión de una extensa colección de imágenes porque proporciona una aproximación al contenido visual basada en el número de aspectos, y estos, al tratarse de varios

ítems de información, ayudan a realizar comparativas entre grupos o imágenes individuales.

Los descriptores de Haralick proporcionan a su vez agrupaciones enriquecedoras, que en muchas ocasiones complementan las evidencias formales extraídas mediante la representación por aspectos latentes. Es especialmente provechoso utilizar la comparativa de ambas informaciones para perfeccionar la interpretación de las constantes visuales compositivas, de textura, de trazo etc, que se pueden valorar en la colección de imágenes.

La representación de aspectos latentes facilita además la comprensión de las imágenes a nivel individual y en grupo, mientras que la clasificación en base a la textura de Haralick, proporciona información de las clases que establecen, pero no a nivel individual de cada imagen. Esta seria una ventaja que poseerían los aspectos latentes sobre los descriptores de Haralick, aunque lo ideal es utilizar todas las herramientas disponibles para sacar el máximo partido al conocimiento de las configuraciones estéticas de nuestra colección.

Por otro lado, las palabras visuales que se han puesto en evidencia configurando los fragmentos de las imágenes que las poseen, constituyen elementos valiosos para utilizarlos en la creación artística, además de proporcionar señales evidentes de las características formales que han servido para llevar a cabo las agrupaciones. Podrían ser utilizados como punto de partida para sistemas informáticos que tengan por objeto el diseño de nuevas posibilidades estéticas.

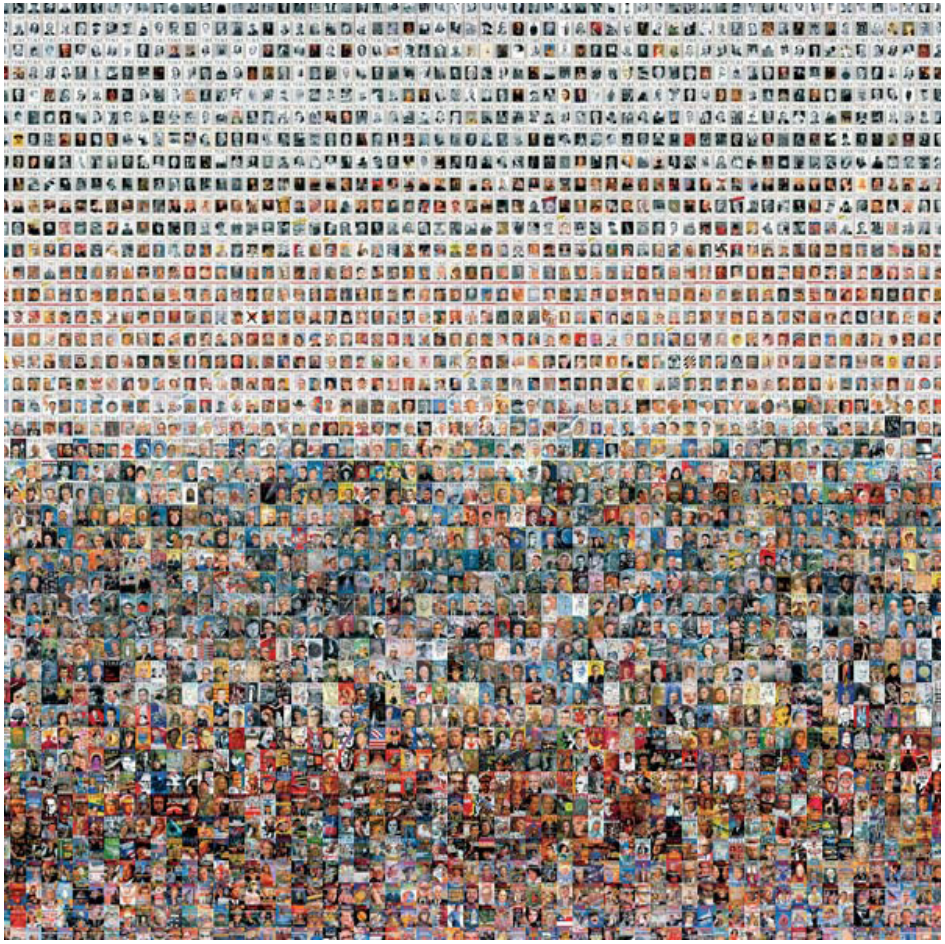


Figura 4.1. Mapping Time. Jeremy Douglass / Lev Manovich, 2009. Las portadas de todos los números de la revista Time publicados entre 1923 y el verano de 2009. Número total de portadas: 4535. Las imágenes están dispuestas de acuerdo con el orden de publicación (de izquierda a derecha y de arriba a abajo) (Manovich, 2012)

4 - CONCLUSIONES

CONCLUSIONES

4.1 APORTACIONES

En nuestro tiempo ha dejado de ser un problema el almacenaje de la información. Disponemos de cantidades ingentes de contenidos, incluso a nivel doméstico, en nuestras casas, guardados en los dispositivos electrónicos que tenemos a nuestro alcance. El verdadero problema es cómo acceder a ellos, pues dependemos de la mirada parcial que nos proporcione el índice de acceso que utilicemos. La visión por computador o artificial, dadas las limitaciones de lo humano, es nuestra única respuesta y por eso se invierten importantes esfuerzos en esta dirección. La presente tesis indaga en los beneficios que puede obtener el mundo del arte de estas investigaciones. Ha concretado su estudio en unos modelos de visión artificial que, de alguna manera, emulan los estadios iniciales de la percepción visual humana y los ha utilizado para categorizar grandes conjuntos de imágenes de obra de artista abstracta. Los resultados obtenidos son considerados satisfactorios por expertos en arte y, lejos de pretender substituir el criterio de los entendidos, el sistema programado propone una herramienta de estudio para establecer analogías y buscar aspectos latentes en grandes colecciones de imágenes. El sistema permite repetir los estudios sobre diferentes periodos del mismo artista, o sobre colecciones de distintos artistas o épocas, con los mismos criterios. De esta forma, los resultados obtenidos se pueden comparar sin riesgo de caer en interpretaciones subjetivas condicionadas por las preferencias o conocimientos previos.

Este tipo de estudios sobre grandes colecciones resulta imposible realizarlo de otra forma por las evidentes limitaciones de accesibilidad a las obras y de capacidad humana de análisis. Y, aunque toda categorización es por definición subjetiva al estar sustentada en un determinado criterio, la tecnología digital nos ofrece esta forma de aproximarnos a una información que de otro modo quedaría imposibilitada por la magnitud de contenido visual almacenado.

> La forma tradicional de visualizar la información ya no es válida. Necesitamos técnicas que nos permitan observar los vastos universos de los media para poder detectar rápidamente multitud de patrones de interés. Estas técnicas tienen que ser compatibles con la capacidad de procesamiento de información del ser humano y, al mismo tiempo, conservar una cantidad suficiente de detalles de las imágenes originales, video, audio o experiencias interactivas para permitir su estudio (Manovich, 2012, p. 1).

Cabe destacar el interés de las herramientas presentadas desde el punto de vista del acceso simultáneo por parte de un artista a su colección de múltiples imágenes para poder analizar su trayectoria creativa, o desde el punto de vista de los teóricos del arte que podrían realizar estudios comparativos entre las obras de arte de diferentes artistas o épocas sin necesidad de mover de su emplazamiento ni una obra. Sin entrar a valorar la calidad estética de las agrupaciones que realiza la maquina, podemos concluir que las relaciones que establece, dada la cualidad matemática que le confiere la metodología utilizada para su realización, proporcionan nuevos puntos de vista libres de preconcepciones historicistas o vivenciales.

A continuación se detalla una relación de las principales aportaciones de la tesis presentada en esta memoria:

4.1.1 Desarrollo de programas informáticos en Matlab extensivos a otras colecciones

Al tratarse de una tesis eminentemente metodológica, la principal aportación consiste en poner a disposición del ámbito artístico una metodología procedente de la inteligencia artificial, que está siendo puesta a prueba en la actualidad en multitud de otros campos del saber con buenos resultados, y que también, como se ha visto, ha demostrado óptimos rendimientos en el descubrimiento de analogías visuales sobre imágenes artísticas abstractas. La dificultad de permitir la ósmosis interdisciplinar se ve superada con este tipo de herramientas que se apropian de modelos de la ingeniería para dar respuesta a cuestiones planteadas desde el arte; de este modo las herramientas se modelan según las necesida-

des y puntos de vista particulares de la disciplina.

Con este objetivo se han desarrollado 4 programas en el lenguaje Matlab que permitirían extender la aplicación de los algoritmos de visión artificial comentados a cualquier colección de imágenes digitales para su estudio: ya se trate de artistas abstractos de la misma o diferente época para realizar comparativas o también de artistas figurativos o de cualquier tipo de imágenes que se deseen estudiar bajo el prisma descrito en la presente tesis. Los programas serían los siguientes:

4.1.1.1 Programa de aprendizaje supervisado discriminativo

El esquema conceptual puede verse en la Fig. 3.1. Este sistema permite buscar y recuperar de una gran colección de imágenes, las que pertenecen a las categorías formales con las que previamente hemos entrenado al sistema. Este entrenamiento se puede llevar a cabo con un mínimo de 15 imágenes de cada clase, etiquetadas de forma manual, y se obtendría un porcentaje de acierto medio del 70 %. Este acierto se refiere a que, del total de imágenes problema que se le presenten al sistema, el 70% será clasificado correctamente en la clase a la que pertenece. Respecto a la velocidad de acceso; sobre una base de datos de 3000 imágenes con 10 clases definidas, el tiempo de recuperación es de aproximadamente 10 minutos, por lo que constituye una herramienta de gran eficacia para permitir, tanto al artista como a los estudiosos del arte, el acceso a grandes colecciones en busca de unas configuraciones formales determinadas que vendrían condicionadas por el entrenamiento previo al que se sometiese al sistema, y que se podrían modificar para cada acceso.

4.1.1.2 Programa de aprendizaje no supervisado generativo

El esquema conceptual puede verse en la Fig. 3.13. Este programa tiene como información de entrada una carpeta con todas las imágenes que se desea estudiar. Sin ningún tipo

de indicación ni anotación externa, el sistema será capaz de agruparlas en función de las constantes o configuraciones visuales que comparten, proporcionando; por un lado una preclasificación formal interesante que puede ser de gran utilidad para tener una idea aproximada del contenido del conjunto, y por otro, aportando un punto de vista auxiliar novedoso y libre de preconcepciones estéticas condicionadas por lo que ya se sabe o se conoce acerca de la muestra de estudio. De esta forma el ordenador pone en evidencia el correlato formal de las intuiciones creativas de un artista para que él, inmerso en su proceso creativo, tome decisiones al respecto. El método así mismo permite desvelar, evidenciar constantes en la obra del creador, aspectos latentes que pueden relacionarse con las obras del mismo artista pertenecientes a otros periodos o también con obras de otros artistas.

En este programa se ha implementado además el cálculo del índice de entropía de Shannon y la posibilidad de visualizar la colección de imágenes tratadas en función de este cálculo. La metodología *pLSA* (ver apartado 4 del Anexo A) proporciona una distribución de probabilidad de determinados aspectos visuales en las imágenes (Fig. 3.19 y 3.20). La contemplación de las obras de arte así representadas nos proporciona información más esquemática que la propia imagen y nos brinda la oportunidad de establecer nuevas relaciones visuales y agrupaciones que favorecen el establecimiento de analogías. Por otro lado, las imágenes calificadas como poco entrópicas en función del índice de Shannon, al tener un solo aspecto representado con mucha probabilidad, ayudan a calificar al aspecto que contienen y con ello simplifican el estudio del resto de imágenes llamadas más entrópicas. Por así decirlo, las imágenes poco entrópicas contribuyen a esclarecer las características formales que definen al aspecto en concreto y esto ayuda a la comprensión del total de la colección. En las imágenes más entrópicas todos los aspectos están representados en mayor o menor medida y sin un índice previo como el que constituyen las imágenes poco entrópicas, resulta complicado determinar en que fragmento concreto de la imagen se presenta un aspecto.

Así pues, la posibilidad de calcular y visualizar las imágenes en función de su índice de Shannon constituye una herramienta de gran utilidad a la hora de analizar el conjunto de obras.

4.1.1.3 Programa de cálculo de distancias y elaboración del dendograma

Este programa calcula la distancia de Bhattacharyya entre los histogramas de frecuencias de las representaciones de las imágenes como probabilidades de aspectos latentes. (Fig. 2. 15). A partir de este cálculo se puede dibujar el dendograma que visualiza las relaciones entre las imágenes estudiadas en función de esta distancia. (Fig. 3.98)

Este gráfico posee un notable interés a nivel informativo ya que constituye una herramienta visual que el artista o el experto puede utilizar para explorar y analizar la complejidad del conjunto de imágenes objeto de estudio, o sea, el total de la obra de artista. El dendograma ofrece la posibilidad de profundizar en su lectura buscando patrones, relaciones entre unas imágenes y otras, o entre diferentes grupos de ellas, así se puede extraer información que enriquezca el conocimiento de la colección de imágenes. Este gráfico no es sólo para ser visto sino que su función es ser leído con el objetivo de ayudar a la comprensión de los datos que presenta.

Permite fácilmente establecer relaciones de parentescos formales entre imágenes y grupos de imágenes. Este gráfico así elaborado facilita:

1. Muestra el parentesco formal entre las imágenes basado en la distancia calculada.
2. Permite la comparación de manera fácil y rápida entre unas imágenes y otras.
3. Ayuda a su clasificación.

4.1.1.4 Programa de agrupación en base a descriptores de textura

La textura en pintura constituye un recurso expresivo importante. Se refiere a la agregación de materiales que aportan variaciones o irregularidades en la superficie de la obra. Estas texturas aumentan las sensaciones que transmite el lienzo o el papel, dan forma y volumen a las creaciones artísticas. Forma parte del lenguaje de la pintura y comienza por la textura

que aporta el propio lienzo, la calidad del hilo, el entramado. En la pintura, para crear efectos de textura, se utilizan desde medios como tierras, arena y pigmentos gruesos mezclados con el óleo, hasta emulsiones de cera, objetos de papel, cartón, madera.

En visión por computador, la textura es una de las características importantes utilizadas para la identificación de objetos o de zonas de interés en una imagen. El programa que se presenta utiliza los descriptores de textura definidos por Haralick (Haralick, 1973) para agrupar el conjunto de imágenes de entrada. También constituye una herramienta de gran utilidad para proporcionar el análisis visual en base a las relaciones de textura entre las distintas imágenes y puede servir de contrapunto para realizar comparaciones respecto a los resultados obtenidos con los descriptores *SIFT* utilizados en el resto de programas.

Los programas desarrollados en esta tesis constituyen una importante aportación al ámbito de las humanidades digitales, en el que se están invirtiendo grandes esfuerzos durante los últimos años[1]. Este área de estudio aplica los conocimientos de las nuevas tecnologías informáticas a la resolución de los problemas clásicos del campo de las humanidades. En palabras de Sandra Álvaro (2013):

> Las técnicas computacionales no son solo un instrumento al servicio de los métodos tradicionales, sino que tienen un efecto en todos los aspectos de las disciplinas. No solo introduciendo nuevos métodos dirigidos a la identificación de nuevos patrones en los datos, que van más allá de la narrativa y comprensión tradicionales, sino permitiendo la modularización y recombinación de las disciplinas, más allá del ambiente académico tradicional. La aplicación de la automatización ligada a la digitalización no solo ofrece nuevas capacidades de análisis de los documentos textuales, sino que también da lugar a nuevas capacidades de recombinación y producción de conocimiento, así como al surgimiento de nuevas plataformas, esferas públicas, donde la distribución de la información ya no puede pensarse de modo independiente a su producción.

Los inicios se remontan a 1987 con la creación de la Text Encoding Iniciative cuya prime-

1 ADHO: Alliance of Digital Humanities Organization, organización que promueve y apoya la investigación y enseñanza digital en todas las disciplinas de las artes y las humanidades. http://adho.org

ra aportación fueron un conjunto de recomendaciones para codificar textos electrónicos (Hockey, 2004). Existe numerosa literatura, que se ha ido enumerando a lo largo del discurso de esta memoria, sobre las metodologías utilizadas en la tesis aplicadas a la recuperación y filtrado de textos, imágenes de escenas naturales e incluso sonidos, pero no se han encontrado referencias de aplicación de estos modelos con el objeto de recuperar o estudiar imágenes digitales pertenecientes a obras de arte abstracto. Por lo tanto el presente estudio contribuye notablemente en el sentido de que valida la solidez de estos protocolos cuando se enfrentan a contenidos del tipo abstracto que presentamos, dado que los resultados han sido considerados como significativos por los expertos en arte consultados.

4.1.2 Extensión de aplicación del modelo *BoW* al análisis de arte abstracto

Las conclusiones de la presente tesis están fundamentadas en lo que han sido capaces de ver los investigadores. El grado de consecución de los objetivos propuestos necesariamente tendrá que ser evaluado por el lector en función de lo que haya sido visto en las figuras presentadas en la tesis. El punto de partida de la búsqueda son las obras analizadas cuyo aval consiste en contener configuraciones valiosas para los artistas quienes, en su momento, vieron en ellas aspectos de interés.

La imagen del mundo de todos ellos a su vez vendrá determinada esencialmente por la vivencia de su vista. A pesar de que existen muchos tipos de percepción, la importancia de lo visual sigue siendo un hecho de validez general para el hombre. Sin embargo, la retina del ojo humano sólo capta con precisión una pequeña zona en el centro de nuestro campo visual. Por este motivo, si queremos contemplar algo con exactitud es necesario que enfoquemos exactamente el objeto. Esto también ocurre con la sensibilidad cromática, que va desapareciendo a medida que aumenta la distancia con respecto al centro del punto visual, hasta desaparecer por completo cuando se forma un ángulo de 80 grados. El hecho de que no seamos conscientes de la ausencia de sensibilidad cromática en las zonas límite de nuestro campo visual constituye un buen ejemplo de hasta que punto interviene la psique en el proceso de la visión.

Si los elementos a evaluar constituyesen conceptos culturalmente asumidos desde la infancia, como paisajes, ciudades, coches, etc, aún albergaríamos la esperanza de que todos los espectadores compartiesen una idea común, de alguna manera acordada. Pero el hecho de que se trate de imágenes abstractas augura con bastante seguridad que no se habrá producido una experiencia visual unívoca en todos los lectores del documento; no habrá visto lo mismo un ingeniero, que un biólogo o un artista en los resultados propuestos. De hecho, al presentar este trabajo en congresos del ámbito de la visión artificial ya se ha planteado esta cuestión, pues allí donde un artista ve un ritmo compositivo interesante y expresivo, un ingeniero puede ver una textura o un adorno de estampado.

Claro está que, como nos dice Arnheim (1980), en presencia del arte lo ideal es "ilustrar y enriquecer nuestra vida a través de la vista y el oído, no analizar minuciosamente los medios formales por los que se efectúan tal ilustración y enriquecimiento" (p. 24). La tarea del genuino espectador consiste en entregarse a la emoción y el estremecimiento que le proporciona la obra de arte. Pero no puede ocurrir así en el caso del propio creador ni del estudioso para los que es de vital importancia elaborar una interpretación que le permita abrir la vista a los mensajes transmitidos por la forma. En la tesis que nos ocupa se ha puesto a prueba la herramienta informática con conjuntos de imágenes de obras abstractas para observar el tipo de analogías que establece. Aclaramos que con el termino "abstracto" no nos referimos a la extracción de un modelo común, sino al arte que no intenta imitar un modelo conocido, o sea, "no objetivo":

> Los esquemas del arte "no objetivo", considerados desde el punto de vista del mundo de las cosas naturales, son extremadamente abstractos. Reducen la representación de la realidad a un equivalente visual de las fuerzas físicas y psicológicas universales que están en la base de la naturaleza y la vida, y de su interacción. Expresan de esta manera armonía y disonancia, dominio y coordinación, contraste y semejanza, movimiento y reposo, equilibrio y desequilibrio. Sin embargo, desde el otro punto de vista, es decir, desde el punto de vista de la forma, los esquemas básicos, no objetivos, no son abstractos. Son los elementos mismos de la comprensión visual, el material de construcción de la composición que el artista crea para representar la estructura del mundo de la manera en que su temperamento le hace verlo (Arnheim, 1980, p. 45).

La dificultad interpretativa de los resultado en esta tipología de obra es mayor ya que se trata de evaluar objetos o escenas concretos conocidos, sino valores compositivos, estructurales y características formales.

En vista de los resultados obtenidos concluimos que el sistema permitiría, dada una colección con gran cantidad de imágenes, realizar una importante preselección formal agrupándolas de forma más objetiva dado que la mirada artificial no está tan sujeta y condicionada por la percepción conceptual humana. Así, un proceso de análisis más objetivo, como el que aquí se presenta, aportaría nuevos principios de agrupación al conjunto. Estos resultados amplían los obtenidos por otros autores en colecciones de imágenes de escenas naturales aplicando el modelo *BoW* y *pLSA*. No hemos encontrado antecedentes de su aplicación en obras abstractas y esta evaluación positiva de su uso contribuiría a acreditar la idoneidad del sistema.

4.1.3 Aproximación matemática al arte y a las formas

"Las máquinas computadoras son fundamentalmente aparatos para registrar números, que operan con números y dan resultados en forma de números" (Wiener, 1998, p. 159). Simplificando mucho el proceso, el objetivo de esta tesis ha sido reducir a números las características visuales de un conjunto de imágenes de artista para después calcular relaciones de similitud entre ellas y así poder establecer analogías formales. El placer estético está asociado con el valor de ciertos números, pero la aportación de la matemática en la presente tesis no es la definición de la belleza a través de una fórmula, sino el establecimiento de relaciones de similitud entre las imágenes de las obras analizadas.

Como apunta Emmer (2005) en su análisis de las relaciones entre arte y matemática, todo el mundo puede mirar una obra de arte o escuchar una sinfonía y emitir un juicio, aunque subjetivo. Pero para juzgar si la matemática posee una belleza propia es imprescindible poseer cierta cultura en esta disciplina; para dominar las ideas matemáticas se necesitan años de estudio y no existe ningún atajo que abrevie materialmente este proceso. En este

sentido, las relaciones de semejanza entre las imágenes de obras de arte que establecen los diferentes sistemas definidos en esta tesis, permiten hacer más inteligibles las cualidades numéricas que las producen. La matemática ejerce de comisaria de exposiciones y es la responsable de establecer el criterio del discurso que se genera: el de la proximidad formal entre las obras. Las analogías formales que se ponen de manifiesto en las colecciones, también ponen en evidencia de manera visual sus distancias matemáticas.

4.2 PROPUESTAS PARA FUTURAS APLICACIONES

4.2.1 Construcción de un Vocabulario visual

El model *BoW* permite visualizar los fragmentos de imágenes que corresponden a las palabras visuales estudiadas. Esta facilidad, aplicada a diferentes colecciones de diferentes artistas, permitiría construir una base de datos a modo de diccionario visual que quedase a disposición de artistas, teóricos y críticos de arte para su consulta. Constituiría así un recurso de utilidad pública tanto para el estudio del arte como para la creación visual.

También el artista podría construir su vocabulario propio, que iría enriqueciendo al alimentar el sistema con nuevas imágenes; las obras que se vayan generando a posteriori se lucrarán de estos resultados, se retroalimentarán con nuevas gradaciones. Esto será de gran ayuda en el proceso creativo, analítico, taxonómico y pedagógico de la obra.

4.2.2 Aplicaciones en la creación artística

Esta nueva forma de analizar información masiva proporciona a su vez formas alternativas de generar propuestas creativas, como ya apuntó Manovich en 2005 con relación a los

nuevos medios. Disponer de la información del conjunto de imágenes en código digital proporciona las siguientes ventajas:

1- *Representación numérica*: La información que contienen las imágenes se puede describir matemáticamente y puede someterse a manipulación algorítmica, lo que implica que es programable.

2- *Modularidad*: Las palabras visuales obtenidas son un conjunto de píxeles con una identidad, pero que pueden someterse a operaciones de combinación o ensamblaje entre sí y/o con otros elementos digitales para obtener nuevas propuestas creativas.

3- *Automatización*: la representación numérica y la modularidad permiten automatizar muchas operaciones en la creación, manipulación y acceso de nuevas obras.

4- *Variabilidad:* La automatización y modularidad permiten la mutación y la adaptación a las necesidades del usuario, así como su interacción. La utilización de los aspectos formales detectados mediante visión artificial en la producción artística junto con el arte evolutivo serán muy útiles para explorar las posibilidades que ofrecen las nuevas tecnologías al enriquecer el proceso creativo con la introducción de criterios computacionales en la generación de posibilidades expresivas.

4.2.3 Aplicaciones en la enseñanza artística

La experiencia de análisis de las imágenes presentadas por grupos es bien curiosa; cuando aparecen ante nuestra vista por primera vez, pueden dejarnos expectantes. Es conveniente en primer lugar haber visualizado todos los grupos para, a partir de la segunda ocasión y en adelante cuando ya teniendo conciencia del total, percibir de inmediato que existen unos ritmos, unas características comunes a todas ellas, pero, si se fija la atención en concreto en alguna de ellas o si se intentan traducir a palabras textuales las cualidades generales percibidas en cada clase, la percepción de conjunto se desvanece. Se podría

utilizar un paralelismo referido a la profundidad de campo en una cámara fotográfica; si el objetivo enfoca un punto, se desenfoca el resto, para tener una visión de conjunto es necesario mirarlas todas a la vez, y de esta forma sí, existe una intuición estética, una analogía de sentido que las atrae entre sí y las vincula en una misma clase. La clase en sí misma forma un corpus contenedor de un sentido que se pierde cuando se focaliza la atención en sólo una de las obras. Si un artista sólo tuviese ocasión de realizar una obra perderíamos la ocasión de percibir los potenciales valores contenidos en sus tanteos, exploraciones y decisiones. Es el conjunto de la obra de un artista consolidado el que nos abre la puerta a la visión de su núcleo íntimo de percepción. Y, a pesar de que toda clasificación por definición es subjetiva ya que se realiza según una determinada directiva, cuando podemos colocar unas obras próximas a otras en nuestro campo visual adquieren juntas un sentido que no poseen por separado.

Este tipo de asociaciones visuales son absolutamente imprescindibles en la enseñanza artística. Basta presenciar alguna clase en alguna institución superior de arte en la que se pretenda transmitir al alumnado ideas acerca de la composición, la armonía o el equilibrio de un conjunto de formas, para percibir lo esencial e informativo que resulta asociar una imagen con otras para que se pueda percibir su potencial por contraposición; ya sea por imágenes que compartan rasgos comunes o todo lo contrario. Así que desde este punto de vista, la herramienta presentada en esta tesis contribuye notablemente a desvelar el potencial de conocimiento que se puede encontrar en una colección de imágenes.

La visión no funciona de un modo linealmente consecuente y el núcleo de la enseñanza artística lo constituye la enseñanza de una serie de valores que no se pueden transmitir por la razón. En el ámbito académico este sistema puede ser de gran utilidad como recurso pedagógico; El docente y el estudiante pueden disponer de una herramienta de trabajo que les permita establecer unos criterios de análisis formales, planteados de manera objetiva, en un conjunto de imágenes o de obras. Este recurso aumentaría la posibilidad de conectar y de establecer relaciones entre imágenes de distintos estilos, épocas, temáticas, etc., ya que se convierte en un mecanismo autónomo y no dependiente de los hechos habitualmente tratados de carácter más historicista.

La tesis que se presenta puede contribuir muy favorablemente al desarrollo del conocimiento científico del ámbito relacionado con el arte. Una de las contribuciones más directas en la que se podría poner en práctica sería en las tesis doctorales dirigidas por los profesores pertenecientes a las áreas de conocimiento de Historia y Arte, favoreciendo sus investigaciones, ya que, exceptuando aquellas centradas en aspectos más técnicos, la mayoría se orientan hacia la investigación más iconográfica, lo que hace realmente necesario una metodología para la catalogación de obras de arte verdaderamente contrastada.

4.2.4 Aplicaciones en la museística

El sistema desarrollado puede constituir una pieza clave de un sistema CBIR (sistema de búsqueda para recuperar imágenes basándose en su contenido) que permitiría una aproximación nueva y diferente hacia instituciones artísticas. El espectador podría fijar su atención en una obra concreta y con un dispositivo móvil capturar la imagen, enviando ésta al sistema que sería capaz de acceder a una gran base de datos de imágenes para mostrarle aquellas de la colección que compartiesen con ella aspectos semánticos. El conocimiento y la interpretación no vendrían sólo marcados por los comisarios de las exposiciones o los responsables de dichos centros culturales, sino también por los mismos visitantes, que hacen sus lecturas e interpretaciones. También se pretende facilitar una difusión de las lecturas personales mediante las nuevas herramientas de comunicación a través de Internet, lo que puede replantear muchas de las normativas museísticas actuales, especialmente en lo que se refiere a la reproducción y difusión de las obras artísticas expuestas o en depósito.

Ante esta evidencia, los resultados que presentamos dinamizan la difusión y el conocimiento del arte, e intentan hacerlo de una manera más atractiva a través de la aproximación participativa del consumidor. En la mayoría de las instituciones artísticas ya sean galerías de arte, museos, pinacotecas, centros artísticos, centros históricos y salas de arte, habitualmente los visitantes se limitan a seguir un recorrido pautado contemplando las obras expuestas. De forma opcional, la visita puede ser acompañada de audio-guía o guía personal que complementa el recorrido visual e introduce al espectador en el universo contempla-

do, generalmente desde una vertiente historicista, y/o acompañado de unas referencias técnicas y estilísticas.

Otro recurso clásico de transmisión del conocimiento de las obras es la palabra escrita en las cartelas y paneles informativos generales. Este sistema, a nuestro entender obsoleto, limita las fuentes documentales de carácter gráfico porque la información se transmite casi exclusivamente a través de la lectura o, como hemos comentado anteriormente, complementada oralmente con ayuda de las audio guías. Con estos formatos y canales predominantes, la información es sin ningún tipo de duda unidireccional y, a menudo, aburrida y poco significativa.

Partimos de la creencia de que si conseguimos hacer más interesante la contemplación de las obras expuestas en las instituciones citadas, probablemente aumente la afluencia de público y su grado de satisfacción. Para ello creemos importante reforzar el rol participativo del visitante. La interacción entre el visitante y la obra artística, generará feedbacks que llevarán a una mayor implicación e interés: El espectador, al intervenir en el proceso, adopta un cierto rol creativo, es un elemento activador.

Al ser un sistema creativo, generador de información, aporta nuevos conocimientos e interpretaciones y contribuye a establecer nuevas posibilidades o registros. Partimos de la idea de que al interactuar con las imágenes, salimos de la mirada arquetipo de la obra de arte, generando percepciones más abstractas e, incluso, inéditas de esa obra. El espectador será el que dirigirá su mirada y definirá los parámetros de la misma. Ya no será la mirada "oficial" o historicista o académica establecida. Entrará en juego una actitud lúdica, creativa, activa y expectativa del visitante.

4.2.5 Aplicaciones en psicología

Los programas desarrollados en la presente tesis, al modelar matemáticamente algunos procesos de percepción visual, también permiten simular los procesos de estas capacida-

des visuales para su estudio. Así, el modelo descrito podría utilizarse en estudios de psicología que pusiesen a prueba la convergencia de la categorización de imágenes llevada a cabo por el sistema computacional con la realizada por individuos seleccionados para este propósito, de forma que los resultados obtenidos pudiesen arrojar alguna luz respecto al modo en que un ser humano realiza este tipo de tareas de clasificación.

Se han realizado contactos con grupos de psicología básica respecto a este tipo de experimentos y han manifestado su interés en realizar estas pruebas con baterías de test especialmente diseñadas para este fin:

1 - Si los humanos realizan la misma clasificación que la metodología, los resultados reforzarían la plausibilidad de los fundamentos de visión artificial utilizados.

2 - Se podrían realizar también estudios comparativos entre individuos sin un entrenamiento específico en análisis de imágenes y artistas, y analizar las diferencias.

4.3 DIFUSIÓN DE RESULTADOS

Las siguientes publicaciones son una consecuencia directa de la investigación llevada a cabo durante la elaboración de la tesis, y dan una idea de la progresión que se ha logrado. Se detallan publicaciones tanto en el ámbito artístico como el de la visión por computador.

4.3.1 Artículos en Revistas Científicas

Rosado, P., Reverter, F., Figueras, E. & Planas, M.A. (2014). Semantic-Based Image Analysis with the Goal of Assisting Artistic Creation. *Lecture Notes in Computer Science. 8671*, p. 526 - 533. Doi: 10.1007/978-3-319-11331-9

Rosado, P., Figueras, E. & Reverter, F. (2014). Intersecciones entre visión artificial y mirada artística. *BRAC-Barcelona, Research, Art, Creation. 2(1)*, p. 1 - 54. Hipatia Press, 2014. Doi:10.4471/brac.2014.01

Reverter, F., Rosado, P., Figueras, E. & Planas, M.A. (2012). Artistic ideation based on computer vision methods. Journal of Theoretical and Applied Computer Science. 6(2), p. 72 - 78. http://www.jtacs.org>.

4.3.2 Ponencias en Congresos Internacionales

Reverter, F., Rosado, P., Figueras, E. & Planas, M.A. (2012). Art images classification using Bag-of-Visualterms representation. Proceedings of Postdigital Art (CAC3). París (França). p.157-160. http://postdigital.eu/program

Rosado, P., Reverter, F., Figueras, E. & Planas, M. (2014). Semantic-Based Image Analysis with the Goal of Assisting Artistic Creation. *In proceedings International Conference on Computer Vision and Graphics 2014 (ICCVG),* Warsaw, Poland, Setember 15-17. Editors: Springer, p. 526 - 533. Doi: 10.1007/978-3-319-11331-9. ccvg.wzim.sggw.pl/default.asp

Rosado, P & Reverter, F. (2015). La mirada in silico. *III Simposio Internacional de la Tipografia al Libro-Arte 2015,* Ciudad Juarez, México, 27 - 29 de Abril. http://3ersimposio.esy.es/index.html

4.3.3 Ponencias en Congresos Nacionales

Rosado, P., Planas, M., Figueras, E. & Reverter, F. (2013). La visión artificial, un nuevo aliado para el análisis de imágenes artísticas. I Congreso de investigadores en arte. El arte necesario. Valencia, España, 11-12 Julio p. 793-798. http://congresoncionaldeinvestigadorese-

narte.blogspot.com.es/p/comunicados_19.html

4.3.4 Libros

Reverter, F., Figueras, E., Planas, M. & Rosado, P. (2012). *Ideación y catalogación artística basada en métodos de visión artificial*. Barcelona: Editorial Raima.

Reverter, F., Figueras, E., Planas, M. & Rosado, P. (2012). *Artístic ideation and cataloging based on artificial vision methods*. Barcelona: Editorial Raima.

CONCLUSIONS

4.1 CONTRIBUTIONS

Data storage is no longer an issue today. As IT users, we currently have access to large quantities of content, even in our homes, stored in electronic devices that are permanently within arm's reach. The real problem is how we access data, working as we do within the confines of our database index of choice. Given our human limitations, computer vision appears to be the only solution and this is why considerable effort has been expended in computer vision research. This thesis explores the benefits the art world can reap from such research. The study uses certain computer vision models, which to some extent simulate the initial stages of human visual perception, to help classify the data in large sets of images of abstract paintings. The results have been considered as satisfactory by art experts. Although the system programmed in the thesis is naturally no substitute for expert criteria, it can support the specialist by establishing analogies between works of art and identifying latent patterns in large collections. It allows researchers to repeat searches on the collections of the same artist in different periods or the collections of different artists or periods using the same criteria. The results can therefore be compared without the risk of subjective interpretation conditioned by personal bias or prior knowledge.

Because our human capacity to analyse certain types of image is limited and because many artworks are simply not available to us, it might be argued that there is no other way to conduct an art study of this kind. And although classification is an implicitly subjective activity which proceeds according to one criteria at the exclusion of others, digital technology still offers a unique means of storing and approaching visual content whose sheer volume would otherwise be impossible to process.

> The basic method [seeing all the media objects and then interpreting them] no longer works. [...] We need techniques which would allow us to observe vast "media universes"

> and quickly detect all interesting patterns [...]These techniques have to compress massive media universes into smaller observable media "landscapes" compatible with the human information processing rates. At the same time, they have to keep enough of the details from the original images, video, audio or interactive experiences to enable the study of the subtle patterns in the data (Manovich, 2012, p.1).

Note that the tools presented here may interest two kinds of user: artists reviewing their own creative process from one period of time to another, and art scholars wishing to study a series of artists within a given period of history or the art of different periods without having to move a single work from its original location. Finally, while the thesis does seek to argue the aesthetic merits of the groupings, their mathematical design does effectively contribute to generating new points of view unconditioned by personal or historical bias.

The text below provides an account of the main contributions made by the thesis presented in this report.

4.1.1 Developing programs in Matlab made extendable to other art collections

Essentially methodological in its focus, this thesis puts artificial intelligence techniques at the service of artists and art scholars. AI applications are producing valuable returns in many fields of knowledge and, as this study shows, they can also be used with optimal results in the discovery of visual analogies in images of abstract works of art. This feat of interdisciplinary osmosis, by which models designed for the purposes of engineering can answer questions about art, is attained by customizing computer vision tools to suit the discipline in question and provide the window it needs on its particular object of study.

To do this, the study developed four programs using the numerical computing environment Matlab, which can extend the application of the computer vision algorithms described to the study of any collection of digital images, whether to compare the works of abstract artists of the same period, works across different periods, works of different kinds (e.g.,

abstract versus figurative), or any other subject the user chooses to study with these tools. The programs are described below.

4.1.1.1 Supervised discriminative learning

See the mind map in Fig. 3.1. The program allows searches within a large collection of pictures and retrieves the images that belong to the formal classes the system has been trained to detect, using a manually labelled training set of as few as 15 images in each class and making correct assumptions in 70 % of the cases (meaning that the system assigns 70% of the images it processes to the right class). Working with a database of 3,000 images and 10 classes, this program offers a retrieval time of approximately 10 minutes, which makes it particularly efficient, and would allow artists and art scholars to search large collections for certain formal configurations determined by training sets that they could modify for each new search.

4.1.1.2 Unsupervised generative learning

See the mind map in Fig. 3.13. Taking as its input a file containing the images the user wishes to study without any external indication or annotation, this program groups images according to visual constants or configurations, thereby generating (a) a formal pre-classification that gives an approximate idea of the set content and (b) a novel and alternative point of view unconstrained by any aesthetic preconceptions conditioning the user's prior knowledge of the study sample. The program provides the formal correlation of an artist's creative intuition so that the artists themselves can take decisions about the creative process in which they are immersed. In general terms, the method allows the user to unveil and describe constants in a single artist's body of work, latent features that can be associated with other works by the same person in other periods or with works by other artists.

In this program we also implement the Shannon entropy so that users may visualise the collection of images according to this computation. *pLSA* methodology (see Paragraph 4 of Appendix A) gives us the probability distribution of certain visual features (aspects) in the images (Fig. 3.19 and 3.20). If we visualise works of art in this way, the information becomes more diagrammatic than it is in the pictures themselves and creates opportunities for us to make new visual associations and groupings with which to posit the presence of analogies between works and artists. The images in which Shannon entropy is low because only one aspect is represented with a high degree of probability help to qualify the aspect they contain and simplify the study of the remaining images that have a higher entropy. In other words, images with a low entropy help to clarify the formal features which define the aspect in particular and this helps the user undertand the set as a whole. In the images with the highest entropy, all the aspects are represented, to a greater or lesser extent, and without the help of an index (like the index for the images with a low entropy) it becomes very difficult to determine which particular image patch contains an aspect.

In short, being able to compute and visualise images according to their Shannon entropy can be particularly useful for analysing a set of works of art.

4.1.1.3 Computing distances and plotting a dendogram

This program computes the Bhattacharyya distance between the frequency histograms of the images as probabilities of latent aspects (Fig. 2.15). This computation can be used to draw a dendogram to illustrate how strongly correlated the images are according to this distance (Fig. 3.98).

This dendogram becomes especially valuable to the artist or art scholar as a visual tool for analysing large and complex sets, such as an artist's complete works. Users can study the pictures in greater depth, look for patterns and associations between paintings or painting series, and extract information to enhance their understanding of a collection as a whole. For this reason the dendogram operates more properly as a document to be read than as

a tree diagram.

The dendogram also helps the user establish the formal relationship between images and groups of images more easily and its facilitating role can be summarised thus:

1. It describes the formal relationship between images according to their distance.
2. It provides a simple method to quickly compare groups of images.
3. It helps the user classify images.

4.1.1.4 Grouping images according to texture descriptors

Texture is an important feature of the art of painting and therefore becomes a useful resource in image classification. Specifically, texture refers to the aggregation of materials that create variations or irregularities in the surface of a painting. These textures enhance the feelings expressed by the canvas or paper on which artists paints, fleshing out the result. Part of the language of painting, texture starts with the look and feel of the canvas itself, its cotton or linen thread and the particular weave of the thread. In order to create effects of texture, artists mix media like earth, sand or thick pigments with drying oil and even wax emulsions and objects made of paper, card and wood.

In computer vision texture is important for object detection and for identifying regions of interest within an image. The program this thesis presents uses Haralick's texture descriptor (Haralick, 1973) to group the set of input images. Texture can also be used to identify similarities or differences between images and draw comparisons with results obtained by the *SIFT* (scale-invariant feature transform) descriptors used in the other programs.

The programs developed for this thesis contribute substantially to the digital humanities, a field which has received much attention in recent years[1]. In this field, new IT knowledge

1 See the report by the Alliance of Digital Humanities Organizations, which promotes digital research

helps solve classical problems in the humanities, as explained by Berry (2011) and Álvaro (2013):

> Computational techniques are not merely an instrument wielded by traditional methods; rather they have profound affects on all the aspects of the disciplines. Not only do they introduce new methods, which tend to focus on the identification of novel patterns in the datas against the principle of narrative and understanding, they also allow the modularisation and recombination of disciplines within the university itself (Berry, p. 13)

> The use of automization in conjunction with digitalisation not only boosts capabilities for analysing text documents, it also creates new capabilities for remixing and producing knowledge, and promotes the emergence of new platforms or public spheres, in which the distribution of information can no longer be considered independently of its production (Álvaro, 2013).

This trend goes back to 1987 and the creation of the Text Encoding Initiative, whose set of Guidelines specified encoding methods for machine-readable texts (Hockey, 2004). As this report has explained, there is extensive literature on the use of computer vision to retrieve and filter text content or to assist natural scene classification and even sound detection; but the researchers have found no reference to its use to retrieve or study digital images of works of abstract art. For this reason the results presented here can make an important contribution to the literature, as the art experts who were consulted have agreed, because they validate the robustness of these methods when applied to the abstract visual content that concerned this study.

4.1.2 Extending the *BoW* model to the analysis of abstract art

The conclusions of this thesis are built on what the researchers have been able to see. The reader will be able to consider the figures in the study to decide how fully the objectives

and teaching in the humanities and arts (http://adho.org).

were attained. The point of departure for the thesis are the works of abstract art which were studied during the research and whose merit consists in the fact that they contain valuable configurations for the artists who created them and who distilled into these paintings, when they were originally created, the aspects of their art they found most interesting.

It is also true that each artist's personal take on the world around them will essentially depend on what their eye experiences. Although perception can take many forms, the form we refer to as our basic, physiological sense of sight remains unquestionably valid for all of us. But we know that the human retina only properly captures a small area in the centre of our field of vision so that when we want to see an object with total clarity, we need to have it in our centre of gaze. This also happens with our appreciation of colour, which declines as the angle from front-on vision increases until this sensibility disappears completely at 80 degrees. And not being aware that our peripheral vision is weaker in distinguishing colour is a fair indication of how the workings of our conscious and unconscious mind intervene in visual perception.

If the media elements in this study were limited to cultural concepts we had absorbed simply by growing up with them (landscapes, cities, houses, cars, etc), we might say that all of us shared a common visual experience, like something we had agreed upon. But our subject was abstract painting and this increased the likelihood that no two spectators would ever share the same visual experience. The engineer, biologist and artist will all see something different in the results we proposed. In fact, in the computer vision conferences where this research has been presented, this was very much the situation: to give an example of sorts, the artist in the audience would detect a recurrent compositional pulse or beat where the engineer, sitting right next to her, saw a certain texture or stamped motif.

Arnheim (1966) has observed that in the presence of art the spectator should "enlighten and enrich his life through seeing and hearing, not […] dissect the formal means by which such enlightenment and enrichment is accomplished" (p. 15). Of course, to be proper spectators of a work of abstract art, we should give ourselves up to the painting's power to stir us emotionally. But the artist or art scholar reviewing the same painting needs to interpret its meaning and find messages in the various shapes and patterns. This thesis

has developed and tested a series of computer vision tools to interpret collections of such paintings and determine the analogies that can be extracted. What we extract when we refer to 'abstract art' is not a single model but the "patterns of 'nonobjective' art" which, as Arnheim observes, do not seek to imitate known models:

> The patterns of "nonobjective " art, if considered from the point of view of the world of natural things, are extremely abstract. They reduce the representation of reality to a visual equivalent of the universal physical and psychological forces that underlie nature and life and of their interplay. In this way they express harmony and disharmony, dominance and coordination, contrast and similarity, movement and rest, equilibrium and disequilibrium, and so forth. From the opposite point of view, however, that is, from the point of view of form, the basic, nonobjective patterns are not abstract. They are the very elements of visual comprehension, the building-stones of the composition the artist creates in order to represent the structure of the world in the way his temperament makes him see it (p. 39).

The difficulty of interpreting the results of searches on "nonobjective art" is compounded by our need to identify formal features and compositional and structural values rather than known objects or scenes. But in view of the results, the researchers conclude that with a large collection of images the system tested would allow us to make a fairly objective formal pre-selection and that this could stand free from the limitations associated with human conceptual perception. A more objective process of analysis like the one described here would provide us with new principles for grouping the images in a set. These results extend previous findings on the use of *BoW* and *pLSA* models to facilitate pattern recognition in image collections of natural scenes. However, the researchers have found no reference in the literature to the use of these models to study patterns in abstract painting and this positive evaluation should contribute to confirming their value in this field.

4.1.3 A mathematical approach to art and to patterns

In the words of Wiener "computing machines are essentially machines for recording numbers, operating with numbers, and giving results in numerical form" (Wiener, 1998, p. 159).

Adopting the same principle but in much simpler terms, this thesis seeks to reduce the visual features of a set of paintings to numbers in order to compute degrees of similarity between different paintings and posit formal analogies.

The properties of certain numbers are often associated with aesthetic pleasure but this thesis uses mathematics not to define beauty through formulae but establish formal relationships between the works in question. As observed by Emmer in his discussion of art and mathematics, "everyone can look at a work of art, listen to a symphony, but one cannot look at or listen to mathematics" (Emmer, 2005). In other words, we can all have a (subjective) opinion of the relative merits of art and music but to appreciate the inherent beauty of mathematics we need training. To master mathematical concepts we need years of study and there is no easy way of speeding this up. On the other hand, these highly complex numerical values can be exploited to make other kinds of meaning more intelligible to us, in this case the formal similarities between abstract paintings evidenced by systems that use those values. In this way, mathematics becomes the curator of collections of images and sets their discourse criteria: the formal proximity between the paintings. And we do actually appreciate the mathematics in visual terms in the sense that the formal analogies that emerge in the collections are revealed by their mathematical distance from each other.

4.2 PROPOSAL FOR FUTURE APPLICATIONS

4.2.1 The construction of a visual vocabulary

The *BoW* model allows us to visualize the image patches that correspond to the visual words (pixel arrays) studied. Applied to the collections of different artists, this could be used to create a database in the form of a visual dictionary for artists and art scholars, a resource for both creating and studying art. Artist users would also be able to construct their own vocabulary and extend this by inputting new images, working from the feedback

this offered them. This would provide valuable material for professionals involved creative processes, the practice of art criticism, art education and the classification of art taxonomies.

4.2.2 Applications in artistic creation

This new method for analyzing large sets of data in turn provides other ways of generating creative activity, as Manovich considered in 2005 in his analysis of the principles of the "new media". Here, we recall the possibilities of four of those principles for the purposes of our research:

1- *The principle of numerical representation*. Composed of digital code, our reproductions of abstract paintings are numerical representations. The information they contain can therefore be described using mathematical functions and be subjected to algorithmic manipulation, which, as Manovich observed, means that "media becomes programmable".

2- *The principle of modularity*. Our visual words are arrays of pixels which can be combined or assembled into larger-scale sequences to create new artworks together or with other media elements, but which continue to have their separate, editable identities.

3- *The principle of automation*. Numerical representation and modularity allow us to automate many of the operations involved in creating, manipulating and having access to new artworks.

4- *The principle of variability*. Automation and the modular structure of a media object allow us to alter data, customize them to our needs and, as Manovich explains, interact with the media insofar as we benefit from "different kinds of interactive structures and operations". The use of formal aspects of artistic production detected by computer vision together with the current trends in evolutionary art will be of major interest to research on the computational criteria that new technologies offer art professionals to enhance the creative process.

4.2.3 Applications in art education

Analysing images presented in groups is a curious experience: when we first see one of them, on its own it can appear wanting. Initially, therefore, it's advisable to have looked at all the groups so that on the second and subsequent viewings, when we are familiar with the whole set, we immediately recognise features in one group that can seen across the entire set. If we focus on just one group or attempt to translate the general properties of each class into words (i.e., words composed of letters rather than pixels), our perception of the set as a whole is dulled. Consider this analogy with the principle of photographic depth-of-field: if we focus on just one part of the scene before us, the rest of the picture will be thrown out-of-focus. To keep as much of the whole picture as possible sharp—to have lots of depth-of-field—we need to make sure all or most of our photograph is in focus. In terms of the groups in our collections of paintings, that means we need to be able to see everything at the same time. And seeing everything at the same time provides us with a kind of aesthetic intuition, an analogy to meaning that attracts objects in the images to one another and gathers them in a single class. The class in itself forms a corpus containing a meaning that is lost when we focus on just one painting. To explain it in another way, if a particular artist had only ever painted one picture in her life we would lose the opportunity to perceive the potential value contained in her preparative studies, investigative projects and basic workaday decisions. We only come closer to appreciating an established artist's private manner of looking at the world by reviewing that person's complete body of work. And although all classifications become subjective by serving one criterion rather than another, when we gather into our field of vision complete sets of data—in this case, different groups of abstract paintings—the whole acquires a sharpness that goes beyond the sum of its parts. The sharpness of the whole is absolutely critical in art education. This is evidenced in the average university art class where teachers are familiarising students with the notions of composition or the harmony of the parts within a whole and where students acquire the ability to compare and contrast different works. In this respect, the models and instrument presented in this thesis can contribute in an important way to tapping the knowledge contained in a collection of images.

The act of seeing is not a linear process and art education is essentially about learning values which cannot be taught by reason. In an academic environment the system described in this research could provide teachers and students alike with an instrument to establish criteria for the formal and objective analysis of a set of images or works. This would make it more possible for teaching programmes to incorporate comparative studies using the associations such a resource generates, autonomously and without human or academic bias, among paintings created in different styles or periods or among paintings that take contrasting objects of study.

The thesis reported on here could also contribute substantially to the development of scientific knowledge in the field of art. For example, it could be practically applied in the writing of theses on art history, supporting the doctoral research conducted by students in conjunction with their thesis directors on artists and works of art (by and large the most common area of such research) and offering researchers a valuable contrastive method for cataloguing works of art.

4.2.4 Applications to museum

The developed project could be the key element of a CBIR (content-based image retrieval) system which would enable a new different approach towards artistic institutions. The viewer could focus attention on a specific artwork, capture the image with a mobile device and send it to the system, which could access a large image database to show them the images in the collection which share semantic aspects with it. Knowledge and interpretation are not established only by the exhibition curators or those responsible for said cultural centres, but also by the visitors themselves with their own readings or interpretations. The project is also intended to facilitate the dissemination of personal readings using the new Internet communication tools, which may raise the question of redefining many of the current museum regulations, particularly with respect to the reproduction and dissemination of exhibited or in store artworks.

In view of this evidence, the project we present is intended to widen art knowledge and dissemination in a more dynamic way, and aims to do this in a more attractive manner, by means of a participative approach from the consumer. In most artistic institutions, be they art galleries, museums, artistic centres, historic centres, art halls, etc., the visitors usually just follow an established tour, contemplating the exhibits. Optionally, the visitor can be accompanied by an audio guide or a personal guide that complements the visual tour and introduces the visitor to the universe they contemplate, generally from a historicist point of view, and/or with the aid of some technical or stylistic references.

The written word on the general information panels and cards is another classical resource in the transmission of knowledge regarding the exhibits. This system, obsolete in our opinion, limits graphic documentary sources since information is almost exclusively transmitted by reading, or, as previously mentioned, orally complemented by audio guides. With these prevailing formats and channels, the information is without doubt unidirectional and, often, boring and of little significance.

We start from the assumption that if we manage to make the contemplation of exhibits in the aforementioned institutions more interesting, the number of visitors and their degree of satisfaction will probably increase. To that end, we think it is important to reinforce the visitors' participative role. The interaction between the visitor and the work of art will generate feedback, which will lead to more participation and interest: by taking part in the process, the viewer adopts a certain creative role, which is a stimulating element.

Being a creative system, which generates information, it provides new knowledge and interpretations and helps establish new possibilities or registers. We start from the premise that when interacting with images, we leave aside the archetypical way of looking at the work of art, by generating more abstract and even unprecedented perceptions of the said work of art. It will be the viewer who will be in control of their vision and define its parameters. It will no longer be the "official" or the established historicist or academic view, but a ludic, creative, active and expectant attitude from the visitor will come into play.

4.2.5 Applications in psychology

Because they mathematically model processes of visual perception, the programs in this thesis can be used in simulation studies in research on human visual abilities. They could therefore be applied in the field of cognitive psychology to study the human brain regions involved in visual categorisation, for example by testing the convergence rates in categorisation tasks performed by computational systems against the performance of human experimental participants. At the time of writing, the researchers have discussed these possibilities with other research teams in this field and a battery of psychological tests is currently being designed to consider the following:

1- If human subjects performed the same classification as the programs, the results could validate the principles of computer vision that were used.

2- Comparative studies could also be conducted with human subjects with no specific training in art.

4.3 PUBLICATIONS DERIVED FROM THIS THESIS

The following publications in the fields of art and computer vision are a direct consequence of the research conducted during the preparation of the thesis and indicate the progress that has been made.

4.3.1 Articles in scientific journals

Rosado, P., Reverter, F., Figueras, E. & Planas, M.A. (2014). Semantic-Based Image Analysis with the Goal of Assisting Artistic Creation. *Lecture Notes in Computer Science*. *8671*, 526 -

533. Doi: 10.1007/978-3-319-11331-9

Rosado, P., Figueras, E. & Reverter, F. (2014). Intersecciones entre visión artificial y mirada artística. *BRAC-Barcelona, Research, Art, Creation. 2(1)*, 1 - 54. Doi:10.4471/brac.2014.01

Reverter, F., Rosado, P., Figueras, E. & Planas, M.A. (2012). Artistic ideation based on computer vision methods. Journal of Theoretical and Applied Computer Science. 6(2), 72 - 78. http://www.jtacs.org>.

4.3.2 Presentations at international conferences

Reverter, F., Rosado, P., Figueras, E. & Planas, M.A. (2012). Art images classification using Bag-of-Visualterms representation. Proceedings of Postdigital Art (CAC3). París (França). p.157-160. http://postdigital.eu/program

Rosado, P., Reverter, F., Figueras, E. & Planas, M. (2014). Semantic-Based Image Analysis with the Goal of Assisting Artistic Creation. *In proceedings International Conference on Computer Vision and Graphics 2014 (ICCVG),* Warsaw, Poland, Setember 15-17. Editors: Springer, p. 526 - 533. Doi: 10.1007/978-3-319-11331-9. ccvg.wzim.sggw.pl/default.asp

Rosado, P & Reverter, F. (2015). La mirada in silico. *III Simposio Internacional de la Tipografia al Libro-Arte 2015,* Ciudad Juarez, México, 27 - 29 de Abril. http://3ersimposio.esy.es/index.html

4.3.3 Presentations at national conferences

Rosado, P., Planas, M., Figueras, E. & Reverter, F. (2013). La visión artificial, un nuevo aliado para el análisis de imágenes artísticas. I Congreso de investigadores en arte. El arte necesario.

Valencia, España, 11-12 Julio p. 793-798. http://congresoncionaldeinvestigadoresenarte.blogspot.com.es/p/comunicados_19.html

4.3.4 Books

Reverter, F., Figueras, E., Planas, M. & Rosado, P. (2012). *Ideación y catalogación artística basada en métodos de visión artificial*. Barcelona: Editorial Raima.

Reverter, F., Figueras, E., Planas, M. & Rosado, P. (2012). *Artístic ideation and cataloging based on artificial vision methods*. Barcelona: Editorial Raima.

ANEXO A

1. DESCRIPTORES *SIFT* (SCALE INVARIANT FEATURE TRANSFORM)

El algoritmo *SIFT* se compone principalmente de cuatro etapas que se describen siguiendo la Implementación de (Lowe , 2004):

1- Detección de extremos en el Espacio de Escala: La primera etapa del algoritmo realiza una búsqueda sobre las diferentes escalas y dimensiones de la imagen identificando los candidatos a *keypoints*. Esto se lleva a cabo mediante la función DoG (*Difference-of-Gaussian)*.

2- Localización de los *keypoints*: Se seleccionan los *keypoints* a partir del conjunto de candidatos encontrados, aplicando una medida de estabilidad sobre todos ellos para descartar los que no sean adecuados.

3- Asignación de la orientación: Se asignan una o más orientaciones a cada *keypoint* basándose en las direcciones locales presentes en la imagen gradiente. Todas las operaciones posteriores serán realizadas sobre los datos transformados según la orientación, escala y localización dentro de la imagen, lo que nos proporcionará la invariancia parcial a distorsiones de forma así como a cambios de iluminación.

4- Descriptor del *keypoint*: La última etapa hace referencia a la representación de los *keypoints* como una medida de los gradientes locales de la imagen en las proximidades de dichos puntos clave y respecto de una determinada escala. Cada punto de interés corresponde a un vector de características compuesto por 128 elementos.

A continuación detallaremos las etapas anteriores de una forma más ilustrativa (Enebral, 2009).

1.1 Detección de extremos en el Espacio de Escala

La primera fase del algoritmo es la encargada de buscar un primer conjunto de *keypoints* de la imagen candidatos a poder ser identificados de forma repetida bajo diferentes vistas

del mismo objeto. La detección de ubicaciones que son invariantes frente a cambios de escala de la imagen se puede lograr mediante la búsqueda de características estables en todas las escalas posibles, utilizando una función continua de escala conocida como función Espacio de Escala.

La función Espacio de Escala de una imagen se define como la función, $L(x,y,\sigma)$, que resulta de la convolución de una función Gausiana, $G(x,y,\sigma)$,con la imagen original $I(x,y)$:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y)$$

donde $*$ denota el operador convolución en *x* e *y*, y:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right)$$

El cálculo de todo el espacio $L(x,y,\sigma)$ se lleva a cabo construyendo una pirámide gausiana (Fig. A.1), convolucionando con diferentes filtros $G(x,y,\sigma)$ al variar el parámetro σ. Las imágenes de la pirámide gausiana se distribuyen según los términos siguientes:

- Octava: Imágenes del espacio $L(x,y,\sigma)$de igual tamaño que difieren en el parámetro de filtrado σ con el que han sido obtenidas.

- Escala: Imágenes del espacio $L(x,y,\sigma)$ filtradas con el mismo parámetro σ pero con diferentes tamaños.

Para mejorar la estabilidad de los puntos de interés que se obtendrán más adelante, es conveniente realizar un pre-procesado. La imagen original $I(x,y)$ se suaviza mediante un filtrado gausiano con $\sigma_0 = 0.5$ y posteriormente se re-escala con un factor 2 usando interpolación lineal. La imagen resultante, al doblar su tamaño, le corresponderá un valor y

$\sigma_0 = 1$ es ésta la que se utilizará como imagen inicial para construir la pirámide.

Los diferentes valores del parámetro σ con los que se configura la pirámide tienen que verificar en cada octava la siguiente condición: el penúltimo, σ_4 en este caso, ha de ser el doble que el primero, $\sigma_0 = 1$. Por consiguiente, dividiremos cada octava en intervalos múltiplos de *k*

$$k = 2^{\frac{1}{(num.escalas)-2}} = 2^{\frac{1}{3}}$$

entonces

$$\sigma_i = k^{i-1} = 2^{\frac{i-1}{3}} \qquad i = 1, \ldots, 5$$

Una vez terminada la primera octava, se elige la imagen con $\sigma_4 = 2$ como imagen inicial de la siguiente octava, de esta manera, al re-escalarla a la mitad su factor de filtrado vuelve a ser $\sigma_1 = 1$. Este proceso se va repitiendo hasta completar toda la pirámide (Fig. A.1).

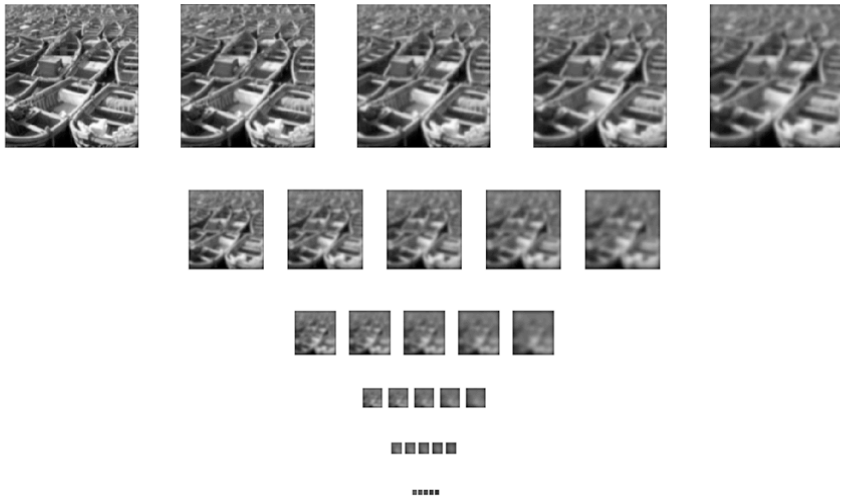


Figura A.1. Pirámide Gausiana

1.2. Localización de keypoints

Para detectar puntos de interés estables en el Espacio de Escala utilizamos la función DoG (Difference-of-Gaussian), definida por:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y)$$
$$= L(x,y,k\sigma) - L(x,y,\sigma)$$

Obsérvese que la función DoG resulta de la función $L(x,y,\sigma)$ calculada en la etapa anterior. La obtención de la función DoG no comporta un incremento considerable del coste computacional total, ya que se calcula simplemente restando imágenes vecinas de una misma octava. En la pirámide DoG tendremos cuatro imágenes-resta por octava (Fig. A.2).

Figura A.2. Imágenes-resta de una primera octava. La pirámide DoG se completa obteniendo las imágenes-resta de las sucesivas octavas.

A partir de los cálculos anteriores, se hallarán los máximos y mínimos locales del espacio $D(x,y,\sigma)$. En esta etapa cada uno de los píxeles de cada imagen de la pirámide se compararán con sus ocho vecinos de la propia imagen y con los nueve vecinos anteriores y posteriores de escala (Fig. A.3).

Un punto quedará seleccionado como keypoint sólo si es mayor que sus 26 vecinos o menor que todos ellos. Observamos que sólo se podrán detectar keypoints en escalas

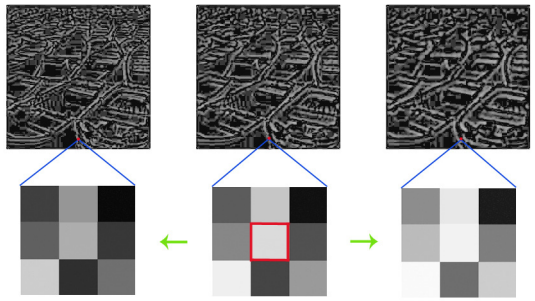


Figura A.3. Vecinos anterior y posterior de escala.

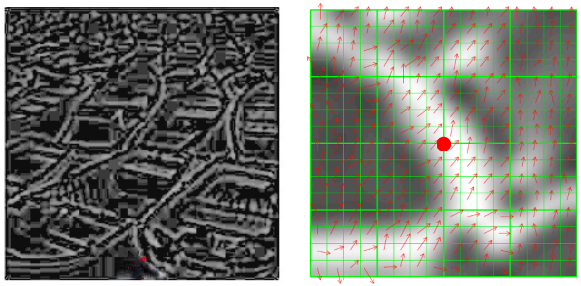


Figura A.4. a) Keypoint. b) Región de 16x16 píxeles alrededor del keypoint y gradiente.

centrales de $D(x,y,\sigma)$ pues no existen imágenes vecinas en las escalas laterales.

En esta fase del método *SIFT* se centra en almacenar toda la información disponible de cada keypoint. Es decir, para cada punto de interés encontrado se guardará a qué escala y octava de la pirámide pertenece, y su posición, es decir la fila y la columna, dentro de la imagen correspondiente.

1.3. Asignación de la orientación

En esta etapa calcularemos las orientaciones de cada punto de interés. Una vez las tengamos podremos construir descriptores invariantes a la rotación, ya que éstos serán referenciados a sus respectivas orientaciones.

Alrededor del punto donde vamos a determinar la orientación definimos una región de 16 x 16 píxeles y a cada uno de los píxeles se le calcula su gradiente (Fig. A.4a y A.4b). El gradiente viene determinado por su módulo $m(x,y)$ e inclinación $\theta(x,y)$, ambos se calculan utilizando deferencias entre pixeles:

$$m(x,y) = \sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1)-L(x,y-1}{L(x+1,y)-L(x-1,y)}\right)$$

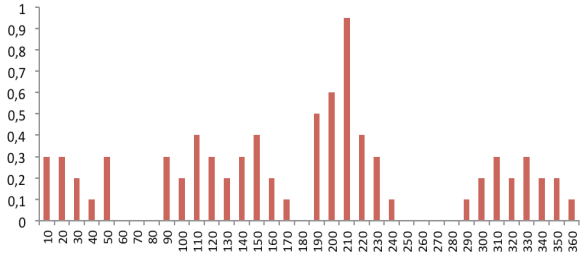


Figura A.5. Histograma de orientaciones. La orientación del keypoint corresponde al valor máximo.

La imagen empleada para obtener $m(x,y)$ y $\theta(x,y)$ será la imagen de la pirámide $L(x,y,\sigma)$ (Fig. A.1) donde se detectó el punto de interés que está siendo analizado.

Después de realizar el proceso anterior, se agrupará la información en forma de histograma, uno para cada punto de interés. De esta manera cada histograma de orientaciones estará formado por 36 bins para completar el rango total de 360º (Fig. A.5). A medida que se añade al histograma cada orientación $\theta(x,y)$, dicho valor se pondera por su módulo $m(x,y)$ y por una ventana circular gausiana con valor σ igual a 1.5 veces la escala del

punto de interés. Los motivos principales para realizar estas dos ponderaciones son: dar mayor peso a las orientaciones con módulos elevados y mayor importancia a los puntos cercanos al punto de interés.

El bin modal de cada histograma corresponderá a la dirección dominante de los gradientes locales, y por lo tanto a la orientación final del punto de interés.

1.4. Descriptores de los keypoints

Las etapas anteriores han dotado a los puntos de interés seleccionados de invariancia respecto de la orientación, escala y localización respecto de la imagen. En esta última etapa se crea un vector de características para cada uno de los puntos de interés que contiene una estadística local de las orientaciones del gradiente.

El proceso parte de las regiones 16 x 16 del apartado anterior ya multiplicadas por la ventana gausiana con σ igual a 1.5 veces la escala del punto de interés (Fig. A.6a). Cada una de estas regiones se divide en subregiones de 4 x 4 píxeles con el objetivo de resumir toda esa información en pequeños histogramas de sólo 8 bins, es decir, 8 orientaciones. Previamente a la realización de esta reconfiguración, cada gradiente de la ventana 16 x 16 se rota tantos grados como especifique la orientación del punto de interés (calculada en la etapa anterior, Fig. A.5), y así será independiente a la inclinación de la imagen.

Para cada punto de interés ahora pasaremos a tener 16 pequeños histogramas de 8 bins cada uno de ellos. Para evitar cambios abruptos entre las fronteras de las subregiones, cada una es filtrada de nuevo por una ventana circular gausiana (en esta ocasión de tamaño 4 x 4) con un factor σ = 0.5 x escala del punto de interés (Fig. A.6b).

Cada uno de los histogramas se compone de 8 bins, que almacenan las orientaciones posibles proporcionales a 45 grados donde la magnitud de cada flecha representa el valor acumulado para cada bin. Por lo tanto se obtienen 16 histogramas respecto de las orientaciones de los puntos de cada región para cada uno de los puntos de interés.

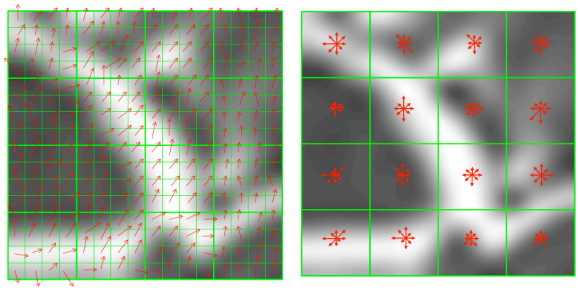


Figura A.6. a) Región de 16 x 16 píxeles alrededor del keypoint y gradiente. b) Subregiones de 4 x 4 píxeles con histogramas de sólo 8 orientaciones.

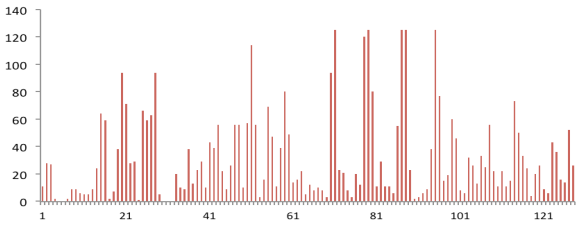


Figura A.7. Descriptor

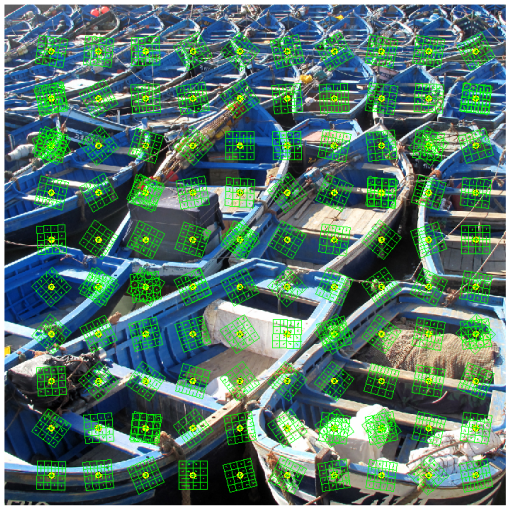


Figura A.8. Imagen con una malla de 10 x 10. Los nodos contienen los valores de las 8 orientaciones de los histogramas 4 x 4. Para cada nodo resulta un vector de características, con 4 x 4 x 8 = 128 elementos.

Finalmente el descriptor de cada punto de interés está formado por un vector que contiene los valores de las 8 orientaciones de los 4 x 4 histogramas componiendo un vector de características de 4 x 4 x 8 = 128 elementos (Fig. A.7).

Las etapas anteriores aseguran que los descriptores obtenidos sean invariantes frente a la luminosidad. Esto es consecuencia de que los gradientes están calculados mediante diferencias entre píxeles vecinos. De esta manera sumar una constante de luz a la imagen no influirá en el resultado final.

Para lograr invariancia frente a cambios de contraste hay que normalizar a la unidad cada uno de los 'sub-histogramas' del total de 16 que tiene cada descriptor.

Por la saturación de la cámara o por cambios de iluminación sobre superficies puede producirse una variación no lineal de luz.

Ante esta variación para reducir sus efectos impondremos un umbral superior de 0.2 a los histogramas normalizados y posteriormente se renormalizará de nuevo a la unidad. Una vez efectuadas estas modificaciones, el proceso de construcción de los descriptores queda completado.

En algunos estudios (Lazebnik, Schmid & Ponce, 2006; Fei-Fei & Perona, 2005) el cálculo de los descriptores locales *SIFT*, en vez de realizarse en los puntos de interés, se efectúa en los nodos de una malla regular sobreimpuesta en la imagen (Fig. A.8). Este enfoque es preferible con el fin de mejorar la capacidad de discriminación en implementaciones orientadas a la clasificación de escenas.

2. DESCRIPTORES DE TEXTURA DE HARALICK

La textura es una de las características importantes utilizadas para la identificación de objetos o de zonas de interés en una imagen. Aunque intuitivamente se pueden asociar diversas propiedades de las imágenes, tales como suavidad, rugosidad, regularidad, etc. (Gonzalez y Woods, 2008; Bharati y col., 2004), realmente no existe una definición formal o completa de textura. Muchos investigadores describen la textura utilizando varias definiciones. Russ

(1999) considera la textura de una imagen como la variación entre píxeles en una pequeña vecindad de una imagen. Alternativamente, la textura puede describirse como un atributo que representa la distribución espacial de los niveles de intensidad en una región dada de una imagen digital (Bharati y col.,2004). Existe, en ambas definiciones, el concepto de variación espacial en un entorno de vecindad.

Los descriptores de textura definidos por Haralick (Haralick, 1973) son un conjunto de medidas de textura basadas en la matriz de co-ocurrencia.

Son de naturaleza estadística y para su cálculo, es necesario asumir que la totalidad de la información textural de una imagen está contenida en las relaciones espaciales que se dan entre los distintos niveles de gris de un objeto. Incorporan información espacial en forma de posición relativa entre niveles de intensidad dentro de la textura.

Para cada imagen se crea una matriz de co-ocurrencia de niveles de gris (*GLCM,* Gray-Level Co-occurrence Matrix) mediante el cálculo de la frecuencia con la que un píxel con un nivel de gris determinado *i* se presenta horizontalmente adyacente a un pixel con el valor de *j*. Cada elemento *(i, j)* de la matriz de co-ocurrencia especifica el número de veces que el píxel con valor *i* ocurrió horizontalmente adyacente a un píxel con valor *j*. Se puede especificar el número de niveles de gris (en nuestro caso hemos utilizado 256 niveles).

La Fig. A.9 muestra la forma en que se calcularían algunos valores de la matriz de co-ocurrencia para una imagen de 4 x 5 píxeles. El elemento (1,1) de la matriz de co-ocurrencia contiene el valor 1 porque sólo hay una instancia en la imagen donde dos píxeles horizontalmente adyacentes tienen los valores 1 y 1 de nivel de gris. Elemento (1,2) de la matriz contiene el valor 2 porque hay dos instancias en la imagen en la que dos píxeles horizontalmente adyacentes tienen los valores 1 y 2.

El número de veces que el valor 1 aparece adyacente al valor 2, es decir los casos (1,2) y (2,1), se cuentan juntos. Así la matriz de co-ocurrencias que se genera es simétrica respecto a su diagonal (Fig. A.11).

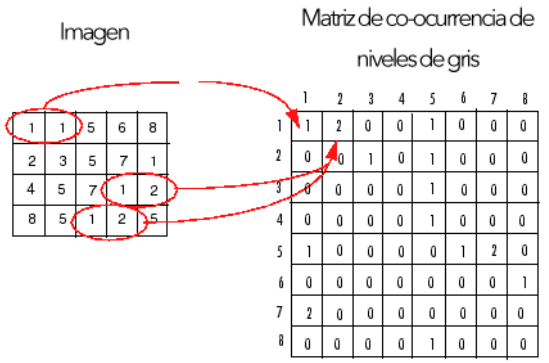


Figura A.9. Cálculo de la matriz de co-ocurrencia de niveles de gris de una imagen de 4 x 5 píxeles.

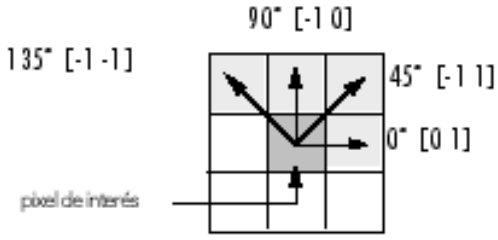


Figura A.10. Ilustra el cálculo del desplazamiento para un sólo píxel.

$$T = \begin{bmatrix} 0 & 0 & 2 & 2 & 3 \\ 1 & 1 & 0 & 0 & 2 \\ 3 & 2 & 3 & 3 & 1 \\ 3 & 2 & 2 & 2 & 0 \\ 0 & 1 & 2 & 3 & 0 \end{bmatrix}$$

$$A_1^0 = \frac{1}{40}\begin{bmatrix} 4 & 2 & 3 & 1 \\ 2 & 2 & 1 & 1 \\ 3 & 1 & 6 & 5 \\ 1 & 1 & 5 & 2 \end{bmatrix} \quad A_1^{45} = \frac{1}{32}\begin{bmatrix} 0 & 1 & 3 & 3 \\ 1 & 0 & 3 & 1 \\ 3 & 3 & 2 & 4 \\ 3 & 1 & 4 & 0 \end{bmatrix}$$

$$A_1^{90} = \frac{1}{40}\begin{bmatrix} 2 & 3 & 2 & 3 \\ 3 & 0 & 3 & 1 \\ 2 & 3 & 4 & 4 \\ 3 & 1 & 4 & 2 \end{bmatrix} \quad A_1^{135} = \frac{1}{32}\begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 0 & 1 & 2 \\ 2 & 1 & 6 & 3 \\ 2 & 2 & 3 & 0 \end{bmatrix}$$

Figura A.11. Ejemplo para cuatro niveles de gris y una imagen de 5 x 5.

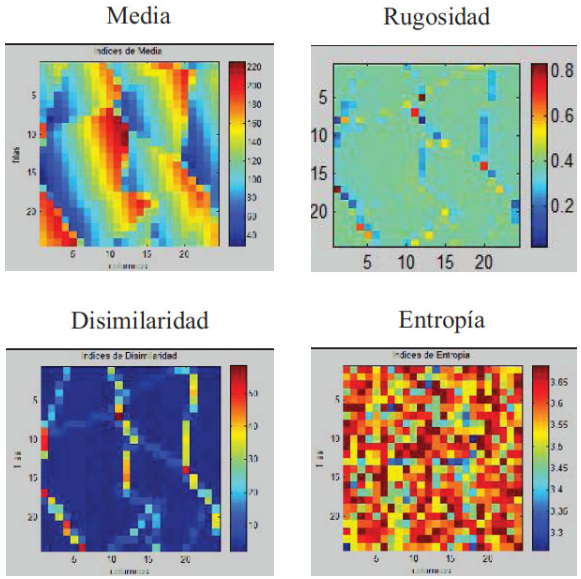


Figura A.12. Media, rugosidad, disimilaridad y entropía, propiedades estadísticas de la imagen de la Fig. A.13.

Debido a que el desplazamiento a menudo se expresa como un ángulo, la Fig. A.10 muestra los valores de desplazamiento para una distancia de 1 píxel especificados como ángulos (0°,45°,90°,135°).

La Fig. A.10 ilustra la distancia D, en número de filas y columnas, entre el píxel de interés y su vecino. En este caso concreto D es igual a 1, pero se podría considerar para mayores distancias (0 D, -D D, -D 0, -D -D). En el caso que nos ocupa hemos considerado 4 distancias diferentes: 1,10, 20 y 50 píxeles. Así, para cada imagen se han calculado en total 16 descriptores de Haralick (basados en 256 niveles de gris) correspondientes a; las 4 distancias consideradas multiplicadas por los 4 ángulos de cada distancia.

A partir de estos descriptores, se pueden inferir unas propiedades estadísticas que proporcionan información acerca de la textura global de la imagen. Para calcularlas se normaliza la matriz de co-ocurrencia de niveles de gris *GLCM* de modo que la suma de sus elementos sea igual a 1. Cada elemento *p(r,c)* en el *GLCM* normalizado es la probabilidad de ocurrencia conjunta de pares de píxeles con una relación espacial definida para los valores de nivel de gris *r* y *c* en la imagen. Se utiliza el *GLCM* normalizado para calcular las siguientes propiedades:

Contraste: El contraste en una imagen se refiere a la diferencia relativa en la intensidad de un punto o zona. Por ejemplo, el contraste entre un objeto de brillo constante sobre un fondo de brillo constante. Si ambas superficies tienen el mismo brillo, el contraste será nulo y si el conjunto está en tonos de gris, el objeto será tanto física como perceptiblemente indistinguible del fondo. A medida que aumenta la diferencia en brillo, el objeto será distinguible del fondo. Para ello es necesario alcanzar el umbral de contraste, que se sitúa alrededor del 0,3 % de diferencia. También se conoce como varianza o inercia. El contraste de intensidad entre un píxel y su vecino en toda la imagen se mide con la siguiente fórmula.:

$$\sum_{i,j} |i-j|^2 p(i,j)$$

El contraste es 0 para una imagen constante.

Correlación: Esta propiedad estadística indica la fuerza y la dirección de una relación lineal y proporcional entre dos variables. Se considera que dos variables cuantitativas están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores de la otra: si tenemos dos variables (A y B) existe correlación si al aumentar los valores de A lo hacen también los de B y viceversa. El grado de correlación entre un píxel y su vecino en toda la imagen se mide con la fórmula:

$$\sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j}$$

donde μi es la media del nivel *i* de gris y σi la desviación teórica del nivel *i* de gris. La correlación es 1 o -1 para una imagen perfectamente correlacionado positiva o negativamente. La correlación no es calculable para una imagen constante.

Energía: También se conoce como uniformidad de la energía o segundo momento angular. Se puede medir la Energía de una textura calculando la suma de elementos cuadráticos en el *GLCM* según la fórmula:

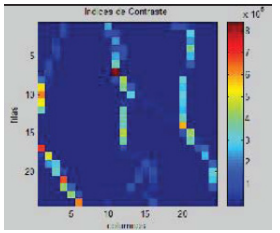
$$\sum_{i,j} p(i,j)^2$$

Para una imagen constante la energía es 1.

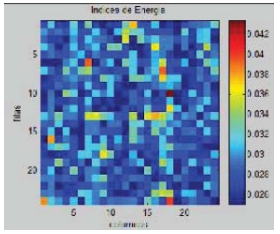
Homogeneidad: Se refiere a la falta de variabilidad. Se calcula el valor que mide la cercanía de la distribución de los elementos de la *GLCM* a la diagonal de la misma, con la fórmula:

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|}$$

Contraste



Energía



Homogeneidad

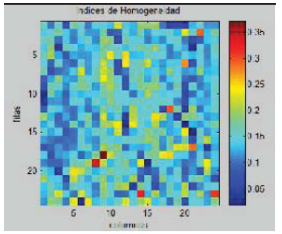


Figura A.13. Ejemplo de algunas propiedades estadísticas que se pueden calcular sobre la textura de una imagen y después visualizar. En primer lugar, arriba, vemos la imagen sobre la que se realizan los cálculos y después propiedades como el contraste, la energía y la homogeneidad. Estas tres propiedades las hemos utilizado en nuestro análisis junto con la correlación.

La homogeneidad es 1 para la diagonal del *GLCM*.

En las Fig. A.12 y A.13 mostramos, simplemente a modo ilustrativo,ejemplos de cómo una vez calculadas las características estadísticas de la textura de una imagen, sería posible tener una representación visual de las propiedades concreta.

Para computar la matriz de co-ocurrencia de niveles de gris *GLCM* y calcular las propiedades estadísticas energía, contraste, homogeneidad y correlación, se han utilizado las funciones graycomatrix y graycoprops de *MATLAB* (2011). Finalmente se realiza la agrupación final con el número de grupos deseados, aplicando el algoritmo *K-means* (ver Anexo A, apartado 5).

3. CONSTRUCCIÓN DEL VOCABULARIO VISUAL

Para la construcción de un vocabulario visual en el que basar la descripción de las imágenes, seguimos un procedimiento análogo al que se utiliza en el análisis automático de textos. Se conoce como modelo *Bag-of-Words (BoW)* porque cada documento está representado como una distribución de frecuencias de las palabras presentes en el texto, sin tener en cuenta las relaciones sintácticas existentes entre ellas.

En el ámbito de las imágenes a veces se refieren a representaciones *Bag-of-Visual Terms (BOV)*. Este enfoque consiste en analizar las imágenes como un conjunto de regiones, describiendo solamente su apariencia e ignorando su estructura espacial. La representación *BoW* se construye a partir de la extracción y cuantización automática de descriptores locales y ha demostrado ser una de las mejores técnicas para resolver diferentes tareas en la visión por computador. La representación *BoW* fue implementada por primera vez (Willamowski, Arregui, Csurka, Dance & Fan, 2004) en el desarrollo de una sistema experto de reconocimiento de objetos.

La construcción *BoW* requiere dos decisiones principales de diseño:
a) La elección de los descriptores locales que aplicamos en nuestras imágenes.

b) La elección del método que se utilice para obtener el vocabulario visual.

Ambas decisiones pueden influir en el rendimiento del sistema resultante, sin embargo la representación *BoW* es robusta, conserva su buen comportamiento en un amplio rango de opciones de los parámetros.

El punto de partida para la construcción de un vocabulario visual es el conjunto de descriptores $F = \{f_i : i = 1, ..., N_F\}$ de la colección de imágenes $D = \{d_1, ..., d_N\}$ y el punto al cual queremos llegar es un vocabulario de visual terms $V = \{v_1, ..., v_M\}$. Utilizaremos la expresión: palabra visual como equivalente a la expresión inglesa visual term. Cada imagen d_i ha quedado descrita mediante los descriptores *SIFT*, denotemos genéricamente por f un descriptor . Considerando toda la colección de imágenes tenemos por tanto una gran colección de descriptores. La construcción del vocabulario requiere la cuantización de cada descriptor local f en su respectiva palabra visual v_i de acuerdo con la siguiente regla de asignación:

$$f \rightarrow Q(f) = v_i \Leftrightarrow dist(f, v_i) \leq dist(f, v_j), \quad j = 1, ..., M \quad , \qquad (1)$$

donde $dist(\cdot, \cdot)$ es una función distancia.

Si indicamos con S el espacio de los descriptores, en nuestro caso al tener descriptores 128 dimensionales podemos asumir que $S \equiv R^{128}$. Una vez fijado el vocabulario $V = \{v_1, ..., v_M\}$ S queda dividido en *M* regiones $S = \{S_1, ..., S_M\}$ de acuerdo con:

$$S_i = \{f \in S : Q(f) = v_i\}$$

La construcción del vocabulario se realiza mediante agrupación (*clustering*). Más específicamente, aplicamos el algoritmo *K-means* (ver Anexo A, apartado 5) a un conjunto representativo de descriptores locales extraídos de la colección de imágenes y tomaremos como palabras visuales los vectores de medias de cada clúster. Usamos la distancia euclidiana en los procesos de agrupación y cuantización y elegimos el número de clústeres dependiendo del tamaño deseado de vocabulario.

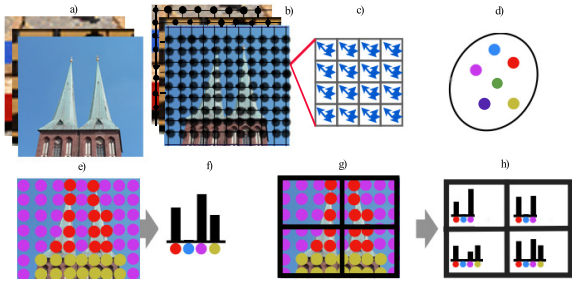


Figura A.14. a) Colección de imágenes. b) Se define una cuadrícula sobre las imágenes c) Se calculan los descriptores. d) A continuación, se cuantizan los descriptores en M clústeres, los cuales definirán un vocabulario visual de M palabras visuales. e) Una vez se dispone del vocabulario, los descriptores de cada imagen se asignan a la palabra visual más cercana. f) Para obtener la representación *BoW* de una imagen dada, se calcula la frecuencia de cada palabra visual en la imagen. g) y h) Secuencia de cuadrículas en la imagen para dibujar las pirámides de histogramas con objeto de tener en cuenta las relaciones espaciales entre palabras visuales.

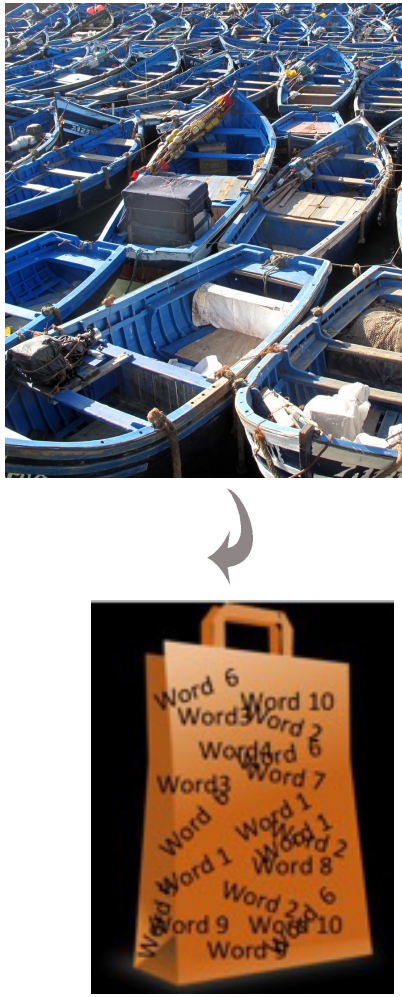


Figura A.15. La representación *BoW* de una imagen no contiene información acerca de las relaciones espaciales entre palabras visuales que la componen.

Dada una imagen d con un conjunto de descriptores $F(d) = \{f_j : j = 1, \ldots, N_{F(d)}\}$ podemos usar los centroides obtenidos en el algoritmo *K-means* para atribuir la palabra visual v_i a todo descriptor f_j para el que el centroide más cercano sea μ_i.

Una vez completada la atribución obtenemos la representación *BoW* de la imagen:

$$h(d) = (h_1(d), \ldots, h_M(d))$$

$$h_i(d) = n(d, v_i)$$

donde $n(d, v_i)$ indica la frecuencia de la palabra visual v_i en la imagen d.

La Fig. A.14 resume el proceso para obtener la representación *BoW* de las imágenes de una colección.

Esta representación de una imagen no contiene información acerca de las relaciones espaciales entre palabras visuales, del mismo modo que la representación *BoW* remueve la información relativa al orden de la palabras en los documentos (Fig. A.15).

No obstante, los métodos *BoW,* que representan una imagen como una colección desordenada de características locales, han demostrado impresionantes niveles de rendimiento en tareas de categorización de imágenes completas.

Sin embargo, debido a que estos métodos no tienen en cuenta toda la información acerca de la disposición espacial de las características, se ha visto limitada su capacidad descriptiva. En particular, son incapaces de capturar formas o de separar un objeto de su fondo.

Para superar las limitaciones del enfoque *BoW* hemos implementado una metodología de histogramas en pirámide que configura una secuencia cada vez más fina de cuadrículas sobre la imagen y lleva a cabo un análisis tipo *BoW* en cada una de las cuadrículas,

obteniendo finalmente una suma ponderada de la cantidad de coincidencias que ocurren en cada nivel de resolución de la pirámide (Grauman & Darrel, 2005; Lazebnik, et al, 2006). (Fig. A.21) Para una información más detallada de este proceso se puede consultar el apartado 9 del Anexo A.

4. REPRESENTACIÓN DE ASPECTOS LATENTES MEDIANTE pLSA

La representación *BoW* es fácil de construir. Sin embargo, adolece de dos inconvenientes (Fig. A.16): polisemia (una sola palabra visual puede representar diferentes contenidos de la escena) y sinonimia (varias palabras visuales pueden caracterizar el mismo contenido de la imagen).

Para solventar en parte los inconvenientes anteriores, encontramos el análisis *probabilístico de aspectos latentes* (*pLSA*), una metodología original de la minería de textos (Hofmann, 2001).

Las aplicaciones del *pLSA* en el análisis estadístico de textos están orientadas a descubrir automáticamente los temas tratados en un documento, tomando como punto de partida la representación *BoW* de documentos.

La extensión del *pLSA* hacia el análisis de imágenes pasa por considerar las imágenes como documentos en un vocabulario visual establecido a partir de un proceso de cuantización como se ha señalado anteriormente. El detectará en las imágenes categorías de objetos, patrones formales, de modo que una imagen que contiene varias tipos de objetos se modela como una mezcla de temas (Fig.15).

Vamos a explicar el modelo en términos de imágenes, palabras visuales y aspectos. Disponemos de una colección de imágenes $D = \{d_1, ..., d_N\}$ y un vocabulario de palabras visuales $V = \{v_1, ..., v_M\}$. Podemos resumir las observaciones en una tabla $N \times M$ de frecuencias $n(d, v_i)$, donde $n(d, v_i)$ indica la frecuencia con que la palabra visual v_i ocurre en la imagen d_i. *pLSA* es un modelo estadístico generativo que asocia una variable latente

Palabra 1

Palabra 2

Figura A.16. Muestras de regiones de imágenes correspondientes a dos palabras visuales de un vocabulario de 300 palabras. Podemos considerar que ambas palabras describen un contenido común; textura rocosa no homogénea, y en este sentido representan una sinonimia. Además, vemos que dentro de una misma palabra hay regiones que representan contenidos distintos, en unos casos el contenido es roca y en otros es muro, por tanto podríamos considerarlas polisémicas.

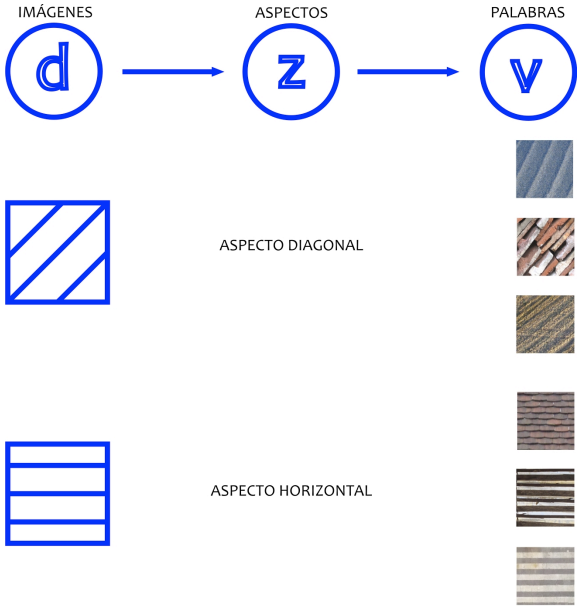


Figura A.17. El método pLSA captura la co-ocurrencia de palabras visuales entre imágenes.

$z_l \in \{z_1, ..., z_K\}$ con cada observación, entendiendo por observación la ocurrencia de una palabra visual en una imagen dada. Estas variables, normalmente llamadas aspectos, se utilizan para construir un modelo de probabilidad conjunta sobre las imágenes y las palabras visuales, definido por:

$$P(d_i, v_j) = P(d_i) \sum_{k=1}^{K} P(v_j|z_k) P(z_k|d_i)$$

donde $P(d_i)$ indica la probabilidad de d_i, $P(v_j|z_k)$ indica la probabilidad condicionada de una palabra visual específica condicionada al aspecto latente z_k, y $P(z_k|d_i)$ indica la probabilidad condicional específica de cada imagen.

El *pLSA* introduce un principio de independencia condicional: asume que la ocurrencia de una palabra visual v_i es independiente de la imagen d_i en la que esta, dado un aspecto z_k.

La estimación de las probabilidades del modelo *pLSA* se llevan a cabo mediante el máximo de la verosimilitud utilizando la colección de imágenes $D = \{d_1, ..., d_N\}$. La optimización se resuelve mediante el algoritmo EM (Dempster, 1977).

El algoritmo EM alterna dos etapas. En la etapa E se calculan las probabilidades a posteriori para los aspectos latentes basándonos en las estimaciones actuales de las probabilidades del modelo, en la etapa M las probabilidades del modelo se actualizan maximizando la llamada *expected complete data log-likelihood*:

Etapa E

$$P(z_k|d_i, v_j) = \frac{P(v_j|z_k) P(z_k|d_i)}{\sum_{l=1}^{K} P(v_j|z_l) P(z_l|d_i)}$$

Etapa M

$$P(v_j|z_k) = \frac{\sum_{i=1}^{N} n(d_i, v_j) P(z_k|d_i, v_j)}{\sum_{m=1}^{M} \sum_{i=1}^{N} n(d_i, v_m) P(z_k|d_i, v_m)}$$

$$P(z_k|d_i) = \frac{\sum_{j=1}^{M} n(d_i, v_j) P(z_k|d_i, v_j)}{n(d_i)} \quad , \quad n(d_i) = \sum_{j=1}^{M} n(d_i, v_j)$$

Las etapas E y M se alternan hasta que se alcanza una cierta condición de terminación. El proceso iterativo se inicia asignando valores aleatorios al conjunto de probabilidades $P(z_k|d_i)$ y $P(v_j|z_k)$

Como consecuencia del proceso anterior obtenemos una nueva representación para las imágenes de la colección basada en la distribución de aspectos,

$$a(d_i) = (P(z_1|d_i), ..., P(z_K|d_i))$$

De hecho, también es posible hallar la distribución de aspectos para una imagen cualquiera que no forme parte de la colección inicial (Quelhas, Monay, Odobez, Gatica-Perez, Tuytelaars & Van Gool, 2005; Bosch, Zisserman & Muñoz, 2006). Basta recurrir de nuevo al algoritmo EM antes descrito pero en este caso en la etapa M sólo se actualizan las probabilidades $P(z_k|d_i)$ y las probabilidades $P(v_j|z_k)$, independientes de la imagen, estimadas a partir de la colección en la fase de aprendizaje, se mantienen fijas.

Si bien la representación de imágenes basada en aspectos se puede usar como punto de entrada para alimentar un clasificador de escenas, nosotros vamos a centrarnos en la utilización de dicha representación para la ordenación o ranking de imágenes basada en la distribución de aspectos subyacentes. Dado un aspecto z, las imágenes pueden

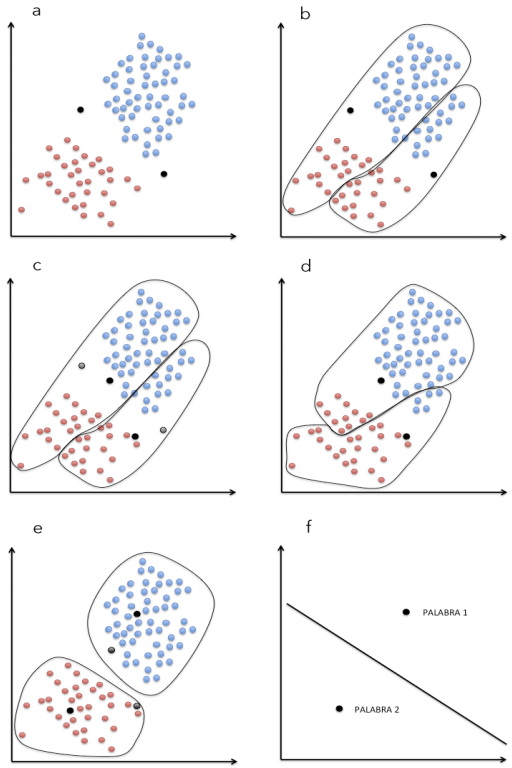


Figura A.18. a) Supongamos que los descriptores de la colección de imágenes configuran dos grupos separados (azul y rojo). El algoritmo empieza estableciendo dos centroides al azar (negro), b) Asignamos cada descriptor al centroide más cercano. c) Recalculamos los nuevos centroides de los grupos formados en la etapa anterior. d) Repetimos la asignación de los descriptores al centroide más cercano. e) El procedimiento prosigue recalculando los nuevos centroides. f) El proceso iterativo se detiene cuando no se produce cambio en los centroides.

ordenarse según los valores:

$$P(d|z) = \frac{P(z|d)P(d)}{P(z)} \propto P(z|d)$$

de esta manera, una vez estimados los valores de $P(z_k|d), k = 1,...,K$, para una imagen dada d, podemos ordenarlos y tener una medida objetiva de la asociación entre la imagen y cada uno de los aspectos. En consecuencia, asociaremos la imagen al aspecto con mayor probabilidad.

A partir de esta metodología nos ha sido posible analizar la colección de imágenes y encontrar aspectos subyacentes mediante los cuales catalogar toda la colección y también contrastar dichos aspectos o patrones con los propuestos por el autor.

5. ALGORITMO *K-MEANS*

El algoritmo *K-means* establece una partición o agrupación del conjunto de descriptores

$$F = \{f_i : i = 1,...,N_F\}$$

en M subconjuntos disjuntos S_i que contienen los descriptores que minimizan la función de error cuadrático:

$$J(F) = \sum_{i=1}^{M} \sum_{j \in S_i} (f_j - \mu_i)^2$$

donde μ_i denota el vector de medias (centroide) del subconjunto e descriptores S_i .

El algoritmo busca la partición mediante la iteración de dos etapas. La primera etapa consiste en asignar cada descriptor al centroide más cercano. En la segunda atapa se recalculan los centroides de cada región, calculando el vector de medias de los descriptores que han sido asignados a cada región. En las Fig. A.18 y A.19 se describe, a modo de ejemplo el

caso de descriptores bidimensionales y de dos palabras visuales. El algoritmo *K-means* establecerá una partición del espacio en dos regiones, cada una asociada a una palabra.

1- Establecer al azar M centroides iniciales.

2- Asignar cada descriptor al subconjunto S_i que tenga el centroide μ_i más cercano, de acuerdo con la fórmula 1.

3- Recalcular el valor del centroide μ_i mediante el vector de medias de los descriptores asignados a S_i.

4- Repetir los pasos 2 y 3 hasta que los valores de los centroides μ_i no se modifiquen.

Dada una imagen d con un conjunto de descriptores $F(d) = \{f_j : j = 1, ..., N_{F(d)}\}$ podemos usar los centroides obtenidos en el algoritmo *K-means* para atribuir la palabra visual v_i a todo descriptor f_j para el que el centroide más cercano sea μ_i. Una vez completada la atribución obtenemos la representación *BoW* de la imagen:

$$h(d) = (h_1(d), ..., h_M(d)) \quad , \quad h_i(d) = n(d, v_i)$$

donde $n(d, v_i)$ indica la frecuencia de la palabra visual v_i en la imagen d.

6. ALGORITMO *SVM* (Support Vector Machines)

Para clasificar una imagen de entrada d representada por su vector *BoW* $h(d)$ empleamos *SVM* (Boser, Guyon & Vapnik, 1992). Se implementan los descriptores de histogramas en pirámide *PHOW* para tener en cuenta la información espacial. Así, la similitud entre un par de imágenes *I* y *J* se calcula utilizando una función *Kernel* entre sus descriptores de histogramas *PHOW* D_I y D_J con ponderaciones adecuadas para cada nivel de la pirámide.

$$K(D_I, D_J) = \exp\left\{\frac{1}{\beta}\sum_{l \in L} \alpha_l d_l(D_I, D_J)\right\}$$

donde β es la media de $\sum_{l \in L} \alpha_l d_l(D_I, D_J)$ sobre los datos de entrenamiento, α_l es el peso

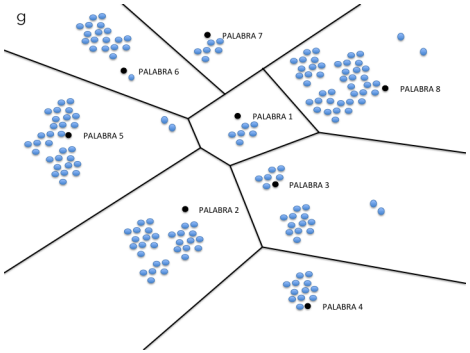


Figura A.19. g) Ilustra la partición del espacio de descriptores en el caso de una vocabulario de más palabras. Dado un descriptor f calcularemos el centroide más cercano, y le corresponderá la palabra representada por dicho centroide.

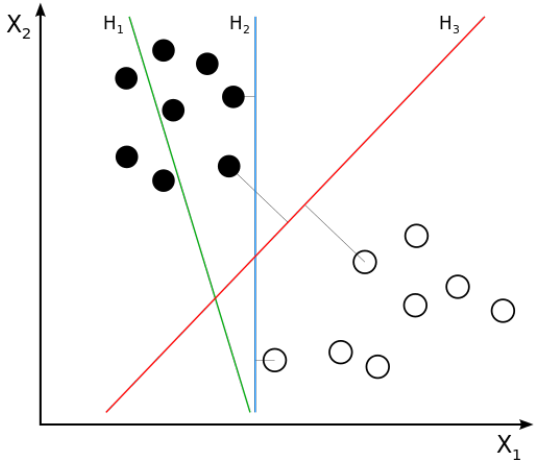


Figura A.20. H1 no separa bien las dos clases, por tanto no sería un buen hiperplano. H2 las separa, pero no es el más indicado, ya que la distancia de los puntos de las clases al plano es muy pequeña. La mejor opción es H3 (rojo) que está más espaciada de las dos clases. © ZackWeinberg (Màquina de vector de suport, 2014)

en el nivel *l* y d_l es la distancia χ^2 (Zhang et al., 2007) entre D_I y D_J en el nivel *l* de la pirámide, calculado utilizando los histogramas normalizados en este nivel.

Los histogramas espaciales podrían ser utilizados como descriptores de imágenes y alimentar a un clasificador *SVM* lineal. Los *SVM* se entrenan muy rápidamente pero también se limitan a usar el producto escalar para comparar descriptores. Se pueden obtener mejores resultados calculando un mapa explícito de características que emulan una χ^2 *-Kernel* no lineal como uno lineal (Vedaldi & Zisserman, 2010).

7. DISTANCIA DE BHATTACHARYYA

En estadística, la distancia de Bhattacharyya (Bhattacharyya, 1943) mide la similitud de dos distribuciones de probabilidad discretas o continuas.

Para distribuciones discretas de probabilidad $p_i = (p_{i1}, \ldots, p_{ik})$ y $p_j = (p_{j1}, \ldots, p_{jk})$ en el mismo dominio de *X*, se define como:

$$d_{ij}^2 = \arccos\left(\sum_{l=1}^{k} \sqrt{p_{il}\, p_{jl}}\right)$$

8. ÍNDICE DE ENTROPÍA DE SHANNON

A raíz del primer análisis de los resultados obtenidos con el *pLSA* considerando 10 aspectos, percibimos que la muestra consta de dos tipologías de imágenes muy marcadas; un tipo de fotografías que presenta un único aspecto muy destacado (las llamaremos imágenes de menos entropía) que son las que aparecen representadas en estas 10 primeras categorías, y otro que presenta varios aspectos asociados simultáneamente (las llamaremos imágenes de más entropía) y que no resulta visible en este primer análisis**.**

Para poder distinguir y tratar separadamente estas dos tipologías hemos utilizado el índice de entropía de Shannon (Cover & Thomas, 2006).

La metodología *pLSA* (ver apartado 4 del Anexo A) proporciona una distribución de probabilidad de los aspectos en las imágenes. Esto es, para una imagen dada *d*, tenemos un vector de probabilidades

$$(p(z_1/d),.p(z_2/d),...,p(z_K/d))$$

De donde, podemos calcular el índice de Entropía de Shannon de la imagen d, mediante

$$H(d) = -\sum_{i=1}^{K} p(z_i/d)\log(p(z_i/d)$$

De esta manera una imagen que esté asociada a un único aspecto, es decir, una imagen con vector de probabilidades con todo ceros excepto un uno, su valor de entropía será mínimo e igual a H(d)=0, contrariamente, una imagen que esté asociada por igual a todos los aspectos, es decir, con vector de probabilidades con 1/10 en cada componente, su valor de entropía será máximo e igual a H(d)=2.3026.

Los rangos de entropía teóricos respecto a 10 aspectos irían de 0 a 2.3026. Los observados en nuestra muestra van de prácticamente 0 a 2,17. Las imágenes que tienen una entropía elevada son aquellas que el procedimiento ha asociado de manera equiprobable a cada uno de los aspectos.

De esta forma se decide seleccionar del total de la muestra las imágenes con un valor de entropía superior a 1,4 y repetir de nuevo la búsqueda de aspectos en este nuevo conjunto formado por 1.482 imágenes. Se repite de nuevo todo el proceso generando los descriptores locales, el vocabulario visual y se intenta así que el sistema sea capaz de establecer nuevas relaciones entre imágenes visualmente más complejas dando lugar a nuevos aspectos latentes distintos de las 10 primeros. La prueba resulta un éxito y se generan otro conjunto distinto de 10 aspectos sobre la nueva muestra. En total el sistema es capaz de categorizar en 20 grupos el total de imágenes analizadas y estos son los resultados que pasaremos a discutir en el resto del capítulo.

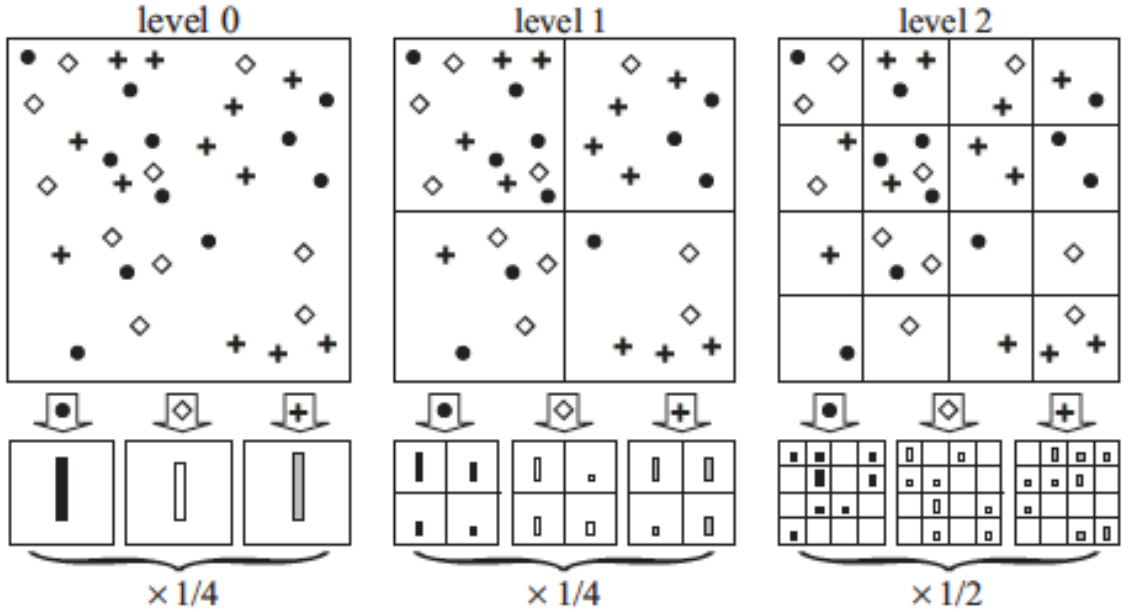


Figura A. 21. Ejemplo ilustrativo de la construcción de una pirámide de tres niveles. La imagen tiene tres tipos de entidades, indicadas por círculos, diamantes, y cruces. En la parte superior, se subdivide la imagen en tres niveles diferentes de resolución. A continuación, para cada nivel de resolución y cada canal, contamos las características que se encuentran en cada bin espacial. Esquema extraído de *Lazebnik et al. (2006).*

9. DESCRIPTOR *PHOW* (Pyramid Histogram Of visual Words)

Para superar las limitaciones del enfoque *BoW,* Lazebnik et al. (2006) proponen un método basado en la pirámide espacial de coincidencias de Grauman & Darrel (2005), que incorpora con éxito información espacial al modelo *BoW*. Se denomina *PHOW (*Pyramid Histogram Of visual Words).

En nuestro trabajo hemos implementado esta metodología (Vedaldi et al., 2008) de histogramas en pirámide que consiste en la colocación de una secuencia de rejillas cada vez más finas sobre la imagen, y en la obtención de una suma ponderada del número de coincidencias que se producen en cada nivel de resolución (L). (Fig. A.21)

Dada una resolución fija, se dice que dos puntos coinciden si están en el mismo bin de la rejilla; las coincidencias encontradas en resoluciones más finas se ponderan más alto que las coincidencias encontradas en resoluciones más gruesas.

Más específicamente, sean *X* e *Y* dos conjuntos de vectores en un espacio de características *p*-dimensional. Vamos a construir una secuencia de rejillas en resoluciones de *0, ..., L* tal que la rejilla en el nivel *l* tenga 2^l celdas a lo largo de cada dimensión, para un total de celdas $D = 2^{p l}$. H_X^l y H_Y^l denotan los histogramas de *X* e *Y* a esa resolución, por lo tanto $H_X^l(i)$ y $H_Y^l(i)$ son el número de puntos de *X* e *Y* que caen dentro de la celda *i-ésima* de cada rejilla. Entonces, el número de coincidencias en el nivel *l* viene dado por la función de *histograma de intersección:*

$$\mathcal{I}(H_X^l, H_Y^l) = \sum_{i=1}^{D} min\,(H_X^l(i), H_Y^l(i)).$$

Con la intención de resumir utilizaremos la abreviatura $\mathcal{I}(H_X^l, H_Y^l) = \mathcal{I}^l$. Nótese que el número de coincidencias encontradas en el nivel *l* también incluye todas las coincidencias encontradas en el nivel más fino *l+1*. Por lo tanto, el número de nuevas coincidencias encontrado en el nivel *l* viene dado por $\mathcal{I}^l - \mathcal{I}^{l+1}$ para *l = 0,...,L -1*. El peso asociado con el nivel *l* se establece con $\frac{1}{2^{L-l}}$, que es inversamente proporcional al ancho de la celda en ese nivel. Intuitivamente, se intenta penalizar las coincidencias encontradas en celdas más

grandes porque implican características cada vez más disímiles. Al poner todas las piezas juntas, el *Kernel* de pirámide de coincidencias (Grauman et al., 2005) está definido por:

$$\mathcal{K}^L(X,Y) = \frac{1}{2^L}\mathcal{I}^0 + \sum_{l=1} L \frac{1}{2^{L-l+1}}\mathcal{I}^l$$

Como se introdujo en (Grauman et al., 2005) un *Kernel* de pirámide de coincidencias trabaja con una representación de la imagen poco ordenada. Esto permite una comparación precisa de los descriptores de las imágenes en un espacio de gran dimensión, pero descartando la información espacial. Lazebnik et al. (2006) defienden un enfoque que tiene la ventaja de mantener la continuidad con el popular paradigma del *vocabulario visual*. Realiza la pirámide de coincidencias en el espacio de imagen bi-dimensional y utiliza técnicas tradicionales de agrupación en el espacio de características. En concreto, se cuantizan todos los vectores de características en un conjunto de *M* tipos discretos, palabras visuales, y asume la simplificación de que sólo las características del mismo tipo pueden ajustarse unas a las otras.

Cada canal *m* nos proporciona dos conjuntos de vectores bi-dimensionales, X_m y Y_m, representando las coordenadas de características de tipo *m* encontradas en las respectivas imágenes. El *Kernel* final es entonces la suma de *Kernels* de canales separados:

$$K^L(X,Y) = \sum_{m=1}^{M} \mathcal{K}^L(X_m, Y_m) \quad (1)$$

La pirámide espacial resultante es una extensión de la representación de la imagen de *Bag-of-Words,* equivaldría a un *Bag-of-Words* normal para L = 0.

El *Kernel* de pirámide de coincidencias es simplemente una suma ponderada de intersecciones de histogramas, y dado que *c* min*(a,b)* = min *(ca,cb)* para números positivos, podemos implementar (1) como una intersección de histograma simple de vectores *largos* formados por la concatenación de histogramas apropiadamente ponderados de todos los canales de todas las resoluciones. Para *L* niveles y *M* canales el vector resultante tiene una dimensionalidad $M\sum_{l}^{L} 4^l = M\frac{1}{3}(4^{L+1}-1)$.

Resumiendo, Lazebnik et al., (2006) extienden el *Kernel* de pirámide de coincidencias a la pirámide de histogramas de palabras visuales. Bosch, Zisserman & Muñoz (2007) implementan una pirámide de histogramas de palabras visuales inspirado en el anterior esquema de coincidencias espaciales pero usando un gaussiano como *Kernel*. En esta implementación de similitud entre un par de imágenes *I* y *J* se calcula usando una función *Kernel* entre su pirámide de histogramas de palabras visuales D_I y D_J, con adecuada ponderación de cada nivel de la pirámide**:**

$$K(D_I, D_J) = \exp\left\{\frac{1}{\beta}\sum_{l \in L} \alpha_l d_l(D_I, D_J)\right\}$$

donde β es el promedio de $\sum_{l \in L} \alpha_l d_l(D_I, D_J)$ sobre los datos de entrenamiento, α_l es el peso en el nivel *l* y d_l es la distancia χ^2 entre D_I y D_J (Zhang, Marszałek, Lazebnik & Schmid, 2007) en el nivel de la pirámide *l* calculado usando los histogramas normalizados en este nivel.

Los histogramas espaciales podrían ser usados como descriptores de imagen y alimentar un *SVM* lineal. Los *SVM* lineales son mucho más rápidos de entrenar pero también están limitados a usar un producto interno para comparar descriptores. Vedaldi y Zisserman (2010) han demostrado que se pueden obtener mejore resultados calculando un mapa de características explícito que emule un *Kernel*-χ^2 no lineal como uno lineal.

ANEXO B

TERMINOLOGÍA

Daremos una definición para evitar confusiones con ciertos términos relacionados con la clasificación de imágenes que se utilizan en la tesis y en la literatura:

Anotación de imágenes (escenas u objetos): También se denomina etiquetado. Consiste en identificar manualmente con una etiqueta a una imagen en función de los elementos que contenga.

Aprendizaje supervisado: Un método de aprendizaje se denomina supervisado si necesita el etiquetado manual del conjunto utilizado para el entrenamiento del sistema.

Aprendizaje no supervisado: En un método de aprendizaje no supervisado, no se proporciona ninguna información acerca de la etiqueta que le pertenece a cada imagen al conjunto de entrenamiento.

Bag-of-Words (BoW): A veces también se denomina Bag-of-features, Bag of visualterms (BoV) o Bag-of-visterms. La imagen se representa como una bolsa de características representativas. De un modo general se refiere al histograma de palabras visuales de una imagen.

Categoría: Conjunto visualmente consistente de imágenes.

Categoría de la imagen: Etiqueta que tiene el total de la imagen. Por ejemplo categoría *ciudad* para una imagen que contiene edificios.

Clasificación de imágenes: Consiste en agrupar un conjunto de imágenes en función de los elementos que contengan.

Clustering: Algoritmos y métodos de agrupación de objetos en categorías.

Modelo discriminativo: Tipo de planteamiento para clasificación de escenas o reconocimiento de objetos que trata las anotaciones de las imágenes como clases y emplea clasificadores entrenados para obtener fronteras que permitan discriminar entre aquellas imágenes en las que aparece un concepto y las que no.

Modelo generativo: Tipo de planteamiento para clasificación de escenas o reconocimiento de objetos que, a diferencia del modelo discriminativo, que sólo diferencia entre casos positivos y negativos, estos métodos tratan de inferir las probabilidades conjuntas entre imágenes y anotaciones. Esta información, si bien más compleja de obtener, proporciona un conocimiento extra sobre la generación de los datos. Introducen variables latentes que asocian a los conceptos semánticos de las etiquetas.

Palabra visual: Es la analogía del término *palabra* en el análisis de textos. Denota partes informativas específicas de una imagen.

Patch: Pequeños fragmentos locales de una imagen.

Procesamiento bottom-up: (De abajo a arriba) Es una estrategia de procesamiento de información. En el diseño bottom-up las partes individuales se diseñan con detalle y luego se enlazan para formar componentes más grandes, que a su vez se enlazan hasta que se forma el sistema completo.
Procesamiento top-down: (De arriba a abajo) En el modelo top-down se formula un resumen del sistema, sin especificar detalles. Cada parte nueva es entonces redefinida, cada vez con mayor detalle, hasta que la especificación completa es lo suficientemente detallada.

Segmentación de la imagen: Es el proceso de asignación de una etiqueta a cada píxel de la imagen de forma que los píxeles que compartan la misma etiqueta también tendrán ciertas características visuales similares.

Vocabulario visual: Está compuesto por un conjunto de palabras visuales.

ABREVIACIONES

A continuación, les resumimos las abreviaturas utilizadas en la tesis:

BoW: Bag of Words. Bolsa de palabras (en nuestro caso palabras visuales)**.**

BRAIN: Brain Research through Advancing Innovative Neurotechnologies (Proyecto EEUU).

CBIR: Content Based Image Retrieval. Recuperación de imagen basada en su contenido visual.

DoG: Difference-of-Gaussian.

GIST: El descriptor de GIST (traducido al castellano, gist se refiere a la esencia, el contexto de la escena) se propuso inicialmente por Oliva & Torralba (2001). La idea es desarrollar una representación de la totalidad de la imagen. Los autores proponen un conjunto de dimensiones perceptibles (naturalidad, apertura, rugosidad, grado de expansión, de robustez) que representan la estructura espacial dominante de una escena. Ellos demuestran que estas dimensiones se pueden estimar de forma fiable utilizando la información espectral y toscamente localizada de las imágenes.

GLCM: Gray-Level Co-occurrence Matrix.

HBP: Human Brain Project (Proyecto e la Unión Europea).

HOG: Histogram of oriented gradients.

IA: Inteligencia Artificial.

MATLAB: Es el lenguaje de alto nivel comercializado por Mathworks con entorno interactivo utilizado para desarrollar algoritmos, analizar y visualizar datos, y realizar cálculos numéricos.

ME: Prefijo que se antepone a los aspectos Más Entrópicos resultantes del análisis con *pLSA*. Por ejemplo; ME1 es el aspecto más entrópico número 1.

PE: Prefijo que se antepone a los aspectos Poco Entrópicos resultantes del análisis con *pLSA*. Por ejemplo; PE1 es el aspecto poco entrópico número 1.

PHOW: Pyramid Histogram Of visual Words. Histograma en pirámide de palabras visuales.

pLSA: probabilistic Latent Semantic Analysis. Anñalisis probabilístico de aspectos latentes.

SIFT: Scale-Invariant Feature Transform. Algoritmo de detección de características invariante tanto a rotaciones como al escalado de las imágenes.

SVM: Support Vector Machine. Máquina de vector de soporte.

ÍNDICE DE FIGURAS

1 - INTRODUCCIÓN

2- METODOLOGÍA

3 - RESULTADOS

BIBLIOGRAFÍA

Alvaro, S. (2013). *Big Data y Humanidades Digitales: de la computación social a los restos de la cultura.* Recuperado el 3 de Diciembre de 2014, de http://blogs.cccb.org/lab/es/article_big-data-i-humanitats-digitals-de-la-computacio-social-als-reptes-de-la-cultura-connectada/

Arnheim, R. (1966). *Towards a Psychology of Art: Collected Essays.* University of California Press. Recuperado el 10 de Mayo de 2015, de https://books.google.es/books?id=0vNcIQSUf0UC&pg=PA39&lpg=%23v=onepage&q&f=false#v=onepage&q&f=false

Arnheim, R. (1980). *Hacia una psicología del arte. Arte y entropía*. Madrid: Alianza forma.

Arnheim, R. (1983). *Arte y percepción visual.* Madrid: Alianza forma.

Arnheim, R. (1986). *El pensamiento visual*. Barcelona: Ediciones Paidós.

Aznar, J.A (2015) *Psicología de la percepción visual*. Recuperado el 1 de Octubre de 2014, de http://www.ub.edu/pa1/node/121

Aznar, J.A. & Moreno, M. (2011). *Neurocomputacion en el Sistema Visual Humano*. Editorial Académica Española (LAP LAMBERT Academic Publishing GmbH & Co. KG.).

Bajcsy P. & Moslemi, M. (2010). Discovering Salient Characteristics of Authors of Art Works, *Proc. SPIE 7531, Computer Vision and Image Analysis of Art*, 75310B. Doi: 10.1117/12.838847

Barnard, K., Duygulu, P., Forsyth, D., Freitas, N., Blei, D.M. & Jordan, M.I. (2003). Modeling words and pictures. J*ournal of Machine Learning Research special issue on machine learning methods for text and images, 3*, 1107–1135.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society, 35,* 99–110.

Berger, J. (2002). *Modos de ver.* Barcelona: Gustavo Gili.

Berry, D.M. (2011) 'The Computational Turn: Thinking about Digital Humanities'. *Culture Machine, 12*. Recuperado el 1 de Enero de 2015, de http://www.culturemachine.net/index.php/cm/issue/view/23

Bharati, M.H., Liu, J.J. & MacGregor, J.F. (2004). Image texture analysis: methods and comparisons. *Chemometrics and Intelligent Laboratory Systems, 72*, 57-71.

Blei, D.M, Ng, A.Y., Jordan, M.I. & Lafferty, J. (2003). Latent Dirichlet Allocation, *Journal of Machine Learning Research, 3*, 993-1022.

Bosch, A., Zisserman, A. & Muñoz, X. (2006) Scene classification via pLSA. *Lecture Notes in Computer Science (3954)*, 517-530. Doi: 10.1007/11744085_40

Bosch, A. (2007). *Image Classification for a largre number of object categories*. Tesis no publicada, Universitat de Girona, España.

Bosch, A., Zisserman, A. & Muñoz, X. (2007). Image Classification using Random Forests and Ferns. *In Proceedings of IEEE International Conference on Computer Vision (ICCV)*. Doi: 10.1109/ICCV.2007.4409066

Boser, B.E., Guyon, I.M. & Vapnik, V.N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Conference on Computational Learning Theory,* 144-152.

Cairo, A. (2011). *El arte funcional. Infografía y visualización de información*. Madrid: Alamut.

Coddington, J., Elton, J., Rockmore, D. & Wang, Y. (2008). Multifractal analysis and autenthication of Jackson Pollock's paintings. *Computer image analysis in the study of art. 6810,* 68100D-1-8.

Cover, T.M. & Thomas, J.A. (2006). *Elements of Information Theory (2on ed.)*. New Jersey: John Wiley & Sons.

Chellappa, R. & Jain, A. K. (1993). *Markov Random Fields: Theory and Ap- plications*. New York: Academic.

Daucher, H. (1978). *Visión artística y visión racionalizada*. Barcelona: Gustavo Gili.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc. B, 39*, 1–38.

Dondis, D. A. (1984) *La sintaxis de la imagen: introducción al alfabeto visua*l. Barcelona: Gustavo Gili.

Enebral, J. (2009). *Detección y asociación automática de puntos característicos para diferentes aplicaciones*. Trabajo Final de Carrera. Escola Politècnica Superior de Castelldefels. Universitat Politècnica de Catalunya.

Emmer, M. (2005). La perfección visible: matemática y arte. *Artnodes, 4,* 1-8. Doi: http://dx.doi.org/10.7238/a.v0i4.731

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99*(14), 9596–9601.

Fei-Fei, L. & Perona, P. (2005). A Bayesian hierarchical model for learning natural scene categories. *In Proc. CVPR*. San Diego, CA, USA**.**

Flickner, M. & al. (1995) Query by image and video content: the QBIC system. *IEEE Computer 28(9)*, 23-32. Doi: 10.1109/2.410146

Flusser, V. (2002). *Filosofía del diseño.* Madrid: Síntesis.

Flusser, V. (2009). *Una filosofía de la fotografía*. Madrid: Síntesis.

Fontcuberta, J. (2010). *La cámara de pandora. La fotografí@ después de la fotografía*. Barcelona: Gustavo Gili.

Forbes, N. (2005) *Imitation of life. How Biology Is Inspiring Computin*g. Cambridge: The MIT Press.

Garcin, G. (2013). *Gilbert Garcin*. Recuperado el 10 de Febrero de 2015, de http://www.gilbert-garcin.com

Gombrich, E.H. (1998). *Arte e ilusión. Estudio sobre la psicología de la representación pictórica*. Madrid: Editorial Debate.

Gombrich, E.H. (2007). *Arte, percepción y realidad*. Barcelona: Ediciones Paidós.

Gonzalez, R.C., Woods, R.E. (2008). *Digital Image Processing, 3rd Ed.*, New Jersey: Prentice- Hall.

Gordon, A. D. (1999), *Classification*, Boca Raton: Chapman and Hal.

GuggheimBilbao (2015). De Durero a Rauschenberg: la quintaesencia del dibujo. Obras maestras de las colecciones Albertina y Guggenheim. Recuperado el 8 de Noviembre de 2015, de http://www.guggenheim-bilbao.es/exposiciones/de-durero-a-rauschenberg-la-quintaesencia-del-dibujo-obras-maestras-de-las-colecciones-albertina-y-guggenheim/

Grauman, K. & Darrel, T. (2005). The pyramid match kernel: Discriminative classification with sets of image features. *In Proceedings of IEEE International Conference on Computer Vision (ICCV),* Beijing.

Guasch, A. M. (1997). *El arte del siglo XX en sus exposiciones. 1945-1995.* Barcelona: Serbal.

Haralick, R. M., Shanmugan, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics, 3*(6), 610-621.

Haralick, R.M. & Shapiro L.G. (1991). *Computer and Robot Vision: Vol. 1*, Boston: Addison-Wesley.

Hegerath, A., Deselaers, T. & Ney, H. (2006). Patch-based object recognition using discriminatively trained gaussian mixtures, *In British Machine Vision Conference (2)*, 519–528. Doi: 10.5244/C.20.54

Hildebrand, A. von (1988). *El problema de la forma en la obra de arte*. Madrid: Visor.

Hind, Ch.L. (eBooks@Adelaide) (2014). *Drawings of Leonardo da Vinc*i. Recuperado el 4 de Enero de 2015, de https://ebooks.adelaide.edu.au/l/leonardo_da_vinci/drawings/index.html

Hockey, S. (2004). *The History of Humanities Computing*. Oxford: Blackwell.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning, 42*,177-196.

Huttenlocher, D.P., Klanderman, G.A., & Rucklidge, W.J. (1993). Comparing images using the hausdorff distance, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*(9), 850–863.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *In Proceedings of the European Conference on Machine Learning. Springer-Verlag.*

Johnson, M.K., Stork, D.G., Biswas, S. & Furuichi, Y. (2008). Inferring illumination direction estimated from disparate sources in paintings: An investigation into Jan Vermeer's Girl with a pearl earring. *Proc. of SPIE-IS&T Electronic Imaging, SPIE (6810),* 1-12.

Kandinsky, V. (1987). *La gramática de la creación. el futuro de la pintura*. Barcelona: Paidós.

Kandinsky, V. (1996). *Punto y línea sobre el plano. Contribución al análisis de los elementos pictóricos.* Barcelona: Paidós.

Koffka, K. (1967). *Principles of Gestalt Psychology*, Mimesis International.

Köhler, W. (1947). *Gestalt psychology: an introduction to new concepts in modern psychology*. New York: Liveright.

Kurzweil, R. (2013). *Cómo crear una mente. El secreto del pensamiento humano*. Berlín: Lola Books GbR.

Laptev I.. (2009) Improvements of object detection using boosted histograms. *Image and Vision Computing 27*(5), 535–544. Doi:10.1016/j.imavis.2008.08.010

Lazebnik, S., Schmid, C. & Ponce, J. (2006). Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2,* 2169-2178. Doi: doi.ieeecomputersociety.org/10.1109/CVPR.2006.68

Leung, T. & Malik, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision,43*(1), 29–44.

Levi, P. (1987). *La machine univers. Creation, cognition et culture informatique*. Paris: La Découverte.

Li, J.,Gray, R. M. & Olshen R. A. (2000). Multiresolution image classification by hierarchical modeling with two dimensional hidden Markov models, *IEEE Trans. Inform. Theory, 46*, 1826–1841.

Li, J. & Wang J.Z. (2003). Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*(9), 1075-1088.

Li, J. & Wang J.Z. (2004). Studying Digital Imagery of Ancient Paintings by Mixtures of Stochastic Models, *IEEE Transactions on Image Processing 13*(3), 340-353.

Li, J. & Wang, J.Z. (2006). Real-Time computerized annotation of pictures. *Proceedings of the ACM International Conference on Multimedia*.

Lowe, D.G. (2000). Towards a Computational Model for Object Recognition in IT Cortex. *Biologically Motivated Computer Vision,* 20-31.

Lowe, D. G. (2004). Distinctive Image Features from Scale Invariant Keypoints. *Int. Journal of Computer Vision, 60,* 2, 91-110.

Manovich, L. (1996). *The Automation of Sight: From Photography to Computer Vision*, En Electronic Culture: Technology and Visual Representation. New York: Ed. Timothy Druckrey, 229-239.

Manovich, L. (2005). *El lenguaje de los nuevos medios de comunicación: la imagen en la era digital*. Barcelona: Paidós.

Manovich, L. (2012). ¿Cómo ver 1000000 de imágenes?. *Deforma cultura on line*. 1-11. http://www.deforma.info/es/product.php?id_product=24

Manovich, L. (2015). *Manovich*. Recuperado el 9 de Mayo de 2015, de http://manovich.net

MATLAB (2011). The Mathworks. http://www.mathworks.com/products/matlab/ (disponible on-line).

Màquina de vector de suport. (2014). Viquipèdia, l'Enciclopèdia Lliure. Recuperado el 8 de Novembre de 2014, de //ca.wikipedia.org/w/index.php?title=M%C3%A0quina_de_vector_de_suport&oldid=14264178

Marr, D. & Poggio, T. (1976). Cooperative Computation of Stereo Disparity, Science,(194), 4262, 283-287. Doi: 10.2307/1742217

Marr, D. E. (1982). *Vision*. San Franciso: Freeman.

Marsland, S. (2009). *Machine learning. An algorithmic perspective*. New York: CRC Press.

Melzer T., Kammerer P. & Zolda E. (1998). Stroke Detection of Brush Strokes in Portrait Miniatures Using a Semi-Parametric and a Model Based Approach, *ICPR 1998, Pattern Recognition International Conference*, 474. Doi: 10.1109/ICPR.1998.711184

Méndez, M.T. (1992). *La mirada inútil*. Madrid: Julio Ollero.

Mitchell, T. (1997). *Machine Learning*. New York: McGraw-Hill.

Mojsilovic, A., Gomes, J., & Rogowitz, B. (2002). Isee: Perceptual features for image library navigation. *In SPIE: Human vision and electronic imaging, 4662*, 266–277.

Mumford, D. (2002). Pattern theory: The mathematics of perception. ICM (3), 1- 21.

Nogué, A. (2013). *Dibuixar un arbre / Drawing a tree*. Barcelona: Comanegra.

Núñez, M. (2008). *Marina Núñez*. Recuperado el 6 de Marzo de 2014, de http://www.marinanunez.net

Oliva, A. & Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145–175. Doi: 10.1023/A:1011139631724

Pinto, A.C. (2006). *Segmentación de imágenes por textura*. Tesis no publicada, Universidad de Concepción, Chile.

Planas, M. A. (2014). *Miquel Planas*. Recuperado el 2 de Enero de 2014, de http://www.miquelplanas.eu

Plensa, J. (2008). *Jaume Plensa*. Recuperado el 4 de Octubre de 2015, de http://jaumeplensa.com

Quelhas, P., Monay, F., Odobez, J.-M., Gatica-Perez, D., Tuytelaars & Van Gool, L. (2005). Modeling scenes with local descriptors and latent aspects. *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), 1*, 883-890. Doi: 10.1109/ICCV. 2005. 152

Quelhas, P. (2007). *Scene Image Classification and Segmentation with Quantized Local Descriptors and Latent Aspect Modeling,* Tesis no publicada, École Polytechnique Fédérale de Lausanne, Suiza.

Pérez, D. (1995). Modelos psicofísicos de discriminación de texturas visuales: evolución y aspectos críticos. *Anuario de Psicología 64*. 1-19.

Reguera, I. (2010). *Aby Warburg, inventor del museo virtual*. Recuperado el 2 de Febrero de 2015, de http://elpais.com/diario/2010/05/01/babelia/1272672757_850215.html

Reverter, F., Figueras, E., Planas, M.A. & Rosado, P. (2012). *Ideación y catalogación artística basada en métodos de visión artificial*. Barcelona: Raima.

Rockmore, D., Lyu, S., & Farid, H. (2006). A digital technique for authentication in the arts, *International Foundation for Art Research (IFAR) Journal, 8*(2), 21–29.

Ruiz, A. (2011). *Comportamiento y análisis de descriptores de texturas en imágenes modis*. Trabajo fin de máster en sistemas inteligentes. Facultad de Informática. Universidad Complutense de Madrid.

Rothko, M. (2004). *La realidad del artista. Filosofías del arte*. Madrid: Editorial Síntesis.

Ruhrberg, K., Schneckenburger, M., Fricke, C & Honnef, K. (2001). *Arte del siglo XX.* Köln: Taschen.

Russ, J.C. (1999). *The Image Processing Handbook, 3rd edition,* Florida: CRC Press.

Shen, J. (2009). Stochastic modeling western paintings for effective classification, *Pattern Recognition, 42*(2), 293–301.

Sivic, J. & Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. *In International Conference on Computer Vision, 2*, 1470–1477.

Sivic, J., Russell, B.,Efros, A.,Zisserman, A. & Freeman, W. (2005). Discovering objects and their location in images. *In International Conference on Computer Vision, (1)*, 370–378.

Smeulders, A., Worring, M., Santini, S., Gupta, A. & Jain, R. (2000). Content based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on , (22)*, 1349-1380.

Svoboda, T., Kybic, J., & Hlavac, V. (2008). *Image processing, analysis, and machine visión. A MATLAB companion.* Madrid: Paraninfo.

Stork, D.G. (2006). Computer vision, image analysis and master art: Part I, *IEEE Multimedia, 13*(3), 16–20.

Stork, D.G., & Johnson, K. (2006). Computer vision, image analysis and master art, Part II: Finding the illuminant in realist paintings', *IEEE Multimedia, 13*(4), 12-17.

Stork, D.G., & Duarte, M. (2007). Computer vision, image analysis and master art, Part III: Quantifying shape in realist art, *IEEE Multimedia, 14* (1), 14–18.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science, (262)*, 685–688.

Tanaka, K. (1997). Mechanisms of visual object recognition: monkey and human studies. *Current Opinion in Neurobiology, (7),* 523–529.

Tàpies, A. (2001). *Fundació Antoni Tàpies*. Recuperado el 12 de Enero de 2015, de http://www.fundaciotapies.org/site/spip.php?rubrique65

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

Torralba, A., Murphy, K.P., Freeman, W.T. & Rubin, M.A. (2003). Context-based vision system for place and object recognition. *In Proc. ICCV (9)*. 273-280.

Treisman, A. & Gelade, G. (1980). A feature-integration theory of atention. *Cognitive Psychology, 12,* 97–136.

Vedaldi, A & Fulkerson, B. (2008). *VLFeat - An open and portable library of computer vision algorithms*. Retrieved from http://www.vlfeat.org.

Vedaldi, A., & Zisserman, A. (2010). Efficient additive kernels via explicit feature maps. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*

Vogel, J. & Schiele, B. (2004). Natural scene retrieval based on a semantic modeling step. *CIVR, Dublin*, Ireland.

Vogel, J. & Schiele, B. (2007). Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision, 72*(2),133–157.

Von Goethe, J.W. (1970). *Versuch die Metamorphose der Pflanzen zu erklären. Gotha:* Ettingersche Buchhandlung.

Wagensberg, J. (1985). *Ideas sobre la complejidad del mundo*. Barcelona: Tusquets editores.

Wagensberg, J. (2007). *El gozo intelectual. Teoria y práctica sobre la inteligibilidad la belleza*. Barcelona: Tusquets editores.

WahooArt (n.d). *Estudio de caballos 1*. Recuperado el 6 de Septiembre de 2014, de http://es.wahooart.com/@@/8EWLC4-Leonardo-Da-Vinci-Estudio-de-los-caballos-1

Wang, J.Z., Li, J., Gray, R.M., and Wiederhold, G. (2001). Unsupervised multiresolution segmentation for images with low depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(1), 85-90.

Wang, J.Z., Li J. & Wiederhold, G., (2001). SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(*9), 947-963.

Warburg, A. (2010). Atlas Mnemosyne. Madrid: Akal.

Warburg, A. (2003) *Der Bilderatlas Mnemosyne*. Berlín: Martin Warnke.

Weber, M., Welling, M. & Perona, P. (2000). Unsupervised Learning of Models for Recognition. *Lecture Notes in Computer Science. (1842)*, 18-32.

Wentworth, D. (1980). *Sobre el crecimiento y la forma*. Madrid: H.Blume.

Wiener, N. (1988). *Cibernética y sociedad*. Buenos Aires; Editorial Sudamericana.

Wiener, N. (1998). *Cibernética o El control y comunicación en animales y máquinas*. Barcelona; Tusquets.

Willamowski, J., Arregui, D., Csurka, G., Dance, C., & Fan, L. (2004). Categorizing nine visual classes using local appearance descriptors. *In Proceedings of LAVS Workshop, in ICPR'04,* Cambridge.

Zambrano, M. (1989). *Algunos lugares de la pintura*. Madrid: Espasa Calpe.

Zhang, H., Berg, A., Maire, M. & Malik, J. (2006). SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2*, 2126–2136.

Zhang J., Marszałek, M. Lazebnik, C. & Schmid, S. (2007). Local features and kernels for classification of texture and object categories: a comprehensive study. *International Journal of Computer Vision, 73*(2), 213–238. Doi: 10.1007/s11263-006-9794-4