Optical Security and Authentication using Nanoscale and Thin Film structures

ARTUR CARNICER,[1] BAHRAM JAVIDI[2,*]

[1]Universitat de Barcelona, Departament de Física Aplicada, Martí i Franquès 1, 08028 Barcelona, Catalunya (Spain)
[2]University of Connecticut, Electrical & Computer Engineering Dept.371 Fairfield Way, Storrs, CT 06269 (USA)

*Corresponding author: Bahram.Javidi@uconn.edu

Authentication of encoded information is a popular current trend in optical security. Recent research has proposed production of secure unclonable ID tags and devices with the use of nanoscale encoding and thin film deposition fabrication techniques which are nearly impossible to counterfeit but can be verified using optics and photonics instruments. Present procedures in optical encryption provide secure access to the information and these techniques are improving daily. Nevertheless, a rightful recipient with access to the decryption key may not be able to validate the authenticity of the message. In other words, there is no simple way to check whether the information has been counterfeited or not. Metallic nanoparticles may be used in the fabrication process because they provide distinctive polarimetric signatures that can be used for validation. The data is encoded in the optical domain that can be verified using physical properties with speckle analysis or ellipsometry. Signals obtained from fake and genuine samples are complex and can be difficult to distinguish. For this reason, machine learning classification algorithms are required in order to determine the authenticity of the encoded data and verify the security of unclonable nano particle encoded or thin film based ID tags. In this paper, we review recent research on optical validation of messages, ID tags, and codes using nano structures, thin films, and 3D optical codes. We analyze several case scenarios where optically encoded devices have to be authenticated. Validation requires the combined use of a variety of multi-disciplinary approaches in optical and statistical techniques and for this reason, the first five sections of this paper are organized as a tutorial. © 2016 Optical Society of America

1. Introduction

The publication of the two seminal papers in optical security in 1994 and 1995 initiated the creation of a new sub-discipline within optics [1, 2]. Twenty years later, papers published in the field of optical security and encryption are numbered and downloaded in thousands [3-6]. Basic concepts behind optical encryption are relatively straightforward since they are based on the use of a 4f coherent processor: the data to be encrypted (plain-text) is phase encoded using phase-only random distribution M_1; a second random phase mask M_2 is located at the Fourier plane of the first lens. When the system is illuminated by means of a laser source, the propagated field at the recording plane (i.e. the Fourier plane of the second lens) is pseudo-random and noise like. The original plain-text can be accessed from the complex cypher-text (not the intensity), provided the conjugate of phase mask M_2 is known.

In order to improve the encryption capabilities and security, multiple enhancements and variations of the original Double Random Phase Encryption setup have been proposed. For instance, encryption systems have been suggested that work in Fresnel domain [7], fractional-Fourier domains [8]; wavelength multiplexing [9] or incoherent light [10] are

also notable design variables. Special mention deserves those systems based on polarization since they dramatically increase the degrees of freedom of the encryption processor [11-14]. Nevertheless, it was proved that the original Double Random Phase Encryption design was flawed due to the linear character of optical transformations involving light propagation [15, 16]. Multiple alternatives have been suggested in order to make optical security systems more robust [17-21]. For instance, the so-called digital optics methods represent a very popular approach in optical-based encryption [22-27]. These procedures mimic physical processes by means of digital calculations or combine numerical operations with optical procedures. Related to these methods, phase-truncation asymmetric keys [28, 29] or optical designs aimed to break symmetries [30] have been proposed. On the other hand, systems that work in low light conditions are being used in security. These arrangements use quantum sources that provide very few photons. Despite operating in photon-starved conditions which makes the decryption very low light level, validation of the authenticity of the message can be achieved. In those scenarios where it is not required to access the information, photon counting encryption is a very good and secure alternative [31-39]. Other designs based on computational ghost imaging have proposed [40, 41].

Most of the optical encryption methods rely on the use of holographic encoding of information. A very recent approach on optical security is based on including an extra security layer in order to authenticate messages. Information is encoded using conventional techniques but the recipient may be able to determine whether the information was sent by a trusted party or not. The physical support used for the message plays a key role in this problem. The recipient has to be able to read or decrypt the data but also can test the material using optical analytical tools in order to validate the authenticity of the sender. For instance, bank notes printed using inks containing certain amounts of nanoparticles are difficult to be counterfeited [42]. Another possible scenario where validation is a critical issue is the electronic industry. A label with the specifications of an integrated circuit is attached to the microchip package. This label can be phase-encoded or produced with metal nanoparticles or thin-films. The user can rely on the contents of the information provided if the response to certain optical measurements (e.g. speckle statistics, polarimetric signatures, etc.) delivers the correct signal [43-45]. Gold nanoparticles (AuNP) can be used in the fabrication of pharmaceutical pills and tablets for consumer safety and protection. These samples become polarimetrically tagged and thus, authentication can be performed using ellipsometry [46].

As explained above, the message has to be encoded on a device that can be authenticated using optical technologies. Accordingly, production of these codes may require the use of fabrication techniques in the nanoscale such as nanostructures [47-49] or thin film deposition [50]. Regarding materials, metallic nanoparticles, and more specifically, AuNP, are suggested in the fabrication of optical codes because they provide distinctive polarimetric signatures that can be used for validation [51-55].

This paper aims to review several scenarios on security and authentication of codes tagged with nanoparticles, phase-encoded or produced using thin-film technologies. Since authentication of such structures requires the combined use of several techniques in different disciplines, sections 2 to 5 are organized as a tutorial. This includes topics such as Fourier optics, polarimetry, and statistical pattern classification using machine learning techniques. Accordingly, the paper is organized as follows: section 2 reviews Fourier optics and wave propagation theory. Within this framework, the behavior of phase encoded information after propagation is analyzed and related with speckle noise. Double Random Phase Encryption is considered as a special case of phase encoding. Section 3 describes basic concepts on polarized light including Stokes parameters and Mueller matrix formalism. Special emphasis is placed to the measurement of the Mueller matrix using ellipsometers. Section 4 describes the important concept of data analysis. Since multidimensional measurements are required for authentication of the samples, we introduce principal component analysis as a method to reduce dimensionality in complex

problems. Section 5 surveys three methods in machine learning classification used in this review: K-Nearest Neighbors, Multiclass Support Vector Machine and Random Forests. The second part of the paper reviews several experiments on authentication for security. Section 6 analyzes how to distinguish among phase encoded QR codes made of gold nanoparticles or silver [44]: the samples are illuminated using polarized light and speckle noise distributions are recorded after propagation. A three-class support vector machine algorithm is used to perform authentication. A similar problem is studied in section 7. In this case, QR codes are produced using thin films of Cr and Ta_2O_5. The dataset is synthetically generated from the measurements of the Mueller matrix. Different scenarios are considered for authentication. For instance, the samples can be imaged using a microscope polarimeter. The information provided by the Mueller components is used to perform authentication by means of correlation. When the samples are phase encoded, the recorded Mueller components of the speckle signal is used to generate a large dataset of Stokes vectors that feeds a K-Nearest Neighbors classification system [45]. In the next section, we describe how to distinguish among pharmaceutical tablets that have been coated with gold nanoparticles of different sizes. In this case, we analyze the polarization signature of these tablets using a Mueller spectro-polarimeter [46]. In section 9, we train a system based on Random Forests classifier to authenticate 3D phase-encoded samples [56]. Finally, conclusions and future perspectives on this topic are presented in section 10.

2. Propagation of phase-encoded information

Scalar diffraction theory provides a simple way of describing the propagation of wave-front E [57]. Without loss of generality, it is assumed that the electric field propagates along the z axis. Then, the mathematical relationship between the propagated wave-front $E(x,y,z)$ and the beam $E(x,y,0)$ at reference plane $z=0$ is:

$$E(x, y, z) = \text{FT}^{-1}\left[\text{FT}\left[E(x, y, 0)\right]H(u, v, z)\right]$$

$$H(u, v, z) = \exp\left(i\frac{2\pi}{\lambda}z\sqrt{1-\lambda^2 u^2 - \lambda^2 v^2}\right).$$

(1)

In this formulae, operators FT and FT^{-1} stand for the Fourier and the inverse Fourier transforms respectively, $H(u,v,z)$ is the transfer function, (u,v) are the spatial frequencies, λ is the wavelength and i is the imaginary unit. Unit vector $\mathbf{s} = (\alpha, \beta, \gamma)$ indicates the normal direction to the wave-front at each point. Components (α, β, γ) are related to the spatial frequencies (u, v) by means of

$$\alpha = \lambda u, \ \beta = \lambda v, \ \gamma = \sqrt{1-\alpha^2 - \beta^2} = \sqrt{1-\lambda^2 u^2 - \lambda^2 v^2}.$$

(2)

Note that for $\alpha, \beta \ll 1$, the wave is paraxial, i.e. vector $\mathbf{s}$ forms small angles with propagation axis z. If this approximation is valid, transfer function $H(u,v,z)$ can be simplified and now reads $H(u, v, z) = \exp(ikz)\exp\left(-i\pi\lambda z\left(u^2 + v^2\right)\right)$. In this case, it is said that propagation is described by means of the **Fresnel diffraction approximation**.

Propagation of light using scalar diffraction theory is calculated by means the following formula:

$$E(x, y, z) = \text{FT}^{-1}\left[\text{FT}\left[E(x, y, 0)\right]\exp\left(\frac{2\pi}{\lambda}z\sqrt{1-\lambda^2 u^2 - \lambda^2 v^2}\right)\right]$$

In Fresnel diffraction conditions, propagation of light is calculated using paraxial approximation:

$$E(x, y, z) = \exp(ikz)\,\text{FT}^{-1}\left[\text{FT}\left[E(x, y, 0)\right]\exp\left(-i\pi\lambda z\left(u^2 + v^2\right)\right)\right]$$

Figure 1 shows an example on how Eq. (1) is used in practice. Figure 1(a) depicts a QR code encoding the message 0034934021143 (a telephone number). This distribution [$Q(x,y,0)$] is illuminated by a coherent light source. The variables used in the calculation are summarized in Table 1.

Table 1: Propagation variables	
Propagated distance	Δz=15 mm
Wavelength	λ=532 nm
Size of the code	0.84 x 0.84 mm²
Number of pixels	256x256 pixels

Figure 1(b) shows the propagated irradiance $I_Q(x,y,z)=|Q(x,y,z)|^2$. Note that the information displayed in this figure still resembles the image of the QR [Fig. 1(a)].
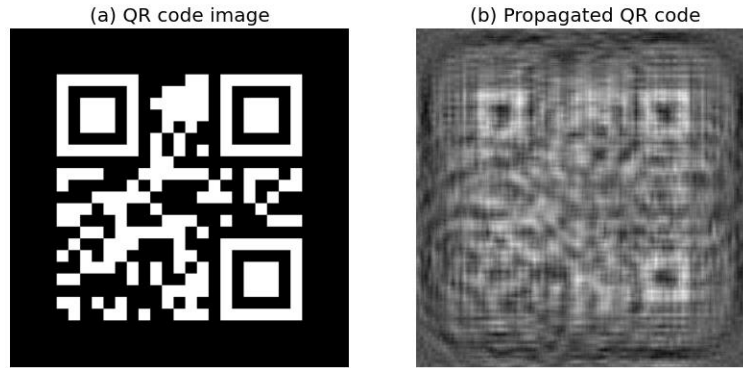

(a) QR code image   (b) Propagated QR code

Fig. 1: (a) $Q(x,y,0)$: reversed QR code encoding the message 0034934021143; (b) propagated irradiance $I_Q(x,y,z)=|Q(x,y,z)|^2$

Random phase-only distributions are commonly used in optical information processing problems (i. e. pattern recognition, optical encryption, etc). They are described by means of $R(x,y,0)=\exp(2\pi i \, \text{rand}(x,y))$, where $\text{rand}(x,y)$ is a function that is uniformly distributed random numbers over [0,1). **Phase encoding** is performed by attaching a diffusor such as adhesive tape to the sample. Note that thick diffusors are random in 3D by nature and thus, thick phase-masks cannot be duplicated.

The analysis of the behavior of propagated random phase-only distributions $R(x,y,0)$ provides interesting information. Figure 2(a) shows the irradiance $I_R(x,y,z)=|R(x,y,z)|^2$ obtained by calculating the propagated field $R(x,y,z)$ using Eq. (2) in Fresnel approximation conditions. The set of values in the calculation are the same used in the previous simulation and summarized in Table 1. Then, $Q(x,y,0)$ is phase-encoded by multiplying this distribution by random phase mask $R(x,y,0)$, i.e. $C(x,y,0) = R(x,y,0)Q(x,y,0)$. Figure 2(b) shows the irradiance of the propagated joint distribution $I_C(x,y,z)=|C(x,y,z)|^2$. Unlike the result presented in Fig. 1(b), $I_c(x,y,z)$ looks noisy and no information from the original image $Q(x,y,0)$ can be inferred.
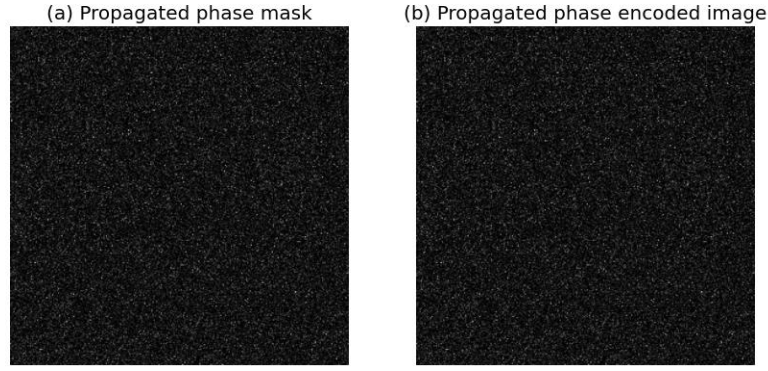
Fig 2: (a) Irradiance of the propagated field $I_R(x,y,z)=|R(x,y,z)|^2$. (b) Irradiance of the propagated field $I_C(x,y,z)=|C(x,y,z)|^2$

Moreover, the appearance of Figs. 2(a) and 2(b) is very similar, that is noise like distributions. Figures 3(a) and 3(b) show the histograms of $I_R(x,y,z)$ and $I_C(x,y,z)$ respectively (blue curves). Note that both histograms display an identical exponential decay behavior. At first sight, propagated phase-encoded images can be described as indistinguishable noisy distributions. Statistical optics [58, 59] describes the intensity distribution of speckle patterns using the following formula:

$$p(I) = \left(\frac{n_0}{\langle I \rangle}\right)^{n_0} \frac{I^{n_0-1}\exp\left(-In_0/\langle I \rangle\right)}{\Gamma(n_0)}, \tag{3}$$

where $I, \langle I \rangle$ and σ are the intensity data points, its average and the corresponding standard deviation respectively; and $\Gamma()$ stands for the Gamma special function. Parameter n_0 is defined as $n_0 = \left(\langle I \rangle/\sigma\right)^2$. Equation (3) corresponds to the **Gamma probability density function** [60]. This curve is superimposed on the histograms of Figs. 3(a) and 3(b). Calculated histograms and curves generated using Eq. (3) are not distinguishable. In general, random phase encoded signals produces speckle-like noise distribution after propagation within the Fresnel domain.
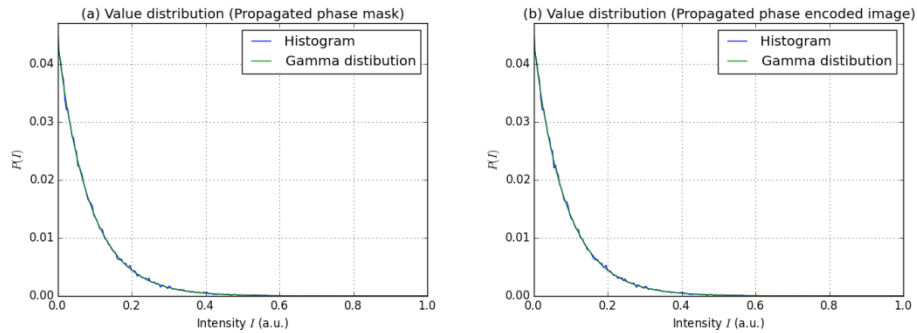

Fig. 3: Histograms of (a) the irradiance of the propagated phase mask $I_R(x,y,z)$ and (b) irradiance of the propagated phase-encoded image $I_C(x,y,z)$.

Random phase encoded images produce a speckle-like noise distribution after propagation. The intensity distribution of a speckle pattern is described using the following formula:

$$p(I) = \left(\frac{n_0}{\langle I \rangle}\right)^{n_0} \frac{I^{n_0-1}\exp\left(-In_0/\langle I \rangle\right)}{\Gamma(n_0)}$$

Despite the fact that $I_C(x,y,z)$ and $I_Q(x,y,z)$ can be easily distinguished by visual inspection, due to the statistical nature of the propagated random phase-encoded beam, the log-likelihood test can be used to discriminate among random phase-encoded and non-phase-encoded propagated distributions using a straightforward calculation [43]. Note that the log-likelihood statistical test is useful for distinguishing among two distributions of probability. Assuming that $I_C(x,y,z)$ and $I_Q(x,y,z)$ can be considered as classes that represent the null (H_0) and alternative (H_1) hypothesis respectively, the log-likelihood condition that validates the null hypothesis (H_0) is

$$\sum_{x,y}\left[\log\left(I_C\left(x,y,z\right)\right) - \log\left(I_Q\left(x,y,z\right)\right)\right] > 0. \tag{4}$$

Calculating the natural logarithm of Eq. (4) for the N pixels of the image,

$$N\ln\left[\frac{1}{\Gamma(n_{0C})}\left(\frac{n_{0C}}{\langle I_C \rangle}\right)^{n_{0e}}\right] + (n_{0C}-1)\sum_{x,y}\ln I_C\left(x,y,z\right) - \frac{n_{0C}}{\langle I_C \rangle}\sum_{x,y}I_C\left(x,y,z\right) -$$

$$N\ln\left[\frac{1}{\Gamma(n_{0Q})}\left(\frac{n_{0Q}}{\langle I_Q \rangle}\right)^{n_0}\right] + (n_{0Q}-1)\sum_{x,y}\ln I_Q\left(x,y,z\right) - \frac{n_{0Q}}{\langle I_Q \rangle}\sum_{x,y}I_Q\left(x,y,z\right) > 0. \tag{5}$$

Using the data obtained from Figs. 1(b) and 2(b) the following values are obtained $\sum_{x,y}\left[\log\left(I_C\left(x,y,z\right)\right)\right]=126629$ and $\sum_{x,y}\log\left(I_Q\left(x,y,z\right)\right)=54165$. This means that the null hypothesis is accomplished. As expected, $I_C(x,y,z)$ has a better fit with the Gamma-distribution model [Eq. (3)] than $I_Q(x,y,z)$.

As explained in the introduction, the **Double Random Phase Encryption** [2] technique provides basic theoretical framework for optical encryption. It can be understood as a generalization of random phase encoding [1]. Figure 4 shows a sketch of a 4f optical encryption system: plain-text image Q is random phase encoded by means of phase-only mask M_1 and phase-only key M_2 is placed at the Fourier plane of the first lens. The scalar electric field at the Fourier plane of the second lens (cypher-text) is:

$$E\left(x,y,4\text{f}\right) = \text{FT}_{\lambda f}\left[M_2\text{FT}_{\lambda f}\left[QM_1\right]\right], \tag{6}$$

where sub-indices λf indicates that Fourier transforms have to be scaled accordingly. $E(x,y,4f)$ is a complex function and has to be recorded using holographical or interferometric techniques. Note that the same optical system can be used to decrypt the message: plain-text Q can be accessed provided that phase key M_2^* is known:

$$Q = \left|\text{FT}_{\lambda f}\left[M_2^*\text{FT}_{\lambda f}\left[E\right]\right]\right|. \tag{7}$$

where * is the complex conjugate operator. Note that Eq. (7) is valid assuming that Q is a real valued, positive distribution image.
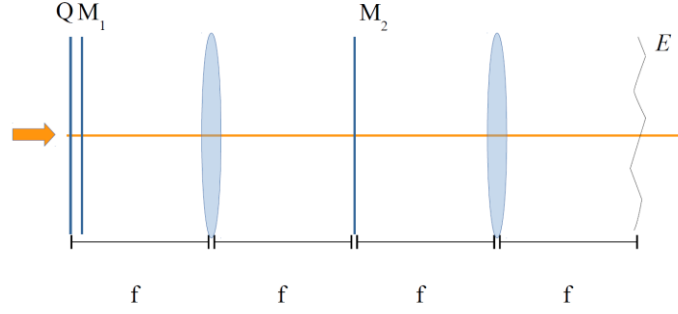
Fig. 4: Sketch of 4f Double Random Phase Encryption system

3. Polarized light

A quasi-monochromatic transverse electromagnetic field propagating in the z-axis direction is described as $\mathbf{E} = \left(E_x(\mathbf{r},t), E_y(\mathbf{r},t), 0 \right)$. The **Stokes vector** $\mathbf{S} = (S_0, S_1, S_2, S_3)$ is a convenient way of describing the state of polarization of the wave-front [61]. The components of this vector (the Stokes parameters) are defined as

$$S_0 = \left\langle E_x^* E_x \right\rangle + \left\langle E_y^* E_y \right\rangle$$
$$S_1 = \left\langle E_x^* E_x \right\rangle - \left\langle E_y^* E_y \right\rangle$$
$$S_2 = \left\langle E_x^* E_y \right\rangle + \left\langle E_y^* E_x \right\rangle \tag{8}$$
$$S_3 = i \left[\left\langle E_y^* E_x \right\rangle - \left\langle E_x^* E_y \right\rangle \right],$$

where $\langle \ \rangle$ stands for temporal average over time interval T, and

$$\left\langle E_i^* E_j \right\rangle = \frac{1}{T} \int_T E_i^* E_j dt. \tag{9}$$

Note that S_0 is a measure of the irradiance of the beam; S_1 compares the irradiance of the wave in the x- and y- directions, S_2 is similar to S_1 but the comparison is performed along two perpendicular directions rotated ±45° with respect to the x-axis. S_3 provides information of the circular content of the wave. For instance, for a linearly polarized beam with a polarization angle ψ with respect of the x-axis, the Stokes vector takes the form $\mathbf{S} = S_0 (1, \cos 2\psi, \sin 2\psi, 0);$. Elliptically polarized light is described by Stokes vector:

$$\mathbf{S} = S_0 \left(1, \cos 2\psi \cos 2\varepsilon, \sin 2\psi \cos 2\varepsilon, \sin 2\varepsilon \right) \tag{10}$$

where ε is the eccentricity angle. The four Stokes parameters are sometimes combined in a single one, called the **Degree of Polarization** (DoP):

$$P = \frac{1}{S_0} \sqrt{S_1^2 + S_2^2 + S_3^2}, \tag{11}$$

where P≤1. Note that for fully polarized beams $S_0^2 = S_1^2 + S_2^2 + S_3^2$ and P=1 whereas for totally depolarized light P=0. Sometimes, the Stokes parameters are normalized to S_0: $\hat{\mathbf{s}} = (S_1/S_0, S_2/S_0, S_3/S_0)$ and only three components are used.

The **Poincaré Sphere**, defined as the surface $|\hat{\mathbf{s}}| = 1$, provides a geometrical interpretation of the states of polarization (see Fig. 5). For instance, linear polarized light beams are located on the equator of the sphere. Circular polarized lights are placed on the poles (clockwise: North Pole, counterclockwise: South Pole). Totally polarized light (P=1) appears on the surface of the sphere whereas unpolarized light (P=0) is set at the very center of the Poincare sphere; partially polarized light states (P<1) are represented inside the sphere. In summary, an arbitrary state of polarization at the Poincaré sphere is described by

$$\mathbf{S} = S_0\left(1, P\cos 2\psi \cos 2\varepsilon, P\sin 2\psi \cos 2\varepsilon, P\sin 2\varepsilon\right). \tag{12}$$

When light interacts with an optical element, the state of polarization of the beam can change. The Stokes parameters of the wave-front are transformed according to a linear law represented by a 4x4 **Mueller matrix**, i.e. **S'=M·S**. If the beam passes through different polarization devices, the final Stokes vectors is obtained by combining the Mueller matrices of each element, **S'=M$_k$·M$_{k-1}$·…·M$_1$·S**. Note that Mueller calculus provides a complete representation of the state of polarization of light because it can be used with incoherent light. The Jones formalism provides a simpler description of polarization, but it can only be used with coherent beams.

> The Stokes vector is a convenient way of describing the state of polarization of the wave-front. Samples and optical devices that interact with light are described by the components of the Mueller matrix. When a light beam interacts with matter, the resulting state of polarization is obtained by multiplying the Mueller matrix by the original Stokes vector. If the beam passes through different optical devices, the overall Mueller Matrix is obtained by combining the Mueller matrices of each element.
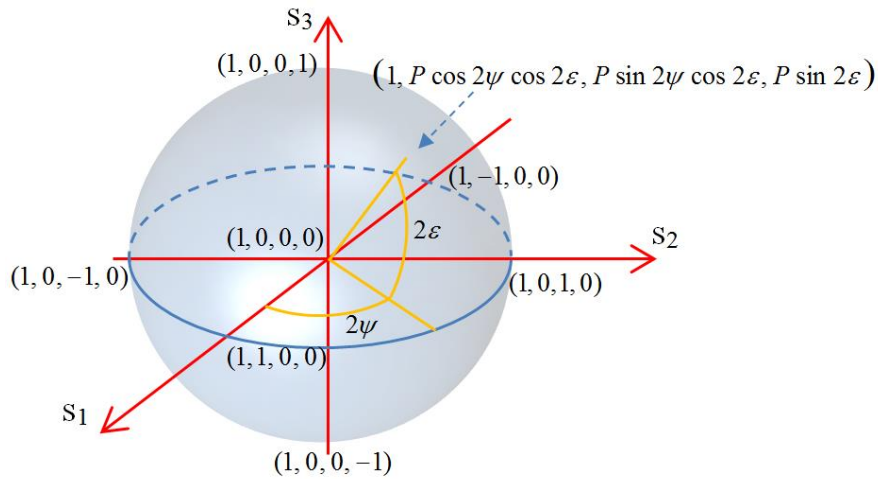

Fig 5. Poincaré Sphere displaying some states of polarization.

Equation (13) below shows the Mueller matrices for a linear polarizer with polarization axis in the 0º, 45º, 90º and -45º directions. The matrix for a quarter wave plate with fast axis in the horizontal direction is also indicated:

$$\mathbf{M}_{0°} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{M}_{\pm45°} = \frac{1}{2}\begin{pmatrix} 1 & 0 & \pm1 & 0 \\ 0 & 0 & 0 & 0 \\ \pm1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{M}_{90°} = \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{M}_{QWP} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

(13)

Note that Mueller matrix for a different direction of polarization θ can be obtained using rotation matrix $\mathbf{R}(\theta)$

$$\mathbf{R}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and}$$

(14)

$$\mathbf{M}_\theta = \mathbf{R}(-\theta)\mathbf{M}_{0°}\mathbf{R}(\theta).$$

(15)

The Stokes parameters can be easily measured with the help of a polarizer and a quarter wave plate in front of the light sensor using the following set of equations:

$$S_0 = I^{0°,0} + I^{90°,0}$$
$$S_1 = I^{0°,0} - I^{90°,0}$$
$$S_2 = I^{45°,0} - I^{-45°,0}$$
$$S_3 = I^{45°,\pi/2} - I^{-45°,\pi/2}.$$

(16)

$I^{\alpha,0}$ stands for the recorded irradiance when the linear polarizer is set at an angle α with respect to the x direction and $I^{\alpha,\pi/2}$ is the image recorded when a quarter wave-plate is used in addition to the polarizer. After recording these six quantities, the Stokes parameters are easily obtained. Mueller calculus provides a simple way of explaining why Eq. (16) provides a way of measuring the Stokes parameters. The corresponding Stokes vectors of the state of polarization of the light recorded in each measurement, namely S_0, S_1, S_2 and S_3, are determined using the Mueller formalism. Let $\mathbf{S} = (S_0, S_1, S_2, S_3)$ be, then:

$$\mathbf{S}_0 = \mathbf{M}_{0°} \cdot \mathbf{S} + \mathbf{M}_{90°} \cdot \mathbf{S} = (S_0, S_1, 0, 0)$$
$$\mathbf{S}_1 = \mathbf{M}_{0°} \cdot \mathbf{S} - \mathbf{M}_{90°} \cdot \mathbf{S} = (S_1, S_0, 0, 0)$$
$$\mathbf{S}_2 = \mathbf{M}_{45°} \cdot \mathbf{S} - \mathbf{M}_{-45°} \cdot \mathbf{S} = (S_2, 0, S_0, 0)$$
$$\mathbf{S}_3 = \mathbf{M}_{45°} \cdot \mathbf{M}_{QWP} \cdot \mathbf{S} - \mathbf{M}_{-45°} \cdot \mathbf{M}_{QWP} \cdot \mathbf{S} = (S_3, 0, S_0, 0).$$

(17)

Note that the first component of the Stokes vector provides the information of the measured irradiance. For instance, vector $\mathbf{S}_2$ is obtained by subtracting the states of polarization when the beam crosses polarizers $\mathbf{M}_{45°}$ and $\mathbf{M}_{-45°}$. In this case, the measured irradiance is S_2 that is the third Stokes parameter of the analyzed beam.

Whereas light is characterized by the four parameters of the Stokes vector, samples and optical devices that interact with light are described by the components of the Mueller matrix. In the general case, the Mueller matrix displays 16 different non-vanishing terms. In fact, due to matrix normalization only 15 components have to be measured. For **isotropic samples**, the Mueller matrix $\mathbf{M}_I$ takes a very simplified form because 8 out of the 16 components are zero [62]. In this case, $\mathbf{M}_I$ reads

$$\mathbf{M}_I = \begin{pmatrix} 1 & -N & 0 & 0 \\ -N & 1 & 0 & 0 \\ 0 & 0 & C & S \\ 0 & 0 & -S & C \end{pmatrix},$$ (18)

where quantities N, C and S are related with ellipsometric angles Ψ and Δ by means of

$$N = \cos(2\Psi),$$
$$C = \sin(2\Psi)\cos\Delta,$$
$$S = \sin(2\Psi)\sin\Delta.$$ (19)

Ψ and Δ are related to the p- and s- reflection Fresnel coefficients (r_{pp} and r_{ss}) by means of ratio ρ:

$$\rho = \frac{r_{pp}}{r_{ss}} = \tan\Psi \exp(i\Delta).$$ (20)

For non-depolarizing anisotropic samples, the **differential Mueller matrix L=ln(M)** provides a simpler description of the media considered [63]. Instead of the 15 independent components of the conventional Mueller matrix, **L** is fully described by means of 6 independent terms:

$$\mathbf{L} = \begin{pmatrix} 0 & l_{01} & l_{02} & l_{03} \\ l_{01} & 0 & l_{12} & l_{13} \\ l_{02} & -l_{12} & 0 & -l_{23} \\ l_{03} & -l_{13} & l_{23} & 0 \end{pmatrix}.$$ (21)

In this case, three generalized Fresnel coefficients have to be defined, namely: $\rho = \frac{r_{pp}}{r_{ss}}, \rho_{ps} = \frac{r_{ps}}{r_{ss}}$ and $\rho_{sp} = \frac{r_{sp}}{r_{ss}}$ where r_{ps} and r_{sp} stand for the reflected light that is transformed from the p- to the s- polarization and vice versa. These ratios can be calculated using the components of the differential Mueller matrix [62]:

$$\rho = \frac{1}{KT}\left[\cos\frac{T}{2} + (l_{01} - il_{23})\sin\frac{T}{2}\right]$$

$$\rho_{ps} = \frac{1}{KT}\left[l_{12} + l_{02} - i(l_{03} + l_{13})\sin\frac{T}{2}\right]$$ (22)

$$\rho_{sp} = \frac{1}{KT}\left[l_{12} - l_{02} - i(l_{03} - l_{13})\sin\frac{T}{2}\right],$$

where

$$K = \frac{1}{T}\left[\cos\frac{T}{2} - (l_{01} - il_{23})\sin\frac{T}{2}\right] \text{ and}$$

$$T = \sqrt{(l_{23} + il_{01})^2 + (l_{13} + il_{02})^2 + (l_{12} - il_{03})^2}.$$

3.1. Measurement of the Mueller matrix

The measurement of the 16 components of the Mueller matrix of a sample is not a simple task. As a general approach, **M** can be determined by subsequently illuminating the sample with four different beams with independent states of polarization. It assumed that the best results are obtained when the four Stokes parameters of the illuminating beam

are on the vertices of a regular tetrahedron inscribed in the Poincaré Sphere [64, 65]. Accordingly, a set of 16 independent linear equations is produced and thus the problem can be solved. To minimize error, more than four different states of polarization measurements can be made. In this case, the components of the Mueller matrix can be determined by calculating the pseudoinverse matrix.

Some different methods for measurement of the Mueller Matrix have been proposed [66]. Among the available techniques, in this paper we used two different approaches adapted to the characteristics of the problem. The first system considered is based on four **photo-elastic modulators** (PEMs) as variable phase retarders. This device presents two main advantages: measurements are made simultaneously and there are no moving parts [67]. A sketch of this setup is presented in Fig. 6(a). PEMs are described as phase plates with horizontal fast axis with retardation of δ. PEMs are electrically driven, with retardation being a periodic function of time, i.e. $\delta(t) \propto \cos(wt + \phi)$:

$$\mathbf{M}_{PEM} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\delta(t) & \sin\delta(t) \\ 0 & 0 & -\sin\delta(t) & \cos\delta(t) \end{pmatrix}. \tag{23}$$

The Mueller matrix of the sample is obtained by writing the polarization state of light recorded by the detector. The light reflected by the sample with matrix $\mathbf{M}_S$ is:

$$\mathbf{S}_S = \mathbf{M}_S \mathbf{R}(-\theta_2) \mathbf{M}_{PEM1} \mathbf{R}(\theta_1) \mathbf{R}(-\theta_0) \mathbf{M}_{PEM0} \mathbf{R}(\theta_0) \mathbf{R}(-\theta_{P1}) \mathbf{M}_{P1} \mathbf{R}(\theta_{P1}) \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{24}$$

where $\mathbf{M}_{PEM0}$, $\mathbf{M}_{PEM1}$, $\mathbf{M}_{P1}$, θ_0, θ_1 and, θ_2 are the matrices and orientation angles of PEM0, PEM1, and P1 respectively. The angles are set in such a way that calculations are simpler. Equivalently, the polarization state at the detector is

$$\mathbf{S}_D = \mathbf{R}(-\theta_{P2}) \mathbf{M}_{P2} \mathbf{R}(\theta_{P2}) \mathbf{R}(-\theta_3) \mathbf{M}_{PEM1} \mathbf{R}(\theta_3) \mathbf{R}(-\theta_2) \mathbf{M}_{PEM1} \mathbf{R}(\theta_2) \mathbf{S}_S, \tag{25}$$

where the first component of $\mathbf{S}_D$ is the measured intensity of the detector. Since $\delta(t)$ is a harmonic function, the components of the sample matrix M can be obtained using Fourier analysis of the recorded signal. The interested reader can access reference [67] for a more comprehensive explanation.
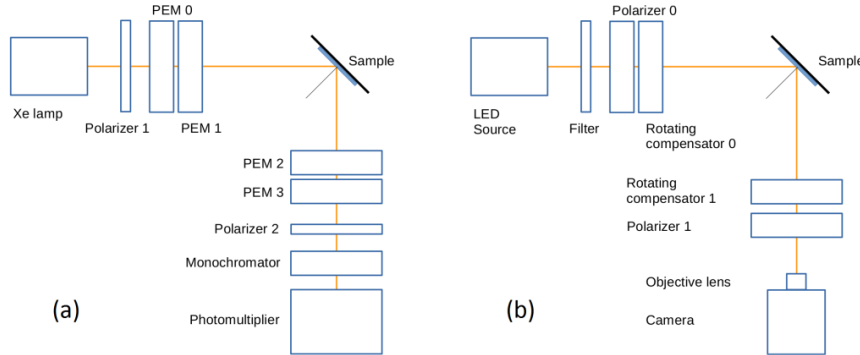

Fig 6: Sketch of (a) four photoelastic modulators Mueller matrix polarimeter and (b) Mueller Matrix imaging ellipsometer. PEMs are photo-elastic modulators. Adapted from [Opt. Lett. ,**41**, 4507 (2016)] [46].

A second instrument for measuring the components of the Mueller matrix has been used. The outline of the apparatus is presented in Fig. 6(b). Again, this device can be described in terms of four main devices: the polarization state generator uses a linear polarizer and a rotating compensator whereas the reflected light beam is analyzed by means of a rotating compensator and a polarizer. The sample is imaged on a camera and thus, the present system is able to perform image polarimetry.

According to Fig. 6(b), the Stokes vector $\mathbf{S}_D$ of the detected light is determined using the following equation

$$\mathbf{S}_D = \mathbf{M}_1\mathbf{M}_{RC1}\mathbf{M}_S\mathbf{M}_{RC0}\mathbf{M}_0\mathbf{S}_S, \tag{26}$$

where $\mathbf{M}_0$ and $\mathbf{M}_1$ are the Mueller matrix of linear polarizers [Eq. (15)] and $\mathbf{M}_{RC}$ is the corresponding matrix for a compensator

$$\mathbf{M}_{RC} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2 2\theta + \cos\delta\sin^2 2\theta & \cos 2\theta\sin 2\theta(1-\cos\delta) & \sin 2\theta\sin\delta \\ 0 & \cos 2\theta\sin 2\theta(1-\cos\delta) & \sin^2 2\theta + \cos\delta\cos^2 2\theta & -\cos 2\theta\sin\delta \\ 0 & -\sin 2\theta\sin\delta & \cos 2\theta\sin\delta & \cos\delta \end{pmatrix}. \tag{27}$$

Note that theta is a function of time $\theta=wt$ and δ is the retardation of the compensator. The angular speeds of the two compensators w_0 and w_1 were related by the relationship $w_0 = \frac{p}{q}w_1$ where p and q are integer numbers. The orientation of polarizers $\mathbf{M}_0$ and $\mathbf{M}_1$ was set at 0º and 90º but other options can be considered. As in the previous case, the components of the Mueller Matrix of the sample are determined using Fourier analysis methods [68].

4. Variables, datasets and reduction of dimensionality.

Classification problems require an accurate selection of the variables used to train and test the system. In some of the practical examples described in this paper, we use different kinds of data: in scenarios where polarization plays a key role, parameters such as the Mueller matrix components, and the Stokes parameters or the Degree of Polarization (DoP) provide additional information. Sometimes these variables are measured for a determined range of wavelengths, providing classification features with a large number of dimensions. In the case of measurements of speckle, recorded data can be adjusted to a Gamma distribution of probability. Here, the histogram of the distribution or statistics such as mean, variance, skewness, and kurtosis can be helpful. These parameters are defined as:

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(x_i - \langle x \rangle \right)^2$$

$$\hat{s}(x) = \frac{\sqrt{N(N-1)}}{N-2} \frac{\frac{1}{N} \sum_{i=1}^{N} \left(x_i - \langle x \rangle \right)^3}{\left(\sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(x_i - \langle x \rangle \right)^2} \right)^3}$$

(28)

$$\hat{k}(x) = 3 + \frac{N-1}{(N-2)(N-3)} \left((N+1) \frac{\frac{1}{N} \sum_{i=1}^{N} \left(x_i - \langle x \rangle \right)^4}{\left(\frac{1}{N} \sum_{i=1}^{N} \left(x_i - \langle x \rangle \right)^2 \right)^2} - 3(N-1) \right),$$

where x_i are the corresponding measurements and N is the total number of pixels. Note that the formulas for the skewness and kurtosis are bias-corrected and differ from the conventional ones. Also, the image entropy [69] is also taken into account:

$$H(x) = -\sum_{i=1}^{N} p_i \log_2 p_i.$$

(29)

Here, p_i is the relative frequency distribution of gray level i.

A **training dataset X** for N samples is described by means of a N x M matrix where row i represents a M-dimensional vector containing the features that describe sample i. Classes are labelled as integer numbers and arranged as a column vector $\mathbf{y} = \left(y_1, y_2, ..., y_n \right)^{\mathrm{T}}$ of dimension N, i.e.;

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & x_{13} & ... & x_{1m} \\ x_{21} & x_{22} & x_{23} & ... & x_{2m} \\ x_{31} & x_{32} & x_{33} & ... & x_{3m} \\ \vdots & \vdots & \vdots & ... & \vdots \\ x_{n1} & x_{n2} & x_{n3} & ... & x_{nm} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}.$$

(30)

4.1 Principal component analysis

Sometimes, feature vectors can have high dimensions. For instance, a conventional 8-bit image histogram has to be stored in a 256 component vector; a wavelength spectrum can require several hundreds of components depending on the analyzed range and the step size. This huge amount of information may complicate the use of classification techniques, especially when the number of measured samples is small when compared with the dimension of feature vectors. This effect is known as the curse of dimensionality [70]. **Principal Component Analysis** (PCA) is popular method for reducing dimensionality [71]. The idea beyond PCA is to find an alternative base that fits better the characteristics of the measured variables of the problem. For an excellent and intuitive tutorial on PCA the reader can access reference [72].

The PCA approach can be derived using just linear algebra. Let $\mathbf{Y}$ be a matrix related to dataset $\mathbf{X}$ [Eq. (30)] by means of linear transformation $\mathbf{P}$, i.e. $\mathbf{Y} = \mathbf{P} \mathbf{X}$. In what follows, the mean of the measurements is zero (i.e. the mean has been previously subtracted). Reduction of dimension is based on the idea that two uncorrelated variables can be described as a new variable which is a linear combination of the original ones. Note that the cross variance between two uncorrelated variables is equal to zero. We introduce covariance matrix for the dataset described in the transformed space $\mathbf{S_y}$ as

$$S_y = \frac{1}{M-1} YY^T. \tag{31}$$

Rewriting Eq. (31) in terms of **P** and **X**, S_y becomes,

$$S_y = \frac{1}{M-1} YY^T = \frac{1}{M-1}(PX)(PX)^T = \frac{1}{M-1}PXX^TP^T = \frac{1}{M-1}PAP^T, \tag{32}$$

where new matrix **A** is $A = XX^T$. Note that the target of PCA is to find the linear transformation that provides the best representation of dataset **X.** Best means that the variables have to be uncorrelated and thus S_y has to be diagonal. Since $A = XX^T$, this matrix can be diagonalizable, i.e $A = EDE^T; D$ is a diagonal matrix and **E** is the transformation (eigenvectors) matrix. Finally, if we set $E = P^T$ then $A = P^TDP$ and S_y becomes diagonal:

$$S_Y = \frac{1}{M-1} PAP^T = \frac{1}{M-1} PP^TDPP^T = \frac{1}{M-1} D. \tag{33}$$

Note that $P^T = P^{-1}$. In summary, the eigenvectors of $A = XX^T$ transforms S_y into a diagonal matrix. It is worth to point out that eigenvectors associated with high eigenvalues (variances) of S_Y provide more information that those directions whose eigenvalues are small. Accordingly, these less relevant directions can be avoided and the dimensionality of the dataset is reduced.

> Principal Component Analysis looks for an alternative base for reducing dimensionality of dataset **X**; this base corresponds to the eigenvectors of $A = XX^T$. Eigenvectors associated to high eigenvalues provide more information that those directions with small eigenvalues: at the end of the day, less relevant directions are avoided and dataset dimensionality is reduced.

5. Classification using machine learning techniques:

In this section we provide a summary of the machine learning classification methods [73] used in the present paper. Among different possible techniques we used K-Nearest Neighbors (K-NN), Support Vector Machines (SVM) and Random Forests (RFs). The three methods are non-parametric and supervised. Non-parametric means that the number of variables involved in the calculation is determined by the training data. Supervised learning means that the system learns from a training set of data in which each member of the training set is labeled with the class this element belongs to.

5.1 K-Nearest Neighbors

K-NN is a machine-learning classification algorithm relatively simple to implement that provides very accurate results with high classification success ratios [74]. To determine the class a test sample belongs, K-NN searches for the k closest points of the training dataset. The analyzed test point is assigned to the class the majority of the k-closest points belong. Sometimes, this voting procedure is weighted according to the distance, that is, the most distant points are less significant and vice-versa. Euclidean distance is the by default option but other definitions of distance or weights can be used as well. Note that the proper selection of the number of neighbors k is highly dependent on the problem considered. In general, small values of k define better boundaries but the system becomes more sensitive to noise

Hold-out validation of classification is performed by splitting the dataset into training and test subsets. To avoid bias, validation is performed several times. In each trial, the samples used in the training or test sets are randomly selected. Classification accuracy is determined by calculating the ratio of the total number of correctly classified samples, divided by the total number of test samples [75]. Several websites provide materials to test and experiment classification methods using real data. For instance, the interested

reader can access the scikit-learn website for useful insight on K-NN classification [76, 77].

5.2 Multiclass Support Vector Machine

SVM is a classification method based on determining the best hyper-surface able to distinguish among two classes [78, 79]. Later on, SVM was generalized in such a way that can be utilized with multiple classes as well [80]. The use of SVM is suggested when the number of samples in the datasets is small but they present high dimensionality.

Let $\mathbf{x}_i$ and y_i be the m–dimensional feature vector of sample i and the label that describes the class $\mathbf{x}_i$ belongs, respectively [$\mathbf{x}_i$ are the rows of Eq. (30)]. Possible values for y_i are +1 and -1. Let $\mathbf{w} \cdot \mathbf{x} + b = 0$ be the equation of the optimum hyperplane (namely, **maximum margin hyperplane**) that separates both classes; $\mathbf{w}$ is the normal vector to the hyperplane, $\mathbf{x}$ is a hyper-point in the feature space and b is the bias. *Optimum hyperplane* means here that the distance between the hyperplane and the nearest point of either group is maximized. These points are the so-called **support vectors.** The hyperplanes that contains the support vectors (**margin hyperplanes**) are $\mathbf{w} \cdot \mathbf{x} + b = \pm 1$ and the distance between them is $2/|\mathbf{w}|$. Note that classes that are not separable in the feature (original) space, can be separated in a transformed higher dimensional space. A 2D sketch of the elements involved in a 2-class SVM is presented in Fig. 7.
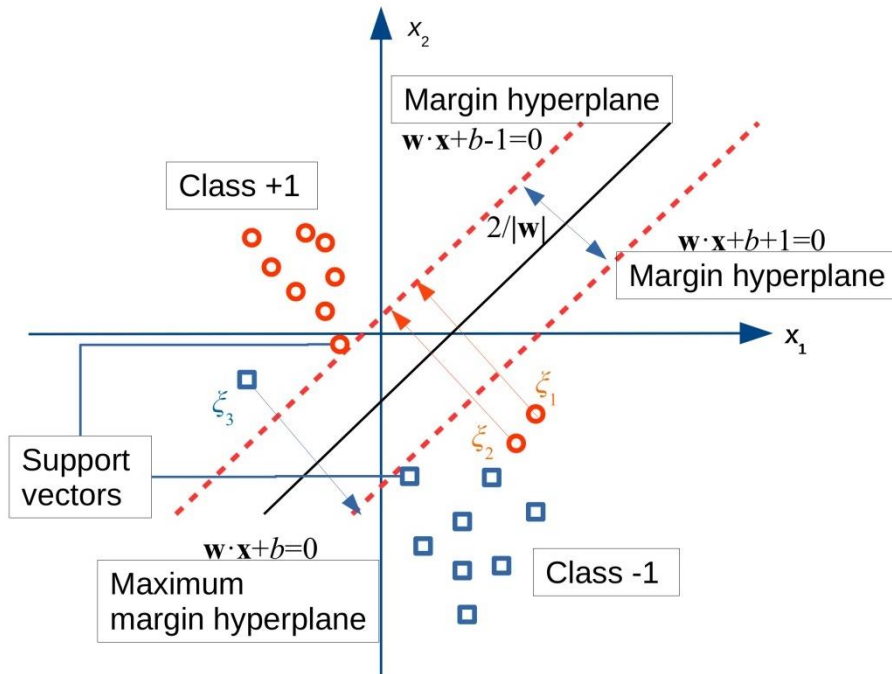

Fig. 7: Two-class SVM: margin hyperplanes and support vectors

Linear SVM is enunciated as follows: All the samples of class +1 have to lay in the semi-space defined by $\mathbf{w} \cdot \mathbf{x} + b \geq 1$. For the class -1 samples, the condition is $\mathbf{w} \cdot \mathbf{x} + b \leq -1$. Accordingly, all the samples of the training set have to satisfy the following conditions

$$y_i \left(\mathbf{w} \cdot \mathbf{x}_i + b \right) \geq 1 \text{ for } i = 1, ..., n. \tag{34}$$

The objective of the algorithm is to find the hyperplane whose distance between margin hyperplanes $2/|\mathbf{w}|$ is maximized. In other words, the system looks for the hyperplane that satisfies:

$$\text{minimize} |\mathbf{w}|^2 \text{ with conditions } y_i \left(\mathbf{w} \cdot \mathbf{x}_i + b \right) \geq 1 \text{ for } i = 1, ..., n. \tag{35}$$

This condition can be refined taking into account that some samples cannot be properly be classified. We introduce parameter ξ as the distance between a misclassified sample and the correct margin hyperplane; $\xi=0$ if the sample is correctly classified. Then, condition (35) is rewritten as

$$\min \left[|\mathbf{w}|^2 + C \sum_i \xi_i \right] \tag{36}$$

with conditions $y_i \left(\mathbf{w} \cdot \mathbf{x}_i + b \right) \geq 1 - \xi_i$ and $\xi_i \geq 0$ for $i = 1, ..., n,$

where C is a regularization constant that has to be determined in such a way SVM reports the best possible results.

In general, classes cannot be separated by means of a hyperplane and a nonlinear approach is required. Feature space is transformed to a new space by means of nonlinear transformation law ϕ. This new space can be of higher dimension but the decision boundary surface in the transformed space is again a hyperplane. Condition (36) becomes

$$\min \left[|\mathbf{w}|^2 + C \sum_i \xi_i \right] \tag{37}$$

with conditions $y_i \left(\mathbf{w} \cdot \phi(\mathbf{x}_i) + b \right) \geq 1 - \xi_i$ and $\xi_i \geq 0$ for $i = 1, ..., n.$

Kernel function K describes a simple way of calculating dot products. Several kernels have been described; in this paper we use the Gaussian Radial Basis Function kernel, defined as

$$K \left(\mathbf{x}_i, \mathbf{x}_j \right) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) = C \exp \left(-\gamma |\mathbf{x}_i - \mathbf{x}_j|^2 \right) \text{ with } \gamma > 0. \tag{38}$$

Again, C and γ have to be found in order for SVM to produce the best classifications results.

SVM has been generalized to K classes. In order to attack this problem, two possible approaches are possible:

(i) one-versus-one (two classes are taken into account each time): 2-class SVM is run $K(K-1)/2$ times: class A vs B, A vs C, B vs C, et cetera. In each run the sample is assigned to a determined class and the class the sample belongs is the one that has been assigned a majority of times. This is the approach used in this paper.

(ii) one-versus-the-rest: one class is tested against the rest of the samples.

5.3 Random forests (RFs)

RFs [81] are based on decision trees. Decision trees are commonly used as selection procedures since they provide a visual perspective of the analyzed problem. The main approach behind RFs is that a group of weak learners or classifiers such as trees can be integrated to produce a strong learner or classifier. Averaging trees is also known as Bagging. Based on this tool, random forests have been demonstrated as a very powerful classification technique. According to some authors, random forests provide the best

possible performance among classification methods [82] but other groups find this statement arguable [83].

As indicated in Eq. (30), $\mathbf{X}$ is the dataset matrix; each of the n samples of the dataset is described by m features whereas $\mathbf{y}$ provides the information of the class each sample belongs to. $\mathbf{X}$ and $\mathbf{y}$ are combined in a new matrix $\widehat{\mathbf{X}}$ (see Fig. 8).

$$\widehat{\mathbf{X}} = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1m} & y_1 \\ x_{21} & x_{22} & x_{23} & \dots & x_{2m} & y_2 \\ x_{31} & x_{32} & x_{33} & \dots & x_{3m} & y_3 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nm} & y_n \end{pmatrix}$$
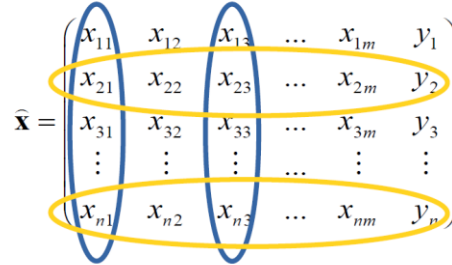
Fig. 8: Combined dataset $\widehat{\mathbf{X}}$. Ellipses indicate a random selection of data (yellow ellipses) and features (blue ellipses) for generating a random tree.

With the information contained in the dataset a complete decision tree is generated (Fig. 9). Each node is split into leaf nodes using the conditions until all features are used.
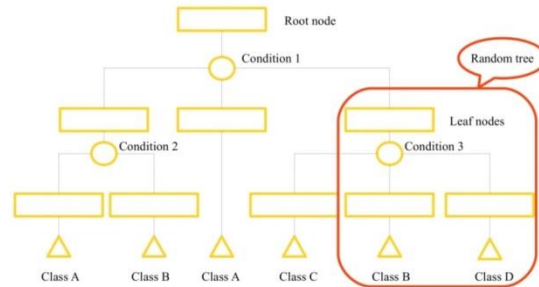

Fig. 9: Example of a complete decision tree. The red box indicates a random selection of nodes and conditions, i.e.: a random tree

RFs are based on selecting at random a limited numbers of features and samples of the dataset (bagging process). Note that the minimum number of features v required to create a random tree is $v = \lfloor \sqrt{m} \rfloor$, where the symbol $\lfloor \ \rfloor$ stands for the floor operator. A random decision tree is created using partial data (for instance, yellow and blue ellipses in Fig. 8). The system is trained by generating a large collection of uncorrelated random trees (such as the red box in Fig. 9). RFs classification is based on the assumption that most of the partial random trees will provide a correct prediction of the class. When a test sample has to be classified, its features are used to check the generated random trees. The output of every random tree provides a vote for a specific class for the test sample. The class the sample is assigned is obtained by the majority of votes of the set of random trees.

Support Vector Machine is a classification method based on determining the best hyper-surface able to distinguish among two or more classes.
Random forests are based on selecting randomly a limited number of features and samples of the dataset (random tree). The system is trained by generating a large collection of uncorrelated random trees. Classification is based on the assumption that most of the partial random trees will provide a correct prediction of the class.

6. Authentication of codes produced with gold nanoparticles

In this section, we demonstrate by experiments that QR codes providing the same information but made of different materials can be distinguished by the combined use of nano particle phase encoding, polarized light, statistical analysis of speckle, and machine learning techniques [44]. The problem consists of authenticating a 21x21 pixel QR code fabricated with gold nanoparticles (AuNP) from another QR code fabricated with platinum nanoparticles. Both gold nanoparticles and platinum (Pt) nanoparticles QR codes are deposited on a glass surface and encoded with the same information, that is the sequence of numbers 12 155 22 13. The size of these codes is 0.840 x 0.840 mm^2 (40 microns/pixel). In order to increase the complexity of the problem, a third code made of AuNP but without QR code was also produced. The image of the three structures is shown in Fig. 10(a-c).
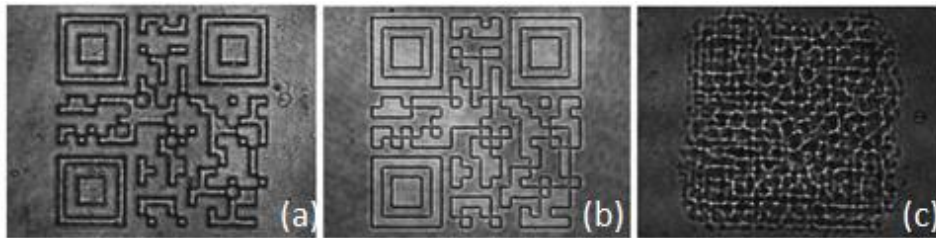

Fig. 10: (a) AuNP (gold) QR code (sample A), (b) Platinum QR code (sample B), and (c): AuNP (gold) structure without QR code (sample C). Reprinted from [Opt. Lett., **40**, 2135 (2015)] [44].

Figure 11 sketches the optical setup. A green laser diode (λ=532 nm) passes through a variable density filter for tuning the total intensity. A linear polarizer and a quarter wave plate are used to generate circularly polarized light. A second polarizer is used to select the direction of polarization. Then, the sample is illuminated and propagated 15 mm where a CCD camera records the light distribution. Note that the light that illuminates the code has to be polarized. Metallic nanoparticles used to produce the code show dichroism and thus, absorption is dependent on the direction of polarization (see, for instance, [51-53]).
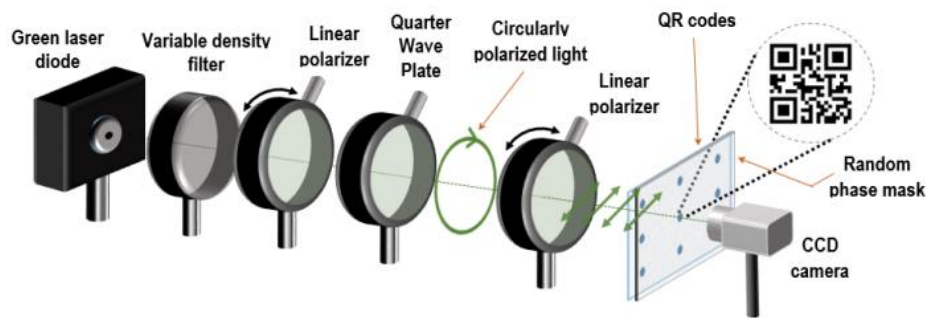

Fig 11: Optical setup for authentication of codes produced with gold nanoparticles. Reprinted from [Opt. Lett., **40**, 2135 (2015)] [44].

The corresponding light distributions after propagation are shown in Fig. 12(a-c) (see the caption for details). Note that Fig. 12(a) and Fig. 12(b) look different despite the initial QRs encode the same information.
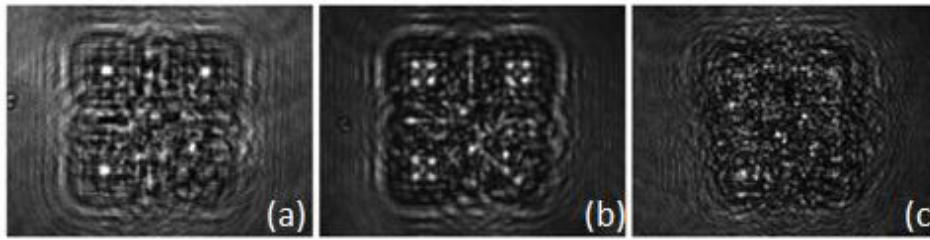
Fig. 12: Propagated QR codes: (a) sample A, (b) sample B, and (c) sample C (see Fig. 10 for details). Reprinted from [Opt. Lett., **40**, 2135 (2015)] [44].

The state of polarization of the light that illuminates the sample is modified by rotating the second polarizer from 0° to 179° at intervals of 1° [Fig. 11]. Accordingly, 180 8-bit images have been recorded and the corresponding histograms are calculated. Note that the variable density filter is set in a way that the camera never saturates. Figures 13(a), 13(b) and 13(c) display the histograms of the propagated distributions of samples A, B and C respectively when the polarization angle is set at 0°, 30°, 60°, 90°, 120° and 150°. For a given sample, all the histograms behave in a very similar way. However, there are slight differences in the histogram when the polarization of the illuminating beam is changed. This fact can be explained in terms of the dichroism present in metal nanoparticles. In other words, optical parameters depend on the direction of polarization of the incident light [51-53].

As discussed in section 4.1, PCA provides a way for reducing the dimensionality of the problem. Accordingly, we used this technique to provide an alternative representation of the histograms. Figure 13(d) displays the results of applying this technique to the histograms. Results are shown using a 3D plot, which means that only the first three principal components are taken into account. It is apparent that using the PCA representation, the set of histograms for each sample look completely disconnected.
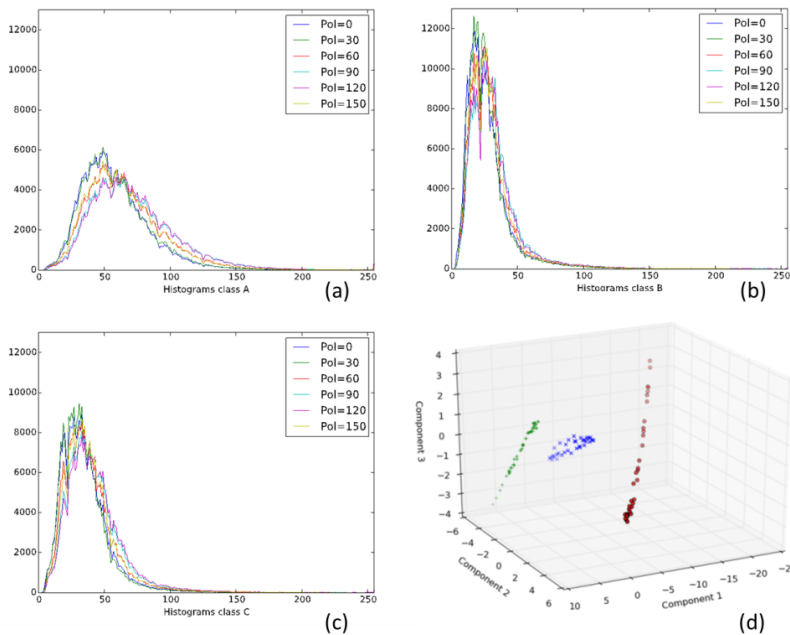

Fig. 13: Histograms of the three samples in Fig. 12(a-c) for different directions of polarization (0º, 30º, 60º, 120º and 150º). (d) Principal Components Analysis of the histograms. Red dots: sample A, green dots: sample B:, and blue dots: sample C. Reprinted from [Opt. Lett., **40**, 2135 (2015)] [44].

The experiment presented above was repeated but now the codes of Fig. 10 are phase encoded. The propagated irradiances are presented in Fig. 14(a-c). At first sight, the images present a speckle-like appearance compatible with the theoretical concepts developed in section 2.

Fig. 14: Propagated random phase-encoded QR codes: (a) sample A, (b) sample B, and (c) sample C. Reprinted from [Opt. Lett., **40**, 2135 (2015)] [44].

Figures 15(a-c) display the histograms measured for the phase-encoded samples. As in the previous experiment the measured histograms display small differences depending on the polarization direction of the illuminating beam. Moreover, the profiles of the histograms look very similar for the three cases considered. Taking into account that the histogram of the propagated irradiances can be described as a speckle-Gamma probability distribution, a Kolmogorov-Smirnov test was used to check the goodness of the fit [84]. All the histograms of the three samples passed the test.

Again, we have performed reduction of the histogram dimensions using PCA. Figure 15(d) shows a plot using the three first principal components. Surprisingly, the clouds of points for each sample are not overlapped.
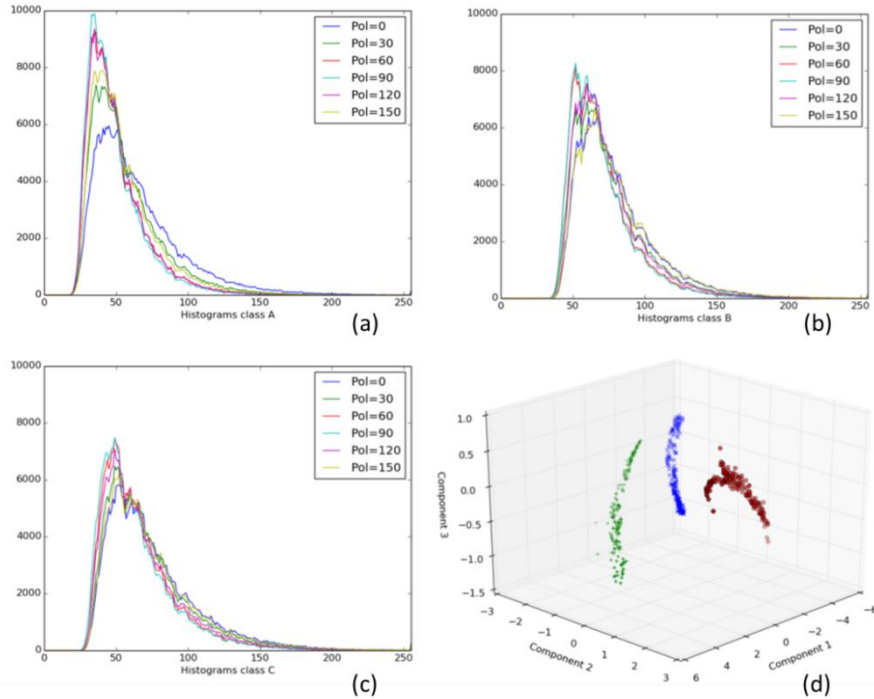

Fig. 15: (a-c) Histograms of the three phase-encoded samples in Fig. 14(a-c) for different directions of polarization (0º, 30º, 60º, 120º and 150º); (d) Principal Components Analysis for the phase encoded samples. Red dots: sample A, green dots: sample B:, and blue dots: sample C. Reprinted from [Opt. Lett., **40**, 2135 (2015)] [44].

Finally, classification among samples using 3-class SVM has been carried out. Since SVM is appropriate for high dimensional problems, we fed the SVM algorithm with the information provided by the histograms. Both cases are considered: phase-encoded and non-phase encoded samples. A hold-out strategy is used to train the classification model (see section 5.1). The dataset is split into training and test subsets and validation is performed 40 times. As explained above, the samples used in the training or test sets were randomly selected. Table 2 shows accuracy results: no errors were detected during classification.

Table 2. SVM Results, three class problem

Phase mask	Accuracy (test set, in %)	# test runs
No	100	40
Yes	100	40

In summary, we demonstrated that polarized light can be used to distinguish and validate phase-encoded QR codes produced with metallic nanoparticles. The histograms of the Speckle distributions obtained after propagation are used to produce a dataset for multi-class SVM. Calculations show that the classification accuracy is 100%.

7. Authentication of codes produced with thin-films techniques

In the previous section, we demonstrated that codes encoding the same information but produced using different nano encoded materials can be distinguished because of their different polarimetric signature. Slight differences in the image histogram are used to train a classification system that is able to perform successful authentication. Now we consider a similar problem: two QR codes providing the same information are made using standard thin-film technology. We demonstrate that they can be distinguished by analyzing the polarimetric properties of the reflected light [45]. In the present study, the measurements were carried out using the Mueller matrix image ellipsometer presented in section 3.1 [Fig. 6(b) [68]].

The QR used is shown in Fig. 16(a). It encodes the numerical sequence 0034934021143. The lithographic mask with the QR image was produced using a high contrast orthochromatic film [Fig. 16(b)]. The two codes were produced using Chromium (Cr) and Tantalum pentoxide (Ta_2O_5) respectively. These materials were deposited on a microscope slide (25 mm x 75 mm) using radio-frequency sputtering [Fig. 16(c)]. The thicknesses of the resulting codes were 120 nm and 20 nm for the Ta_2O_5 and Cr samples respectively. Note that the Ta_2O_5 and Cr codes look very different whereas the former is very transparent, the latter is quite absorbent [Fig. 16(c)]. More information about the fabrication procedure of these codes can be found in [45, 85, 86].
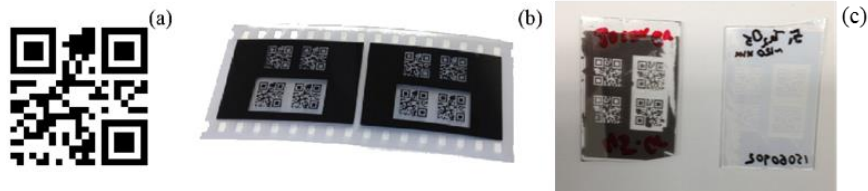

Fig. 16: QR code encoding the message '0034934021143'; (b) high-contrast thin film used as lithographic mask; (c) Ta_2O_5 QR (left) code and Cr QR (right). Adapted from [Opt. Lett., **40**, 5399 (2015)] [45].

The system is illuminated by means of a LED source in combination with an interferometer filter (λ=535 nm). The state of polarization of the beam is determined by means of a linear polarizer and a rotating achromatic compensator. The beam interacts

with the sample with an angle of incidence of 56º and passes through a rotating compensator and a linear polarizer. The resulting distribution of irradiance is imaged using an objective lens and recorded by a CMOS camera. This instrument produces the 16 Muller component images of the sample.

Figure 17(a-b) displays the 16 Mueller images for the two samples considered. It is quite apparent that 8 out of the 16 components of the matrix are zero and, as expected, the two samples are isotropic [Eq. (18)]. The Mueller matrix for isotropic samples is highly symmetrical with only three independent terms (N, C and S). These quantities are related to ellipsometric angles Ψ and Δ [Eq. (19)], and ratio ρ between the p- and s- reflection Fresnel coefficients [Eq. (20)]. This parameter can be used to distinguish among the two samples.
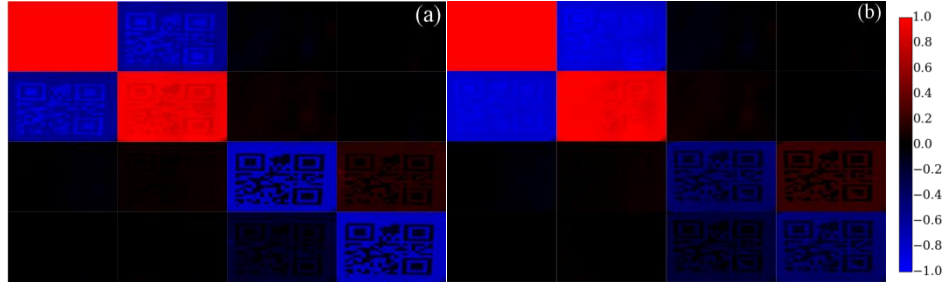

Fig. 17: Mueller matrix images for samples (a) Cr and (b) Ta$_2$O$_5$. Reprinted from [Opt. Lett., **40**, 5399 (2015)] [45].

We calculated coefficient ρ for the two samples namely ρ_{Cr} and $\rho_{Ta_2O_5}$ and phase-only filtering is used as a way to distinguish among the two samples [87]. This operation is defined as:

$$\rho_{Cr} \otimes \rho_{Ta_2O_5} = \left| FT^{-1} \left[\frac{FT[\rho_{Cr}]}{|FT[\rho_{Cr}]|} FT\left[\rho_{Ta_2O_5} \right]^* \right] \right|, \tag{39}$$

where $\otimes$ stands for the phase-only filter correlation operator. Figure 18(a-b) shows a 3D representation of correlations $\left| \rho_{Cr} \otimes \rho_{Cr} \right|$ and $\left| \rho_{Cr} \otimes \rho_{Ta_2O_5} \right|$. Note that the correlation peak for the latter is very small when compared with the auto-correlation. This means that the chromium sample is detected whereas the Ta$_2$O$_5$ is rejected. Therefore, polarimetric coefficient ρ can be used for distinguishing two identical images made of different materials using correlation.
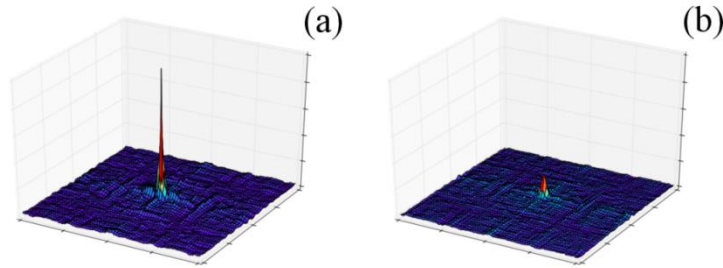

Fig. 18: Phase-only correlation: (a) $\left| \rho_{Cr} \otimes \rho_{Cr} \right|$ and (b) $\left| \rho_{Cr} \otimes \rho_{Ta_2O_5} \right|$. Reprinted from [Opt. Lett., **40**, 5399 (2015)] [45].

Anisotropic materials such as cellophane [88] can be used for producing more complex structures. Anisotropy has been induced by attaching small strips of adhesive tape on the two samples; note that the orientation, length and width of the strips are placed at random. The resulting Mueller matrices are presented in Fig. 19(a-b). Note that in both cases, these matrices do not show vanishing terms or symmetries.
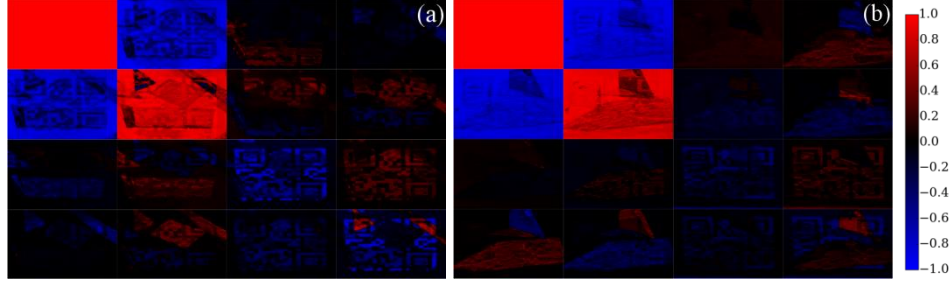

Fig. 19: Mueller matrix images for samples with adhesive tape: (a) Cr and (b) Ta$_2$O$_5$. Reprinted from [Opt. Lett., **40**, 5399 (2015)] [45].

In this second experiment, recognition based on phase-only correlation of Fresnel ratios ρ is not as simple as in the scenario presented above. With non-isotropic samples, two new complex-valued ratios, namely ρ_{sp} and ρ_{ps}, can be used to discriminate among samples. The three Fresnel ratios for anisotropic non-depolarizing media are calculated using Eqs. (21) and (22) (see section 2 and reference [62]).

Table 3 shows the value of the cross-correlation maxima for ρ, ρ_{sp} and ρ_{ps}. Values are normalized to the corresponding autocorrelation maximum. In anisotropic media, the three cross-correlations have to be very small in order reject that sample. In this example, the three cross-correlation values are very small and thus the Cr code is detected.

<div align="center">

Table 3. Phase-only cross-correlation maxima.
Samples with adhesive tape

</div>

$\dfrac{\left\| \rho_{Cr} \otimes \rho_{Ta_2O_5} \right\|(0,0)}{\left\| \rho_{Cr} \otimes \rho_{Cr} \right\|(0,0)}$	0.040
$\dfrac{\left\| \rho_{ps\text{-}Cr} \otimes \rho_{ps\text{-}Ta_2O_5} \right\|(0,0)}{\left\| \rho_{ps\text{-}Cr} \otimes \rho_{ps\text{-}Cr} \right\|(0,0)}$	0.020
$\dfrac{\left\| \rho_{sp\text{-}Cr} \otimes \rho_{sp\text{-}Ta_2O_5} \right\|(0,0)}{\left\| \rho_{sp\text{-}Cr} \otimes \rho_{sp\text{-}Cr} \right\|(0,0)}$	0.014

A third scenario has been considered. The objective lens of the ellipsometer was removed and thus, the samples cannot be imaged [see Fig. 6(b)]. Note that the QRs are phase-encoded and therefore, the recorded components of the Mueller matrix are speckle-like noise distributions. In this conditions, the code cannot be read but we demonstrate that the samples can be distinguished using only the polarimetric signature. This is very advantageous from the security point of view because samples can be authenticated as true or counterfeit without accessing the QR code or the information encoded.

Using the Mueller matrix, we can produce a large set of resulting Stokes parameters **S**'=**MS**. We take into account linearly polarized input light $\mathbf{S} = S_0 \left(1, \cos 2\psi, \sin 2\psi, 0\right)$ with ψ ranging from 0º to 179º at intervals of 0.1º. Note that this large set of Stokes vectors is numerically produced. The 256-bins histograms of the four Stokes parameters of the 1800 synthetically generated distributions were calculated. The results are presented in Fig. 20(a-d) (Cr sample) and Fig. 21(a-d) (Ta$_2$O$_5$ sample) (see caption for

details). At first sight these histograms look different and could be used for classification purposes using machine learning algorithms.
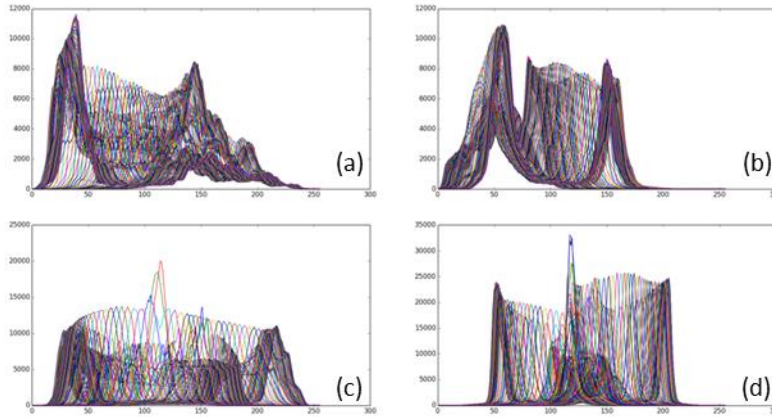

Fig. 20: Histograms (x-axis: grey levels and y-axis: frequency) of the 4 Stokes components distributions for class Cr: (a) S_0, (b) S_1, (c) S_2, and (d) S_3. Adapted from [Opt. Lett., **40**, 5399 (2015)] [45].

Dimensionality is reduced using Principal Component Analysis. Figure 22(a-d) shows the representation of the three largest principal components for S_0, S_1, S_2, and S_3. Note that with only three principal components the histograms corresponding to S_0 appear separated.

Components S_0, S_1, S_2, S_3 and the DoP were used to produce 5 independent datasets. Using a hold-out strategy, the datasets were divided at random into training and test sets. A two-class K-NN method with k=1, was used to perform classification within the five groups of data. The algorithm was run 50 times. The results are presented in Table 4: in summary, K-NN provides error-free classification using the histograms of S_0, S_2, S_3 or the DoP.

Table 4. K-NN classification mean accuracy, k=1

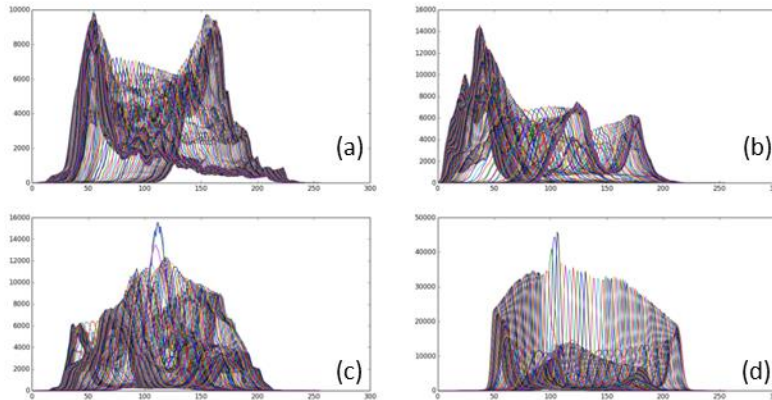| Parameter	S_0	S_1	S_2,	S_3	DoP	# test runs
Accuracy (test set, in %)	100	78	100	100	100	50

Fig. 21: Histograms (x-axis: grey levels and y-axis: frequency) of the 4 Stokes components distributions for class Ta_2O_5: (a) S_0, (b) S_1, (c) S_2, and (d) S_3. Adapted from [Opt. Lett., **40**, 5399 (2015)] [45].
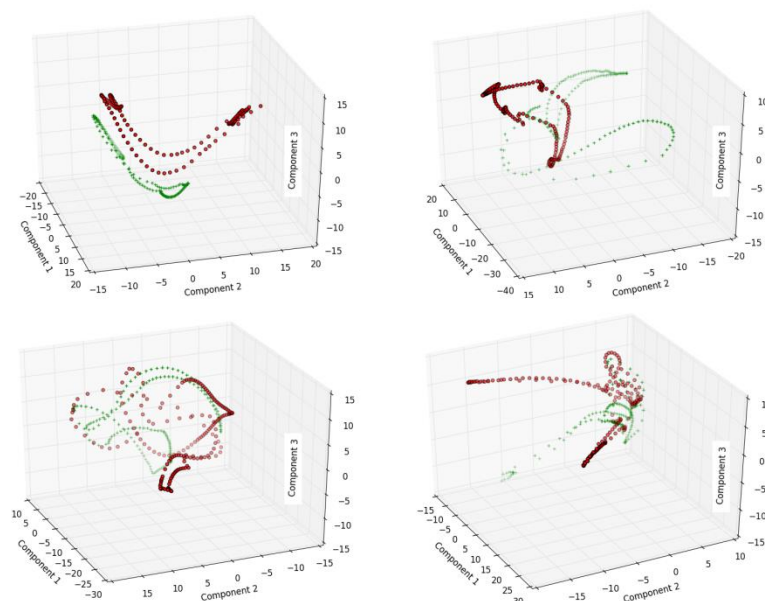
Fig. 22: Principal Components Analysis for the phase encoded samples Cr (red dots) and Ta$_2$O$_5$ (green dots): (a) S_0, (b) S_1, (c) S_2, and (d) S_3.

8. Authentication of pharmaceutical pills coated with gold nanoparticles.

Pharmaceutical products are produced under strict safety measures following well established protocols. Falsification of specifics could be a major problem for the industry, drugs agencies and general public. Criminal activity related with counterfeit medicines produces financial losses for companies and can become a health risk for patients who acquire medicines outside of the regulated commercial networks. Both public administrations and companies are pursuing methods for detecting counterfeiting products in a secure and fast way [89-93].

AuNP are very common in biomedicine: since AuNP are non-toxic and biocompatible, multiple applications have been described [94-98]. Among many other applications, AuNP are used as drug carriers and labeling, delivery, heating, or gene-regulating agents. Also, AuNP are used in imaging diagnostics [99, 100].

In this section we introduce an approach for tagging pharmaceutical tablets in such a way that falsification becomes virtually impossible. AuNP are used to produce nano particle coated tablets that become polarimetrically labeled. As explained in section 6, AuNP present dichroic behavior and thus, the properties of the scattered light may depend on the state of polarization of the light source. Tablets can be analyzed by means of polarized light and according to the polarimetric signature detected they are classified as true of counterfeit [46].

We produced flat placebo pharmaceutical tablets according to the specifications of Table 5. The size of the tablets is 10 mm. The coating is composed by deionized water, AuNP of different sizes and hydroxypropyl methylcellulose (HPMC), a very common excipient used in the pharmaceutical and food industries. Three commercial solutions containing AuNP with different sizes and concentrations (provided by Endor Nanotechnologies, Barcelona) were used. The nanoparticles of the solutions used were very homogenous with a Dispersity index Đ close to 1. A short description on how the tablets were prepared can be found in [46]. In summary, three sets of gold-coated tablets were

produced containing respectively 4, 12 and 25 nm AuNP. In addition to these sets, we prepared a fourth group of tablets containing no nanoparticles.

Table 5: Pharmaceutical tablets and coating components

Tablet components:	Lactose monohydrate (Tablettose® 80)	35%
	Microcrystalline cellulose (Vivapur® 101)	59%
	Talc	4%
	Magnesium stearate	2%
Tablet coating:	Hydroxypropyl methylcellulose (HPMC) (Pharmacoat® 606)	3%
	Deionized water	97%
Nanoparticles	4 nm AuNP. 0.05 mg/ml	
	12 nm AuNP, 0.03 mg/ml	
	25 nm AuNO, 0.01 mg/ml	

The tablets of the four sets look identical and cannot be distinguished by visual inspection (see Fig. 23). A couple of pills of each group were selected at random for characterization. The Mueller matrix components were obtained using a four-PEM Mueller Matrix Polarimeter; this instrument is depicted in Fig. 6(a) [67]. The light source was able to illuminate the sample in the range 280–700 nm. Results for the 8 measured tablets are presented in Fig. 24.
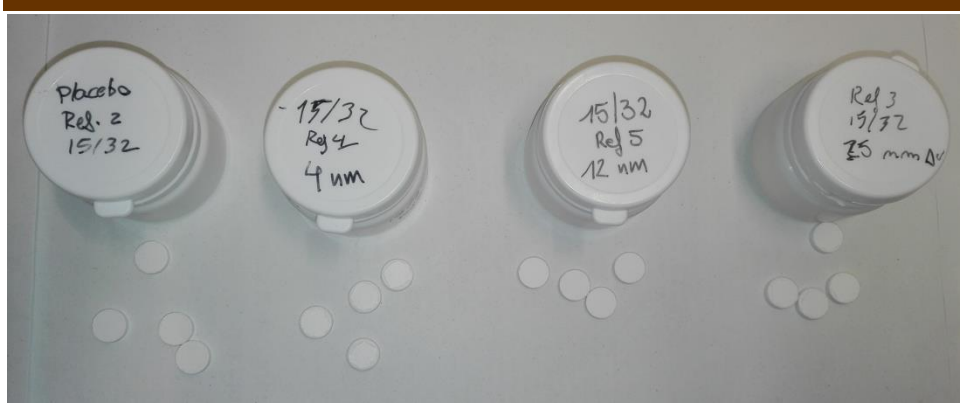

Fig. 23: Produced tablets used in the experiments.

Note that some of the Mueller components in Fig. 24 are very close to zero for any wavelength. Accordingly, the tablets display a behavior that can be identified as isotropic because the measured Mueller matrix can be approximately described by means of Eqs. (18) and (19) (see section 2).

Figure 25 shows the DoP as a function of the wavelength for some particular states of polarization of the incident light, that is natural, linear (horizontal, 45º, vertical and -45º) and circular. For instance if the sample is illuminated with natural light $\mathbf{S'}=(1,0,0,0)$ the DoP of the reflected beam verifies $0.6<P<0.7$. On the other hand, when $\mathbf{S'}=(1,-1,0,0)$, $P>0.9$. In any case, there is not a strong dependence of the DoP as a function of the wavelength.

Then, wavelength average of the DoP as a function of polarization angle ψ $(0º<\psi<90º)$ was calculated. Results are presented in Fig. 26. The error bars indicate the standard deviation. Note that the maximum distance between the curves appears for polarization angles $20º<\psi<35º$. Nevertheless, red curves (4nm) and black curves (no nanoparticles) are very close to each other and almost overlap.
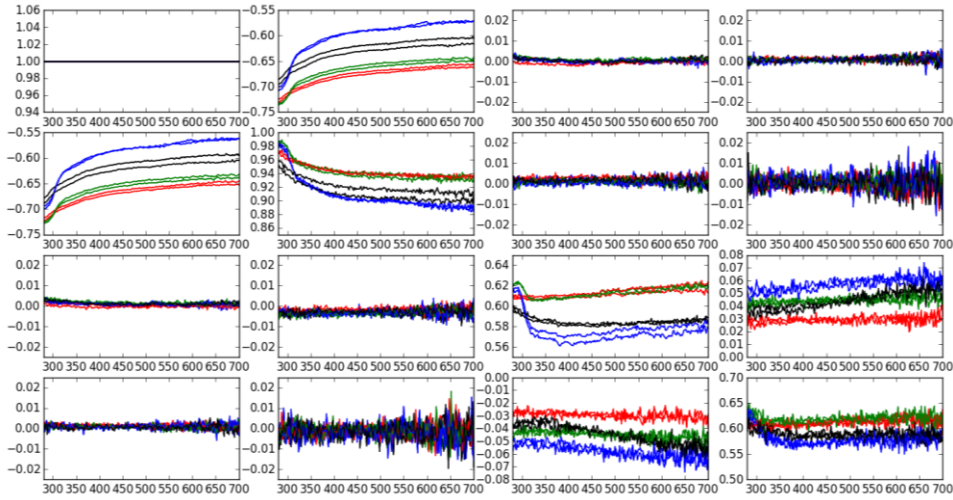
Fig. 24: Mueller matrix components as a function of the wavelength. The wavelength and the Mueller components range from 280 to 700 nm and [-1, 1] respectively. For the sake of clarity, the y-scale of the graphs is adapted to the dynamic range for each component. Two tablets of each class were used. Black: tablets without nanoparticles; red: tablets with 4nm AuNP; green: tablets with 12nm AuNP; blue: tablets with 25nm AuNP. Reprinted from [Opt. Lett., **41**, 4507 (2016)] [46].

Fig. 25: DoP as a function of the wavelength for several input states of polarization: S=(1,0,0,0), S=(1,1,0,0), S=(1,-1,0,0), S=(1,0,1,0), S=(1,0,-1,0) and S=(1,0,0,1). Two tablets of each class were used. Black: tablets without nanoparticles; red: tablets with 4nm AuNP; green: tablets with 12nm AuNP; blue: tablets with 25nm AuNP. Reprinted from [Opt. Lett., **41**, 4507 (2016)] [46].
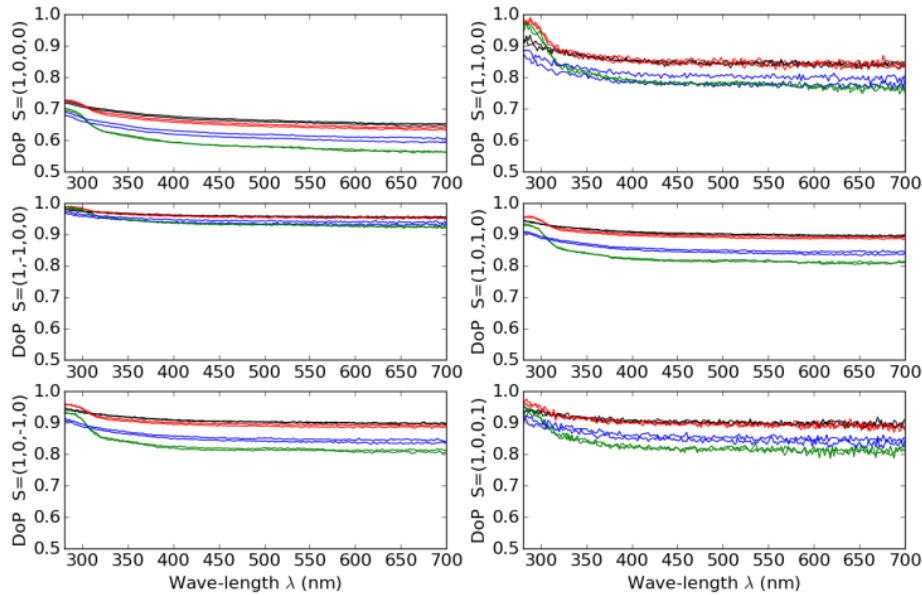
The calculation of the average of the DoP is not enough to perform a trustable classification of the sample. For this reason we used a classification algorithm such as K-NN. Since the sample is characterized by the Mueller matrix, a dataset with an arbitrary number of states of polarization can be generated. We calculate the state of polarization

of the reflected light when the polarization angle ψ ranges from 20° to 35° with a step size of 0.057° (0.001 rad). Then, the DoP as a function of the wavelength for each angle ψ was obtained. Note that DoP curves can cross each other when λ<350 nm (see Fig. 25). For this reason, the wavelengths are limited to $\lambda \in$ [350,700]. The system was trained using a hold-out strategy: half of the values of DoP are used to train the system and the other half for testing purposes. The training and test sets are generated at random. The classifier was trained 1000 times with k=1 (i.e. only the closest neighbor is taken into account). The accuracy of the classification was 1 and no errors were detected.

We performed a second classification test using K-NN (k=1). The DoP was calculated in the same way as explained above, but in this case the sets were not generated at random: the training set was produced by selecting one of the pills of each class; the data obtained from the remaining pill was used as the test set. The 16 possible combinations where taken into account. The accuracy of the classification was again 1 for all cases considered.

In summary, we demonstrated that pills coated with nano particles (AuNP) of different sizes can be successfully classified using polarimetric techniques. This result may open a possible trend in research of anti-counterfeiting of pharmaceutical products.
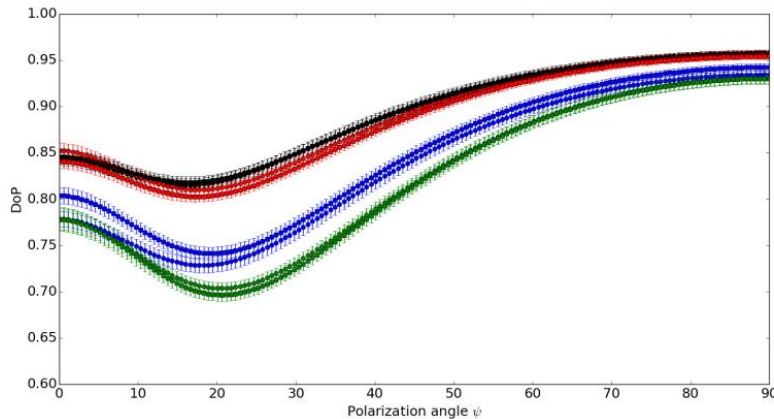

Fig. 26: Averaged DoP as a function of the polarization angle ψ. Two tablets of each class were used. Black: tablets without nanoparticles; red: tablets with 4nm AuNP; green: tablets with 12nm AuNP; blue: tablets with 25nm AuNP. Reprinted from [Opt. Lett., **41**, 4507 (2016)] [46].

9. Authentication of Three Dimensional phase encoded samples

In the previous sections, we analyzed the effect of polarized light on the behavior of codes or pharmaceutical pills produced with AuNP, platinum, chromium, et cetera. The polarimetric signature is used to distinguish among different classes of samples and authentication is performed using machine learning algorithms such as SVM or K-NN, etc. Now, the problem is slightly different because polarization does not play a key role in validation. As we discussed in section 1, phase encoding using adhesive tape produces speckle-like distributions after propagation. In this section, we extend conventional 2D phase encoding to a more general 3D encoding [101-103]. We show that 3D phase masks (3DPM) of different thickness will produce different speckle signals. Consequently, 3D masks attached to codes can be used for authentication purposes [56].
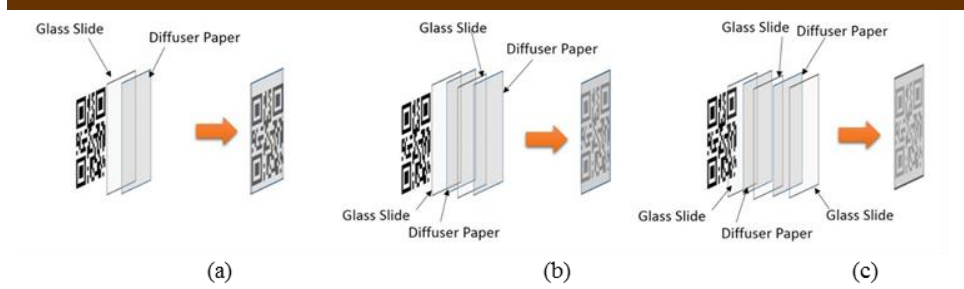
Fig. 27: 3D phase encoded codes: (a) two layers 3D code (code A); (b) four layer s3D code (code B), and (c) five layers 3D code (code C). Adapted from [J. Opt. Soc. Am. A, **33**, 1160 (2016)] [56].

Figure 27(a-c) shows how the samples were prepared. Three 3D phase masks of different thickness were attached to a printed QR code. The 3D phase masks were produced by combining glass and diffuser paper. Note that we use the same QR code for the three cases considered. According to Fig. 27, we refer to the three samples as code A, B, and C. For illustrative purposes, Fig. 28(a) and 28(b) shows the QR code printed on transparency film and the two-layers 3D code (code A), respectively. The size of the codes is 4 mm x 4 mm.

The optical set-up for measuring the speckle signal is presented in Fig. 28(c). An expanded laser diode with wavelength λ=455 nm illuminates the sample. A linear polarizer is used as attenuator. The sample is placed in front of a CCD camera with the following specifications (Table 6):

Table 6: camera specifications

Sensor size	14.9 mm x 22.3 mm
Pixels	2784 x 1856

The camera records the speckle distributions. Distance d is set at 30, 70, 110, and 150 mm. Because speckle patterns can be influenced by vibrations and other surrounding variables, twenty images were recorded for each distance d [59],
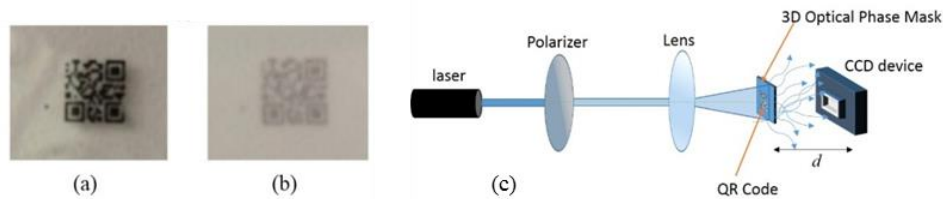

Fig. 28: (a) conventional QR code printed on transparency film, (b) two layer 3D code (class A) and (c) optical setup. Adapted from [J. Opt. Soc. Am. A, **33**, 1160 (2016)] [56].

Fig. 29(a-c) shows recorded images and histograms at d= 110 mm for samples of classes A, B and C. Images are presented in false color in order to better visualize the speckle-like character of these distributions. Note that in Fig. 29(b) the illumination was not uniform due to a slight misalignment of the optical setup. The three histograms look slightly different. It seems that when the number of layers increases, the histogram tends to be narrower. Accordingly, distinctive information could be derived to perform classification. As discussed in section 1, speckle values follow a Gamma distribution probability [Eq. (3)]. Note that the curve that fits each gray level distribution is superimposed on the histograms.
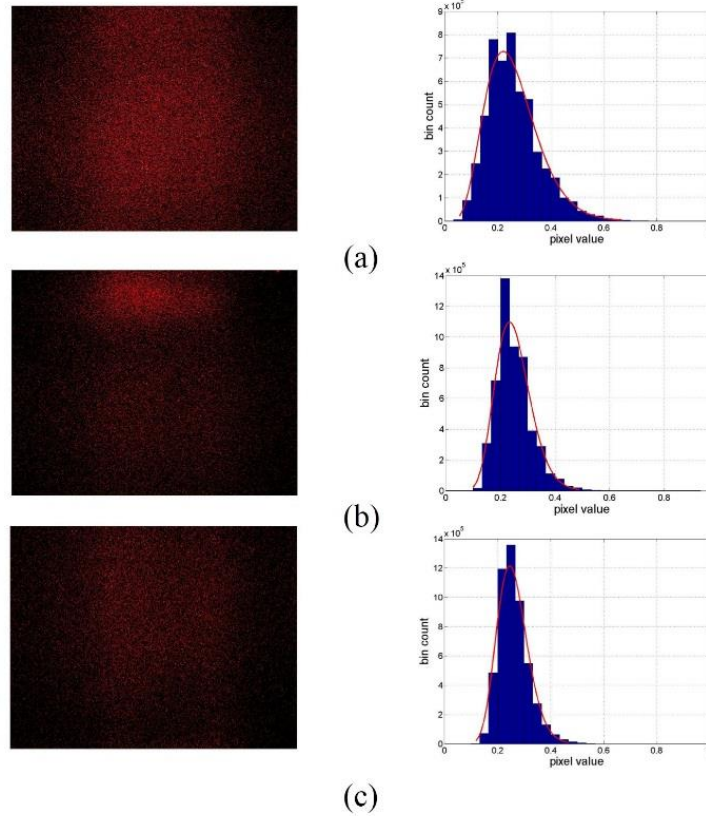
Fig. 29: Recorded speckle distributions and the corresponding histograms for $d = 110$ mm. Gamma distributions fitted to match the histograms are depicted in red: (a) 3D code class A, (b) 3D code class B and, (c) 3D code class C. Reprinted from [J. Opt. Soc. Am. A, **33**, 1160 (2016)] [56].

In the present problem, instead of using the 256-bin histograms to feed the classification system, we calculated the following statistics: mean, variance, skewness, kurtosis and entropy [see Eqs. (41) and (42)]. These features are used to train a Random Forest classification system. A total of twelve classes have been considered according to the information presented in Table 7:

Table 7: 12 classes considered

	Code A	Code B	Code B
d=30 mm	Class 1	Class 2	Class 3
d=70 mm	Class 4	Class 5	Class 6
d=110 mm	Class 7	Class 9	Class 10
d=150 mm	Class 10	Class 11	Class 12

In summary, the system has to determine the code and the distance the measurement was carried out. Each class is described by 20 sets of m=5 features: 10 are used to train the classifier and the remaining 10 are used to test the system. According to the formula that estimates the minimum number of features ν required to create a random tree $v = \left\lfloor \sqrt{m} \right\rfloor$, is v=2 (see section 5.3). The number of random trees is selected by calculating the classification error as a function of the number of trees. The error decays with the number of trees used (see Fig. 30). Note that when the system is trained with 100 random binary trees, the error is less than 1%. Finally, the model is checked with the members of the test set. All test images were successfully classified.
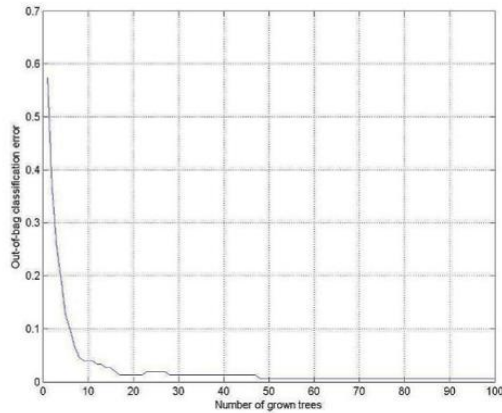
Fig 30: Classification errors as a function of the number of trees used. Reprinted from [J. Opt. Soc. Am. A, **33**, 1160 (2016)] [56].

A second experiment was taken into account. 50 extra speckle patterns were recorded at arbitrary distances *d,* ranging from 30 to 150 mm. In particular, distances *d*=30, 70, 110 and 150 were explicitly excluded. The features obtained from these new images were used to test the classification system. The classifier tried to assign the images to one of the 12 classes of Table 7 using the following criteria: a sample image is assigned to one of the classes if and only if gets 95% of the votes; otherwise, this speckle image is rejected. As a result, none of these 50 speckle images were assigned to one of the 12 possible classes.

In summary, we demonstrate authentication with 3D codes in scattering medium that is identical QR codes, encoded with different 3D phase masks can be distinguished using a Random Forest classifier.

10. Concluding remarks.

In this paper, we have reviewed recent progress on validation and authentication using optics and photonics with nano particle and thin film encoding. Optically encoded data produced using nanoparticles or thin film technologies can be verified using polarimetric speckle analysis and pattern recognition techniques. Experimental methods include Mueller matrix polarimetry and image analysis. Since measured signals for true and counterfeit samples are very similar and difficult to distinguish, statistical and machine learning methods are required to perform successful classification. Future trends on this topic may include preparation of more complex samples with nanoparticles and simplified measuring methods that do not require sophisticated laboratory equipment.

REFERENCES

[1] B. Javidi and J. L. Horner, "Optical pattern recognition for validation and security verification," Opt. Eng. **33**, 1752-1756 (1994).

[2] P. Réfrégier and B. Javidi, "Optical image encryption based on input plane and Fourier plane random encoding," Opt. Lett. **20**, 767-769 (1995).

[3] O. Matoba, T. Nomura, E. Perez-Cabre, M. S. Millan and B. Javidi, "Optical techniques for information security," Proc. IEEE **97**, 1128-1148 (2009).

[4] A. Alfalou and C. Brosseau, "Optical image compression and encryption methods," Adv. Opt. Photonics, **1**, 589-636 (2009)

[5] W. Chen, B. Javidi and X. Chen, "Advances in optical security systems," Adv. Opt. Photonics **6**, 120-155 (2014).

[6] B. Javidi, A. Carnicer, M. Yamaguchi, T. Nomura, E. Pérez-Cabré, M. S. Millán, N. K. Nishchal, R. Torroba, J. F. Barrera, W. He, X. Peng, A. Stern, Y. Rivenson, A. Alfalou et. al. "Roadmap on Optical Security," J. Opt. **16**, 083001 (2016).

[7] O. Matoba and B. Javidi, "Encrypted optical memory system using three-dimensional keys in the Fresnel domain," Opt. Lett. **24**, 762-764 (1999).

[8] G. Unnikrishnan, J. Joseph and K. Singh, "Optical encryption by double-random phase encoding in the fractional Fourier domain," Opt. Lett. **25**, 887-889 (2000).

[9] O. Matoba and B. Javidi, "Encrypted optical storage with wavelength-key and random phase codes," Appl. Opt. 38, 6785-6790, (1999).

[10] E. Tajahuerce, J. Lancis, B. Javidi and P. Andrés, "Optical security and encryption with totally incoherent light," Opt. Lett. **26**, 678-680 (2001).

[11] B. Javidi and T. Nomura, "Polarization encoding for optical security systems," Opt. Eng. **39**, 2439-2443 (2000).

[12] X. Tan, O. Matoba, Y. Okada-Shudo, M. Ide, T. Shimura and K. Kuroda, "Secure optical memory system with polarization encryption," Appl. Opt. **40**, 2310-2315 (2001).

[13] O. Matoba and B. Javidi, "Secure holographic memory by double-random polarization encryption," Appl. Opt. **43**, 2915 -2919 (2004).

[14] J. F. Barrera, R. Henao, M. Tebaldi, R. Torroba and N. Bolognini, "Multiplexing encrypted data by using polarized light," Opt. Commun. **260**, 109-112 (2006).

[15] A. Carnicer, M. Montes-Usategui, S. Arcos and I. Juvells, "Vulnerability to chosen-cyphertext attacks of optical encryption schemes based on double random phase keys," Opt. Lett. **30**, 1644-1646 (2005).

[16] X. Peng, P. Zhang, H. Wei and B. Yu, "Known-plaintext attack on optical encryption based on double random phase keys," Opt. Lett. **31**, 1044 -1046 (2006).

[17] Y. Frauel, A. Castro, T. J. Naughton and B. Javidi, "Resistance of the double random phase encryption against various attacks," Opt. Express **15**, 10253-10265 (2007).

[18] P. Kumar, A. Kumar, J. Joseph and K. Singh, "Impulse attack free double-random-phase encryption scheme with randomized lens-phase functions," Opt. Lett. **34**, 331-333 (2009).

[19] T. J. Naughton, B. M. Hennelly and T. Dowling, "Introducing secure modes of operation for optical encryption," J. Opt. Soc. Am. A **25**, 10, 2608-2617 (2008).

[20] H. Tashima, M. Takeda, H. Suzuki, T. Obi, M. Yamaguchi and N. Ohyama, "Known plaintext attack on double random phase encoding using fingerprint as key and a method for avoiding the attack," Opt. Express **18**, 13772 -13781 (2010).

[21] W. Chen, X. Chen and C. J. Sheppard, "Optical image encryption based on diffractive imaging," Opt. Lett. **35**, 3817-3819 (2010).

[22] W. Chen and X. Chen, "Space-based optical image encryption," Opt. Express **18**, 27095-27104 (2010).

[23] K. Nakano, M. Takeda, H. Suzuki and M. Yamaguchi, "Evaluations of phase-only double random phase encoding based on key-space analysis," Appl. Opt. **52**, 1276-1283 (2013).

[24] F. Mosso, J. F. Barrera, M. Tebaldi, N. Bolognini and R. Torroba, "All-optical encrypted movie," Opt. Express **16**, 5706–5712 (2011).

[25] T. Yatagai, T. Ochiai and D. Barada, "Encryption based on vector wave holography," J. Opt. **14**, 094005 (2012).

[26] N. K. Nishchal and T. J. Naughton, "Flexible optical encryption with multiple users and multiple security levels," Opt. Commun. **284**, 735-739 (2011).

[27] P. Kumar, J. Joseph and K. Singh, "Optical image encryption using a jigsaw transform for silhouette removal in interference-based methods and decryption with a single spatial light modulator," Appl. Opt. **50**, 1805-1811 (2011).

[28] W. Qin and X. Peng, "Asymmetric cryptosystem based on phase-truncated Fourier transforms," Opt. Lett. **35**, 118–120 (2010).

[29] S. K. Rajput and N. K. Nishchal, "An optical encryption and authentication scheme using asymmetric keys," J. Opt. Soc. Am. A **31**, 1233-1238 (2014).

[30] J. Liu, X. Xu, Q. Wu, T. S. J and G. Situ, "Information encryption in phase space," Opt. Lett. **40**, 859–862, (2015).

[31] E. Pérez-Cabré, M. Cho and B. Javidi, "Information authentication using photon-counting double-random-phase encrypted images," Opt. Lett. **36**, 22-24 (2011).

[32] E. Pérez-Cabré, H. C. Abril, M. S. Millán and B. Javidi, "Photon-counting double-random-phase encoding for secure image verification and retrieval," J. Opt. **14**, 094001 (2012).

[33] M. Cho and B. Javidi, "Three-dimensional photon counting double-random-phase encryption," Opt. Lett. **38**, 3198-3201 (2013).

[34] A. Markman and B. Javidi, "Full-phase photon-counting double-random-phase encryption," J. Opt. Soc. Am. A **31**, 394-403 (2014).

[35] D. Maluenda, A. Carnicer, R. Martínez-Herrero, I. Juvells and B. Javidi, "Optical encryption using photon-counting polarimetric imaging," Opt. Express **23**, 655 -666 (2015).

[36] W. Chen, X. Chen, A. Stern and B. Javidi, "Phase-modulated optical system with sparse representation for information encoding and authentication," IEEE Photonics J. **5**, 6900113 (2013).

[37] S. A. Goorden, M. Horstmann, A. P. Mosk, B. Skoric and P. W. Pinkse, "Quantum-secure authentication of a physical unclonable key," Optica **1**, 421-424 (2014).

[38] A. Carnicer, I. Juvells, B. Javidi and R. Martínez-Herrero, "Optical encryption in the longitudinal domain of focused fields," Opt. Express **24**, 6793 -6801 (2016).

[39] A. Carnicer, I. Juvells, B. Javidi and R. Martínez-Herrero, "Optical encryption in the axial domain using beams with arbitrary polarization," Opt. Laser Eng. **89**, 141-145 (2016).

[40] P. Clemente, V. Durán, E. Tajahuerce and J. Lancis, "Optical encryption based on computational ghost imaging," Opt. Lett. **35**, 2391-2393 (2010).

[41] W. Chen and X. Chen, "Ghost imaging for three-dimensional optical security," Appl. Phys. Lett. 103, 221106 (2014).

[42] J. M. Meruga, W. M. Cross, P. S. May, Q. Luu, G. A. Crawford and J. J. Kellar, "Security printing of covert quick response codes using upconverting nanoparticle inks," Nanotechnology **23**, 395201 (2012).

[43] A. Markman, B. Javidi and M. Tehranipoor, "Photon-counting security tagging and verification using optically encoded QR codes," IEEE Photonics J. **6**, 1-9 (2014).

[44] A. Carnicer, A. Hassanfiroozi, P. Latorre-Carmona, Y.-P. Huang and B. Javidi, "Security authentication using phase-encoded nanoparticle structures and polarized light," Opt. Lett. **40**, 2135-2138 (2015).

[45] A. Carnicer, O. Arteaga, E. Pascual, A. Canillas, S. Vallmitjana, B. Javidi and E. Bertran, "Optical security verification by synthesizing thin films with unique polarimetric signatures," Opt. Lett. **40**, 5399-5402, (2015).

[46] A. Carnicer, O. Arteaga, J. M. Suñé and B. Javidi, "Authentication of gold nanoparticle encoded pharmaceutical tablets using polarimetric signatures," Opt. Lett. **41**, 4507-4510 (2016).

[47] M. Naruse, N. Tate and M. Ohtsu, "Optical security based on near-field processes at the nanoscale," . J. Opt. **14**, 094002 (2012).

[48] T. Matsumoto, M. Hoga, Y. Ohyagi, M. Ishikawa, M. Naruse, K. Hanaki, R. Suzuki, D. Sekiguchi, N. Tate and M. Ohtsu, "Nano-artifact metrics based on random collapse of resist," Sci. Rep. **4**, 6142 (2014).

[49] T. Matsumoto, N. Yoshida, S. Nishio, M. Hoga, Y. Ohyagi, N. Tate and M. Naruse, "Optical nano artifact metrics using silicon random nanostructures," Sci. Rep. **6** , 32438 (2016).

[50] H. A. Macleod, *Thin-film optical filters* (CRC, 2001).

[51] W. Gotschy, K. Vonmetz, A. Leitner and F. Aussenegg, "Optical dichroism of lithographically designed silver nanoparticle films," Opt. Lett. **21**, 1099-1101 (1996).

[52] U. Guler and R. Turan, "Effect of particle properties and light polarization on the plasmonic resonances in metallic nanoparticles," Opt. Express **18**, 17322-17338 (2010).

[53] B. K. Canfield, S. Kujala, K. Jefimovs, J. Turunen and M. Kauranen, "Linear and nonlinear optical responses influenced by broken symmetry in an array of gold nanoparticles," Opt. Express **12**, 5418-

5423 (2004).

[54] J. Aaron, E. de La Rosa, K. Travis, N. Harrison, J. Burt, M. J. Yacamán and K. Sokolov, "Polarization microscopy with stellated gold nanoparticles for robust monitoring of molecular assemblies and single biomolecules," Opt. Express **16**, 2153-2167 (2008).

[55] T. Ming, L. Zhao, Z. Yang, H. Chen, L. Sun, J. Wang and C. Yan, "Strong polarization dependence of plasmon-enhanced fluorescence on single gold nanorods," Nano letters **9**, 3896-3903 (2009).

[56] A. Markman, A. Carnicer and B. Javidi, "Security authentication with a three-dimensional optical phase code using random forest classifier," J. Opt. Soc. Am. A **33**, 1160-1165 (2016)

[57] J. W. Goodman, Introduction to Fourier optics (Roberts and Company, 2005).

[58] J. C. Dainty, "The statistics of speckle patterns," Progress in Optics **14**, 1-46 (1977).

[59] J. W. Goodman, Statistical optics (John Wiley & Sons, 2015).

[60] E. W. Weisstein, "Gamma Distribution, From MathWorld--A Wolfram Web Resource.," [Online]. Available: http://mathworld.wolfram.com/GammaDistribution.html . [Accessed 26 09 2016].

[61] E. Collett, *Field guide to polarization* (SPIE, 2005).

[62] O. Arteaga, "Useful Mueller matrix symmetries for ellipsometry," Thin Solid Films **571**, 584-588, (2014).

[63] R. M. A. Azzam, "Propagation of partially polarized light through anisotropic media with or without depolarization: A differential 4 × 4 matrix calculus," J. Opt. Soc. Am. **68**, 1756-1767 (1978).

[64] D. S. Sabatke, M. R. Descour, E. L. Dereniak, W. C. Sweat, S. A. Kemme and G. S. Phillips, "Optimization of retardance for a complete Stokes polarimeter," Opt. Lett. **25**, 802–804 (2000).

[65] F. Goudail and M. Boffety, "Optimal configuration of static polarization imagers for target detection," J. Opt. Soc. Am. A **33**, 9-16 (2016).

[66] R. M. A. Azzam, " Stokes-vector and Mueller-matrix polarimetry [Invited]," J. Opt. Soc. Am. A **33**, 1396-1408 (2016).

[67] O. Arteaga, J. Freudenthal, B. Wang and B. Kahr, "Mueller matrix polarimetry with four photoelastic modulators: theory and calibration," Appl. Opt. **51**, 6805-6817 (2012).

[68] O. Arteaga, M. Baldrís, J. Antó, A. Canillas, E. Pascual and E. Bertran, "Mueller matrix microscope with a dual continuous rotating compensator setup and digital demodulation," Appl. Opt. **53**, 2236-2245 (2014).

[69] R. C. Gonzalez and R. E. Woods, *Digital image processing* (Pearson-Prentice Hall, 2008).

[70] R.E. Bellman, *Adaptive Control Processes*, (Princeton University, 1961).

[71] W. R. Dillon and M. Goldstein, *Multivariate analysis: Methods and applications* (John Wiley & Sons, 1984).

[72] J. Shlens, "A tutorial on principal component analysis: Derivation, Discussion and Singular Value Decomposition," arXiv preprint, arXiv:1404.1100 (2014).

[73] M. Kantardzic, *Data mining: concepts, models, methods, and algorithms* (John Wiley & Sons, 2011).

[74] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory **13**, 21-27 (1967).

[75] N. Japkowicz and M. Shah, *Evaluating learning algorithms: a classification perspective* (Cambridge University, 2011).

[76] "Scikt-learn," [Online]. Available: http://scikit-learn.org/stable/auto_examples/neighbors/plot_classification.html. [Accessed 12 08 2016].

[77] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res **12**, 2825-2830 (2011).

[78] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond* (MIT, 2002).

[79] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," ACM Trans Intell Syst Technol, **2**, 27 (2011).

[80] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," IEEE Trans. Neural Netw. **13**, 415-425 (2002).

[81] L. Breiman, "Random forests," Machine learning, **45**, 5-32 (2001).

[82] M. Fernández-Delgado, E. Cernadas, S. Barro and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems," J. Mach. Learn. Res. **15**, 3133-3181 (2014).

[83] M. Wainberg, B. Alipanahi and B. Frey, "Are Random Forests Truly the Best Classifiers?," J. Mach. Learn. Res. **17**, 1-5 (2016).

[84] F. J. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," JASA **46**, 68-78 (1951).

[85] V.-M. Freire, C. Corbella, E. Bertrán, S. Portal-Marco, M. Rubio-Roy and J.-L. Andújar, "Anisotropic

surface properties of micro/nanostructured aC: H: F thin films with self-assembly applications," J. Appl. Phys. **111**, 124323 (2012).

[86] C. Corbella, B. Echebarria, L. Ramírez-Piscina, E. Pascual, J. Andújar and E. Bertran, "Spontaneous formation of nanometric multilayers of metal-carbon films by up-hill diffusion during growth," Appl. Phys. Lett. **87**, 213117 (2005).

[87] J. L. Horner and P. D. Gianino, "Phase-only matched filtering," Appl. Opt. **23**, 812-816 (1984).

[88] E. F. Fahy and M. A. MacConaill, "Optical Properties of 'Cellophane'," Nature **178**, 1072-1073 (1956).

[89] World Health Organization, "Growing threat from counterfeit medicines," Bull. World Health. Organ. **88**, 247–248 (2010).

[90] D. Bansal, S. Malla, P. Gudala and K. Tiwari, "Anti-counterfeit technologies: a pharmaceutical industry perspective," Sci. Pharm. **81**, 1-4 (2014).

[91] K. Dégardin, Y. Roggo and P. Margot, "Understanding and fighting the medicine counterfeit market," J. Pharm. Biomed. Anal. **87**, 167-175 (2014).

[92] S. Kovacs, S. E. Hawes and S. N. Maley, "Technologies for detecting falsified and substandard drugs in low and middle-income countries," PLoS One **26**, e90601, (2014).

[93] E. Medina, E. Bel and J. Suñé, "Counterfeit medicines in Peru: a retorspective review (1997-2014)," BMJ Open **4**, e010387 (2016)

[94] R. Shukla, V. Bansal, M. Chaudhary, A. Basu, R. R. Bhonde and M. Sastry, "Biocompatibility of gold nanoparticles and their endocytotic fate inside the cellular compartment: a microscopic overview," Langmuir **21**, 10644-10654 (2005).

[95] P. Ghosh, G. Han, M. De, C. K. Kim and V. M. Rotello, "Gold nanoparticles in delivery applications," Adv. Drug Deliv. Rev. **60**, 1307-1315 (2008).

[96] E. Boisselier and D. Astruc, "Gold nanoparticles in nanomedicine: preparations, imaging, diagnostics, therapies and toxicity," Chem. Soc. Rev. **38**, 1759-1782 (2009).

[97] R. A. Sperling, P. R. Gil, F. Zhang, M. Zanella and W. J. Parak, "Biological applications of gold nanoparticles," Chem. Soc. Rev. **37**, 1896-1908 (2008).

[98] D. A. Giljohann, D. S. Seferos, W. L. Daniel, M. D. Massich, P. C. Patel and C. A. Mirkin, "Gold nanoparticles for biology and medicine," Angewandte Chemie International Edition **49**, 3280-3294 (2010).

[99] A. J. Mieszawska, W. J. Mulder, Z. A. Fayad and D. P. Cormode, "Multifunctional gold nanoparticles for diagnosis and therapy of disease," Mol. Pharm. **10**, 831-847 (2013).

[100] P. K. Jain, K. S. Lee, I. H. El-Sayed and M. A. El-Sayed, "Calculated absorption and scattering properties of gold nanoparticles of different size, shape, and composition: applications in biological imaging and biomedicine," J. Phys. Chem. B **110**, 7238-7248 (2006).

[101] O. Matoba, T. Sawasaki and K. Nitta, "Optical authentication method using a three-dimensional phase object with various wavelength readouts," Appl. Opt. **47**, 4400-4404 (2008).

[102] O. Matoba, S. Matsuki and K. Nitta, "Secure data storage by three-dimensional absorbers in highly scattering volume medium," in Journal of Physics: Conference Series **139,** 1 (2008).

[103] O. Matoba, K. Nitta and W. Watanabe, "Secure Data Storage using 3D Scattering Medium," in Conference on Lasers and Electro-Optics CLEO/Pacific RIM (2009)

Artur Carnicer (Barcelona, 1965) received the Licentiate degree and the Ph.D. in Physics from the University of Barcelona (UB). Currently, he is an associate professor in optics with the Applied Physics Department, UB. From 2003 to 2006, Prof. Carnicer was appointed as director of studies at the UB School of Physics. In 2010, he was awarded with the MERLOT Classic Award (Physics) for the JOptics Course, a collection of open source simulations that cover important undergraduate topics in optics. His scientific interests include, but are not limited to, polarization and polarimetric techniques, highly focused fields, optical security and 2D and 3D visualization systems. He is member of OSA, SPIE, SedOPTICA (Spanish Optical Society) and Catalan Physical Society.

Prof. Bahram Javidi received the B.S. degree from George Washington University and the Ph.D. from the Pennsylvania State University in electrical engineering. He is Board of Trustees Distinguished Professor at University of Connecticut. Prof. Javidi's interests are in a broad range of transformative imaging approaches using optics and photonics, and he has made seminal contributions to passive and active multi-dimensional imaging from nano- to micro- and macroscales. His recent research activities include 3D visualization and recognition of objects in photon-starved environments; automated disease identification using biophotonics with low cost compact sensors; information security, encryption, and authentication using quantum imaging; non-planar flexible 3D image sensing, and bio-inspired imaging. Prof. Javidi has been named Fellow of eight societies, including IEEE, OSA, SPIE, EOS, and IoP. Early in his career, the National Science Foundation named him a Presidential Young Investigator. Prof. Javidi has received the Prize for Applied Aspects of Quantum Electronics and Optics of European Physical Society (2015), the SPIE Dennis Gabor Award in Diffractive Wave Technologies (2005), and the SPIE Technology Achievement Award (2008). In 2008, he was awarded the IEEE Donald G. Fink Paper Prize, and the John Simon Guggenheim Foundation Fellow Award. In 2007, the Alexander von Humboldt Foundation (Germany) awarded Prof. Javidi the Humboldt Prize. His papers have been cited over 31000 times according to Google Scholar citation report.