

# On the Mahalanobis Distance Classification Criterion for Multidimensional Normal Distributions

Guillermo Gallego, Carlos Cuevas, Raúl Mohedano, and Narciso García

**Abstract**—Many existing engineering works model the statistical characteristics of the entities under study as normal distributions. These models are eventually used for decision making, requiring in practice the definition of the classification region corresponding to the desired confidence level. Surprisingly enough, however, a great amount of computer vision works using multidimensional normal models leave unspecified or fail to establish correct confidence regions due to misconceptions on the features of Gaussian functions or to wrong analogies with the unidimensional case. The resulting regions incur in deviations that can be unacceptable in high-dimensional models.

Here we provide a comprehensive derivation of the optimal confidence regions for multivariate normal distributions of arbitrary dimensionality. To this end, firstly we derive the condition for region optimality of general continuous multidimensional distributions, and then we apply it to the widespread case of the normal probability density function. The obtained results are used to analyze the confidence error incurred by previous works related to vision research, showing that deviations caused by wrong regions may turn into unacceptable as dimensionality increases. To support the theoretical analysis, a quantitative example in the context of moving object detection by means of background modeling is given.

**Index Terms**—multidimensional signal processing, uncertainty, classification algorithms, Gaussian distribution, Chi-squared distribution, Mahalanobis distance.

## I. INTRODUCTION

**I**N recent years, countless scientific and engineering works in several areas proposing strategies based on probabilistic analyses [1] have been developed. Many of these works [2] use multivariate normal distributions, or mixtures of them [3], due to the satisfactory continuity, differentiability and locality properties of the Gaussian function [4].

Often, the algorithms proposed in these works require the computation of cumulative probabilities of the normal distributions for different purposes such as, for example, disregarding the data that does not contribute significantly to the distributions [5] or evaluating how well a normal distribution represents a data set [6].

The computation of these cumulative probabilities on multivariate normal distributions is not trivial due to misconceptions on their features or to wrong analogies with the unidimensional case. Thus, several authors establish erroneous criteria to compute such probabilities, which, in high-dimensional models,

can produce unacceptable results. On the one hand, some authors [7][8][9] erroneously consider that the cumulative probability of a generic  $n$ -dimensional normal distribution can be computed as the integral of its probability density over a hyper-rectangle (Cartesian product of intervals). On the other hand, in different areas such as, for example, face recognition [10] or object tracking in video data [11], the cumulative probabilities are computed as the integral over (hyper-)ellipsoids with inadequate radius.

Among the multiple engineering areas where the aforementioned errors are found, moving object detection using Gaussian Mixture Models (GMMs) [12] must be highlighted. In this field, most strategies proposed during the last years take as a starting point the work by Stauffer and Grimson [13], which is a seminal work with more than one thousand citations. Their algorithm states that a sample is correctly modeled by a Gaussian distribution if it is within 2.5 standard deviations from the mean of the distribution. However, the authors leave unspecified two key details: *i*) the way the distance from a sample to the mean is measured and *ii*) the relation of the 2.5 threshold with respect to both a target confidence value and the dimension of the GMM model.

In spite of the omitted details, a very significant amount of recent scientific works rely on [13] to set their classification criterion. Some of these works [14][15][16] establish hyper-rectangular decision regions by imposing the condition established by [13] separately on each channel (dimension). Other approaches [17][18][19] use the Mahalanobis distance to the mean of the multidimensional Gaussian distribution to measure the goodness of fit between the samples and the statistical model, resulting in ellipsoidal confidence regions. Finally, a third group of treatises [20][21][22][23][24] mimic the description in [13] and therefore do not disambiguate the shape of the decision region. Similarly to [13], most of these works set the classification threshold to 2.5 regardless of the dimensionality of the Gaussian model and the shape of the confidence region considered. However, the use of a fixed threshold causes larger deviations of the confidence level with respect to the one-dimensional case as the dimensionality increases, which may produce unacceptable results.

Here, to prevent the propagation of these errors into future works, we present a novel and helpful analysis to determine, given a target cumulated probability, the correct confidence region for a multivariate normal model of arbitrary dimensionality and general covariance matrix. In Section II-A we discuss the unidimensional case, while in Sections II-B and II-C we extend this discussion to multiple dimensions. Firstly, in Section II-B, we prove that, for a broad class of distributions,

Manuscript received in xxx x, 2012. This work has been supported in part by the Ministerio de Economía y Competitividad of the Spanish Government under project TEC2010-20412 (Enhanced 3DTV).

G. Gallego, C. Cuevas, R. Mohedano, and N. García are with Grupo de Tratamiento de Imágenes (GTI), ETSI Telecomunicación, Universidad Politécnica de Madrid (UPM), 28040 Madrid, Spain (e-mail: ggb,ccr,rmp,narciso@gti.ssr.upm.es)

the optimal region accumulating a target confidence level is within an equidensity contour of the probability density function (PDF). Secondly, in Section II-C, we illustrate how to determine the probability accumulated inside an equidensity contour of a multivariate normal distribution. In Section III we compare and analyze the proposed classification criteria to those applied in previous works, showing that the deviations caused by erroneously selected confidence regions may result unacceptable as dimensionality increases. Finally, the conclusions of this work are presented in Section IV.

## II. ANALYSIS OF A GENERIC MULTIDIMENSIONAL NORMAL DISTRIBUTION

### A. Motivation from the unidimensional normal distribution

The 68-95-99.7 rule states that 68%, 95% and 99.7% of the values drawn from a normal distribution are within 1, 2 and 3 standard deviations  $\sigma > 0$  away from the mean  $\mu$ , respectively. In general, for non-integer  $w$ , the probability contained in the symmetric interval (region)  $R := [\mu - w\sigma, \mu + w\sigma]$  around the mean is

$$P\left(\frac{|x - \mu|}{\sigma} \leq w\right) = \operatorname{erf}\left(\frac{w}{\sqrt{2}}\right), \quad (1)$$

where  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$  is the error function. Observe that, since the PDF  $f_X$  of a normal distribution is symmetric about its mean  $\mu$  and  $R$  has been chosen symmetric about  $\mu$ ,  $R$  is equivalently characterized by the value of the PDF at the endpoints according to

$$R = \{x \in \mathbb{R} \text{ such that } f_X(x) \geq f_0\}, \quad (2)$$

where  $f_0 := f_X(\mu - w\sigma) = f_X(\mu + w\sigma)$ .

Equation (1) establishes a one-to-one correspondence between the region  $R$  (characterized by  $f_0$ ) and the probability contained within it. Hence, given a probability value, it is possible to find the corresponding region.

Motivated by the previous choice of region  $R$ , we consider the optimality criterion for the determination of confidence regions. First, we focus on distributions whose PDFs have a single mode (local maximum), such as the normal distribution or the beta distribution  $\operatorname{Beta}(\alpha, \beta)$  with  $\alpha > 1, \beta > 1$ . Then, we discuss the case of multimodal distributions.

Consider the context of interval estimates [25, p. 307] of an unknown parameter  $\theta$  from noisy observations  $z_i = \theta + \nu_i$ . To draw a conclusion about the true value of the parameter, the goal is the determination of the smallest interval  $[\theta_1, \theta_2]$  accumulating a given probability (or confidence level) that the parameter is contained in it,  $P(\theta \in [\theta_1, \theta_2]) = P_0$ .

Similarly, this notion of *smallest size* also drives the determination of optimal decision regions for data classification. Observe that the above  $R$  for a normal distribution satisfies such optimality condition: it is the smallest interval  $[a, b]$  containing a target confidence level  $P_0$ . This is easy to prove since  $R$  is the solution of the constrained optimization problem

$$\min_{a,b} (b - a) \text{ subject to } P(x \in [a, b]) = P_0 \quad (3)$$

and assuming  $b > a$ . Using a Lagrange multiplier  $\lambda \in \mathbb{R}$ , the minimizer of (3) is among the extremals of

$$L(a, b, \lambda) := (b - a) + \lambda (P_0 - P(x \in [a, b])), \quad (4)$$

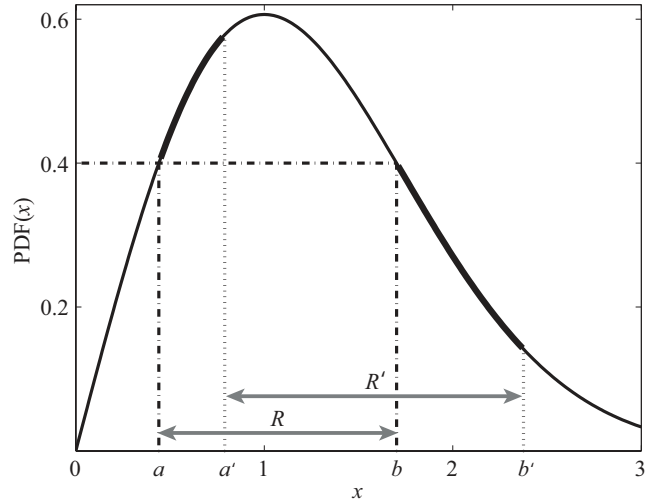


Fig. 1. Optimal confidence region in one-dimensional PDFs. The optimal interval  $R = [a, b]$  containing a given confidence  $P_0$  is compared to a different interval  $R' = [a', b']$  having the same confidence  $P_0$ . Interval  $R$  is smaller than interval  $R'$ , and the value  $f_X(a) = f_X(b)$  fully characterizes  $R$ .

where  $P(x \in [a, b]) = \int_a^b f_X(x) dx$ . The extremals of (4) are among the solutions of the system of equations

$$\begin{aligned} 0 &= \partial L / \partial \lambda = P_0 - P(x \in [a, b]), \\ 0 &= \partial L / \partial a = -1 + \lambda f_X(a), \\ 0 &= \partial L / \partial b = 1 - \lambda f_X(b), \end{aligned}$$

which is obtained using Leibniz's integral rule. Since, in general  $\lambda \neq 0$ , we conclude from the last two equations that  $f_X(a) = f_X(b)$ , i.e., the endpoints (boundary) of the optimal  $R$  have the same density value. Since  $R = [a, b]$  must contain the regions of maximum density (otherwise its size would not be minimal), we conclude that it is of the form (2) with  $f_0 := f_X(a)$ . This is illustrated in Fig. 1 for a non-symmetric unimodal distribution. It compares the optimal interval  $R = [a, b]$  to a different interval  $R' = [a', b']$ , both accumulating a given probability  $P_0$ . Omitting the common part,  $R \cap R' = [a', b]$ , it is clear that  $R$  is smaller than  $R'$  due to the size of the non-overlapping parts: interval  $[a, a']$  is smaller than interval  $(b, b']$ , since, both accumulating the same probability, the PDF is higher in  $[a, a']$  than in  $(b, b']$  (highlighted PDF segments in bold solid line). Fig. 1 also shows the optimality condition satisfied by  $R$ , i.e.,  $f_X(a) = f_X(b)$ . Moreover, such density value completely characterizes the optimal interval. For any symmetric PDF about its mean  $\mu$  (e.g. normal distribution), the previous result implies that the interval  $[a, b]$  is centered at  $\mu$ , i.e.,  $a = \mu - \Delta$  and  $b = \mu + \Delta$ , for  $\Delta > 0$ . In contrast, the optimal interval in Fig. 1 is not symmetric about the mean of the distribution since we considered an asymmetric PDF.

Let us now consider distributions with one or more modes, say  $\operatorname{Beta}(\alpha, \beta)$  for almost arbitrary  $\alpha > 0, \beta > 0$ . Then, problem (3) should be restated as the search of the smallest union of intervals accumulating a target probability. Still, it would be required that the PDF of the distribution have no regions of constant value (i.e., flat regions) within its support so that an optimal region of the form (2) exists for every

possible value of  $0 < P_0 < 1$ . This condition would rule out distributions such as  $\text{Beta}(\alpha, \beta)$  with  $\alpha = \beta = 1$  (the uniform  $[0, 1]$  distribution).

### B. Equidensity contours: extremal volume property

We will extend the previous discussion to multiple dimensions. The next result will prove that, for a broad class of distributions, the optimal region of  $n$ -space accumulating a target confidence level is contained inside an equidensity contour of the PDF. This is the generalization to multiple dimensions of the solution to problem (3).

**Result 1.** *The smallest region  $R \subseteq \mathbb{R}^n$  containing a target probability  $0 < P_0 < 1$  of a distribution whose PDF  $f_X$  has no constant regions within its support is given by the interior of an equidensity contour of the PDF:*

$$R = \{\mathbf{x} \in \mathbb{R}^n \text{ such that } f_X(\mathbf{x}) \geq f_0\}, \quad (5)$$

where  $P_0 = \int_R f_X(\mathbf{x}) d\mathbf{x}$  and  $f_X(\partial R) = f_0$  at the boundary of  $R$ .

*Proof:* Let us solve the constrained optimization problem

$$\min_{R \subseteq \mathbb{R}^n} \text{Vol}(R) \quad \text{subject to} \quad P(R) = P_0, \quad (6)$$

where  $\text{Vol}(R) = \int_R d\mathbf{x}$  measures the size of  $R$  by means of its volume, the natural measure in  $\mathbb{R}^n$ , and  $P(R) = \int_R f_X(\mathbf{x}) d\mathbf{x}$ . Using a Lagrange multiplier  $\lambda \in \mathbb{R}$ , the solution of (6) is among the extremals of

$$\begin{aligned} F &:= \text{Vol}(R) + \lambda (P_0 - P(R)) \\ &= \lambda P_0 + \int_R (1 - \lambda f_X(\mathbf{x})) d\mathbf{x}. \end{aligned}$$

The extremals satisfy a vanishing necessary optimality condition with respect to the variables  $\lambda$  and  $R$ . As expected,  $\partial F / \partial \lambda = 0$  yields the constraint,  $P(R) = P_0$ . To compute the sensitivity of  $F$  with respect to  $R$ , assume that  $R$  is a region varying smoothly with respect to a parameter  $t \in (-\epsilon, \epsilon)$ ,  $\epsilon > 0$ ,  $R(t) \subset \mathbb{R}^n$ , such that  $R(0) = R$  is the original region. Then,  $F$  also depends on  $t$  and, assuming  $F$  and  $dF/dt$  are both continuous in an open set containing  $(-\epsilon, \epsilon)$ , we may compute  $dF/dt$  using Leibniz's integral rule

$$\begin{aligned} &\frac{d}{dt} \int_{R(t)} g(\mathbf{x}, t) d\mathbf{x} \\ &= \int_{R(t)} \frac{\partial}{\partial t} g(\mathbf{x}, t) d\mathbf{x} + \int_{B(t)} \left\langle \frac{\partial \mathbf{x}}{\partial t}(\sigma), g(\mathbf{x}(\sigma), t) \mathbf{N}(\sigma) \right\rangle d\sigma, \end{aligned}$$

where  $g(\mathbf{x}, t) \in \mathbb{R}$ ,  $B(t) := \partial R(t)$  is the boundary of  $R(t)$ ,  $\langle \mathbf{x}, \mathbf{y} \rangle$  is the Euclidean inner product in  $\mathbb{R}^n$ ,  $\mathbf{N}$  is the outward unit normal to  $B$ ,  $\sigma$  is a local parametrization of  $B$ ,  $(\partial \mathbf{x} / \partial t)(\sigma)$  is the velocity of the boundary, and  $d\sigma$  is the area element on  $B$ .

In our case,  $g(\mathbf{x}, t) := 1 - \lambda f_X(\mathbf{x})$  does not depend on  $t$ , so only the boundary integral (flux) survives in Leibniz's rule:

$$\frac{dF}{dt} = \int_{B(t)} \left\langle \frac{\partial \mathbf{x}}{\partial t}(\sigma), g(\mathbf{x}(\sigma), t) \mathbf{N}(\sigma) \right\rangle d\sigma.$$

The first order Taylor expansion of  $F$  with respect to  $t$  is  $F \approx F|_{t=0} + (dF/dt)|_{t=0} t$ . Letting  $\mathbf{W} := \frac{\partial \mathbf{x}}{\partial t}|_{t=0}$  be the

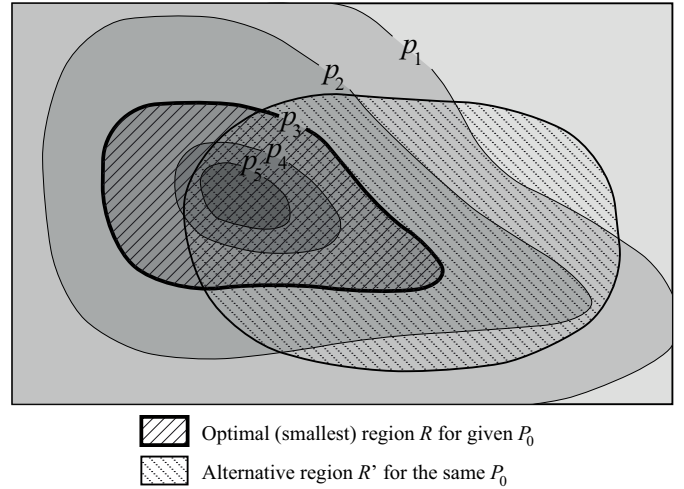


Fig. 2. Graphical illustration of Result 1. A bidimensional PDF is represented by its equidensity contours corresponding to a monotonically increasing sequence of density values ( $p_i < p_j, \forall i < j$ ). The optimal confidence region  $R$  for a given probability  $P_0$  (i.e., the interior of an equidensity contour), is compared to an arbitrary region  $R'$  accumulating the same probability  $P_0$ .

velocity field of the boundary  $B(0)$ , the necessary optimality condition of  $F$  with respect to  $R$  is

$$0 = \frac{dF}{dt} \Big|_{t=0} = \int_{B(0)} \langle \mathbf{W}(\sigma), g(\mathbf{x}(\sigma), 0) \mathbf{N}(\sigma) \rangle d\sigma.$$

for every admissible (smooth)  $\mathbf{W}$ . This implies that  $0 = g(\mathbf{x}(\sigma), 0) = 1 - \lambda f_X(\mathbf{x}(\sigma)) \forall \sigma$ , and, since  $\lambda$  is a scalar, we conclude that  $f_X(B) = f_0$  for some constant  $f_0$ . If we restrict our attention to the class of density functions  $f_X$  that have no constant regions in its support, the set given by the inverse image  $B = f_X^{-1}(f_0)$  will have a zero  $n$ -dimensional measure, a required condition to be a properly defined boundary of some volume in  $\mathbb{R}^n$ . Consequently, the boundary of  $R$  is an equidensity contour of  $f_X$ . Observe that  $R$  must contain the regions of maximum density since otherwise the volume would not be minimal; therefore we conclude that  $R$  is of the form (5). The previous assumption on the lack of constant  $f_X$  regions also implies that an optimal region of the form (5) exists for every possible value of  $0 < P_0 < 1$ .

To show that  $R$  in (5) is not only an extremal but also minimizes the volume while accumulating a probability  $P_0 = \int_R f_X(\mathbf{x}) d\mathbf{x}$ , take any other region  $R'$  of the same size as  $R$  and show that it contains less density mass than  $R$ . Firstly, define new regions by removing the common part,  $R_1 := R \setminus (R \cap R')$ ,  $R_2 := R' \setminus (R \cap R')$ , that is,  $R_2 = \{\mathbf{x} \in R' \subset \mathbb{R}^n \text{ such that } f_X(\mathbf{x}) < f_0\}$ . Note that  $V := \text{Vol}(R_1) = \text{Vol}(R_2)$  because  $\text{Vol}(R) = \text{Vol}(R')$ . Secondly, compare the density mass within  $R$  and  $R'$ . Removing the common part, it is easier to compare  $R_1$  and  $R_2$ ,

$$\begin{aligned} P_1 &= \int_{R_1} f_X(\mathbf{x}) d\mathbf{x} \geq \int_{R_1} f_0 d\mathbf{x} = f_0 \text{Vol}(R_1) = f_0 V, \\ P_2 &= \int_{R_2} f_X(\mathbf{x}) d\mathbf{x} < \int_{R_2} f_0 d\mathbf{x} = f_0 \text{Vol}(R_2) = f_0 V. \end{aligned}$$

Therefore,  $P_2 < P_1$ , and consequently, adding common mass  $P(R \cap R')$  to both sides gives  $P(R') < P(R) = P_0$ . ■

At the end of Section II-A, we showed a simpler version of Result 1 and its application to normal distributions. Observe that, in the unidimensional case (2), the interval  $R$  complies with Result 1. In the  $n$ -dimensional case, Result 1 allows us to state that the optimal confidence region (in the sense of minimum volume) accumulating a given probability is the interior of an equidensity contour, and not an hyper-rectangle, as assumed in many previous works [7][8][14][15] that use normal distributions, as we will see.

Figure 2 illustrates Result 1 for an arbitrary bidimensional PDF. In the figure, the optimal region  $R$  (of volume, i.e., area  $V$ ) accumulating a given confidence  $P_0$  is depicted: the interior of an equidensity contour, as stated in Result 1. A different region  $R'$  (of volume  $V'$ ) accumulating the same confidence is also shown. To compare both regions, we obviously exclude the common part  $R \cap R'$ . Then, it becomes visually clear that points in the non-overlapping subregion  $R_2 = R' \setminus (R \cap R')$  have lower probability density than points within  $R_1 = R \setminus (R \cap R')$ , by the definition of  $R$ . The equidensity contours of the example PDF show this fact. Because of this and the fact that  $R$  and  $R'$  accumulate the same density mass (and consequently so do  $R_1$  and  $R_2$ ), we conclude that  $R_2$  must be bigger than  $R_1$ . Therefore,  $V < V'$ , i.e.,  $R$  is the smallest region for the given level  $P_0$ .

Result 1 matches an insightful argument: given a region  $R' \subset \mathbb{R}^n$  whose boundary does not have a constant density mass  $f_X(\partial R') \neq f_0$ , it is possible to shrink the region while holding a constant mass by swapping large portions of small density for small portions of high density. This deformation process stops if it is no longer possible to exchange regions, i.e., if compared regions have equal size and mass within, even at infinitesimal scale. Hence, the resulting region ( $R$ ) will have a constant density at the boundary. Moreover, the minimum density within the region will be achieved at the boundary.

### C. Probability accumulated in an optimal confidence region of a generic $n$ -dimensional normal distribution

Let us focus on the multivariate normal distribution [26], and, given that the optimal confidence regions are equidensity contours (see Result 1), let us compute the probability accumulated inside such regions in a self-contained manner and show well known connections with the Mahalanobis distance and the  $\chi^2$  distribution. The result will be the generalization of (1), origin of the 68-95-99.7 rule.

Let  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$  follow an  $n$ -dimensional normal distribution with mean  $\boldsymbol{\mu} \in \mathbb{R}^n$  and (symmetric, positive definite) covariance matrix  $\Sigma \in \mathbb{R}^{n \times n}$ . Let the Mahalanobis distance between two vectors  $\mathbf{x}, \mathbf{y}$  according to the weighting matrix  $\Sigma$  be

$$d_{\text{Mah}}(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|_{\Sigma}, \quad (7)$$

with

$$\|\mathbf{u}\|_{\Sigma} := \sqrt{\mathbf{u}^{\top} \Sigma^{-1} \mathbf{u}},$$

then the PDF of  $\mathbf{X}$  can be written as

$$f_X(\mathbf{x}) := (2\pi)^{-\frac{n}{2}} (\det(\Sigma))^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \|\mathbf{x} - \boldsymbol{\mu}\|_{\Sigma}^2\right). \quad (8)$$

In the one-dimensional (1-D) case, the standard score (or  $z$ -score)  $z = (x - \mu)/\sigma$  is a dimensionless quantity whose absolute value measures how far an observation or datum is from the mean. Similarly, in the multivariate ( $n$ -D) case, the Mahalanobis distance  $d_{\text{Mah}}(\mathbf{x}, \boldsymbol{\mu})$  provides such a measure. The equidensity contours of  $f_X$ , characterized by constant  $d_{\text{Mah}}(\mathbf{x}, \boldsymbol{\mu})$ , are ellipsoids centered at  $\boldsymbol{\mu}$  whose axes are aligned with the eigenvectors of  $\Sigma$  [27]. The specific equidensity contour accumulating a target probability can be found using the property that the squared Mahalanobis distance follows a  $\chi^2$  distribution [26, p.86]. The next result shows an alternative and self-contained derivation of these statements.

**Result 2.** For the normal distribution  $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ , whose PDF is (8), the probability accumulated in the region

$$R := \{\mathbf{x} \in \mathbb{R}^n \text{ such that } d_{\text{Mah}}(\mathbf{x}, \boldsymbol{\mu}) \leq d\} \quad (9)$$

is

$$P(R) := P(\|\mathbf{x} - \boldsymbol{\mu}\|_{\Sigma} \leq d) = \frac{\gamma\left(\frac{n}{2}, \frac{d^2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \quad (10)$$

where  $\Gamma(p)$  is the Gamma function and  $\gamma(p, x)$  is the lower incomplete Gamma function.

*Proof:* Let us directly evaluate  $P(R) = \int_R f_X(\mathbf{x}) d\mathbf{x}$ . Just as in the 1-D case it is possible to relate any normal random variable  $\mathcal{N}(\mu, \sigma^2)$  to the standard normal  $\mathcal{N}(0, 1)$  by a standardization process, so it is in the  $n$ -D case. This is implemented by substituting  $\mathbf{y} = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu})$ , whose Jacobian is  $|\det(\partial \mathbf{x} / \partial \mathbf{y})| = \det(\Sigma^{\frac{1}{2}})$ , in the integral  $P(R)$ . The matrix of the change of variables, which satisfies  $\Sigma^{\frac{1}{2}}(\Sigma^{\frac{1}{2}})^{\top} = \Sigma$ , can be obtained by Cholesky or singular value decompositions. The geometric interpretation of the standardization is the concatenation of three operations: a translation of  $\boldsymbol{\mu}$  to the origin, a rotation that aligns the normal PDF with the coordinate axes, and a different scaling in each axis. In the new coordinate system,  $R$  becomes the spherical region  $\{\mathbf{y} \text{ such that } \|\mathbf{y}\| \leq d\}$ , where  $\|\cdot\|$  stands for the Euclidean norm in  $\mathbb{R}^n$ . Hence,

$$\begin{aligned} & \int_R f_X(\mathbf{x}) d\mathbf{x} \\ &= (2\pi)^{-\frac{n}{2}} (\det(\Sigma))^{-\frac{1}{2}} \int_{\|\mathbf{y}\| \leq d} \exp\left(-\frac{1}{2} \|\mathbf{y}\|^2\right) \det(\Sigma^{\frac{1}{2}}) d\mathbf{y} \\ &= (2\pi)^{-\frac{n}{2}} \int_{\|\mathbf{y}\| \leq d} \exp\left(-\frac{1}{2} \|\mathbf{y}\|^2\right) d\mathbf{y}. \end{aligned} \quad (11)$$

Let  $I$  be the integral in (11) and observe that its integrand is rotationally invariant, corresponding to the standard normal distribution  $\mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$ . Hence, in spherical coordinates, with  $r = \|\mathbf{y}\|$ , we have

$$I = \int_0^d \int_{S^{n-1}(r)} e^{-r^2/2} dA dr,$$

where  $S^{n-1}(r)$  is the  $(n-1)$ -sphere of radius  $r$ , and the  $n$ -dimensional volume element  $d\mathbf{y} := dA dr$  splits into the length and area elements. Since the surface area of an

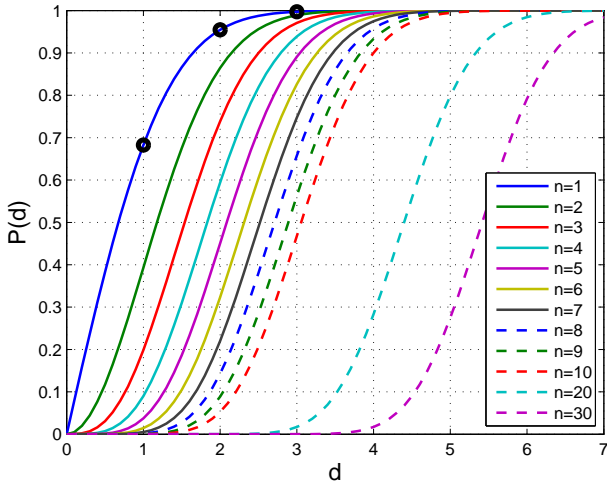


Fig. 3. Curves representing Eq. (10) for  $n = \{1, 2, \dots, 9, 10, 20, 30\}$ . The three circular markers illustrate the 68-95-99.7 rule on the curve representing Eq. (1).

$(n-1)$ -sphere of radius  $r$  is  $A_{n-1}(r) = r^{(n-1)} A_{n-1}(1)$ , with  $A_{n-1}(1) = 2\pi^{n/2}/\Gamma(n/2)$ , substituting  $t = r^2/2$ , we have

$$\begin{aligned} I &= \int_0^d e^{-r^2/2} A_{n-1}(r) dr \\ &= A_{n-1}(1) 2^{\frac{n}{2}-1} \int_0^{d^2/2} t^{\frac{n}{2}-1} e^{-t} dt, \end{aligned} \quad (12)$$

where the last integral is a form of the lower incomplete Gamma function

$$\gamma(p, x) = \int_0^x t^{p-1} e^{-t} dt. \quad (13)$$

Collecting results (11), (12) and (13) gives formula (10). ■ The right hand side of (10) coincides with the value of the cumulative distribution function (CDF) of a (centered) chi-squared distribution with  $n$  degrees of freedom,  $\chi_n^2$ , for a  $d^2$  abscissa, i.e.,  $F_{\chi_n^2}(d^2)$ . This is due to the fact that, after the standardization, the sum of the squares of  $n$  i.i.d. Gaussians is a chi-squared distribution with  $n$  degrees of freedom. Hence, as already announced, the squared Mahalanobis distance follows a  $\chi_n^2$  distribution.

Equation (10) generalizes (1) to the  $n$ -dimensional setting. If  $n = 1$ , the Mahalanobis distance reduces to the z-score ( $\|\mathbf{x} - \boldsymbol{\mu}\|_{\Sigma} = |x - \mu|/\sigma$ ) and, using formulas for special values of the Gamma functions, (10) reduces to (1) as expected. Moreover, since  $\gamma(p, x)$  in (13) is a strictly increasing function of  $x$ , (10) also establishes a one-to-one correspondence between the region  $R$  (characterized by  $d > 0$  in (9)) and the probability it contains, as it is shown in Fig. 3 for different dimensions  $n = \{1, 2, \dots, 9, 10, 20, 30\}$ . Thus, given a probability value  $0 < P_0 < 1$ , the value  $d > 0$  that specifies the equidensity contour bounding the region  $R$  for which  $P_0 = P(R)$  can be obtained by inverting (10). Scientific software (such as the GNU Scientific Library) provides routines to evaluate the well known Gamma functions in (10).

Some special cases deserve further comments. If  $n$  is even, the  $\chi_n^2$  distribution coincides with the Erlang distribution of shape parameter  $n/2$  and scale parameter 2, and (10) admits

further simplifications. In case  $n = 2$ , a Rayleigh distribution (of parameter  $\sigma = 1$ ) is obtained by taking the square root of the  $\chi_2^2$  distribution, i.e.,  $d_{\text{Mah}} \sim \text{Rayleigh}(1)$ , and a closed form solution exits for  $d$  in terms of  $P$ :  $d = \sqrt{-2 \log(1 - P)}$ .

### III. COMPARISON OF CLASSIFICATION CRITERIA

While the theory behind (10) should be known to researchers and it is properly used in many computer vision works [28][29][30][31][32], to our surprise, there is a large amount of works that do not follow such approach. This fact motivates the following analysis of the different incomplete or wrong classification criteria found in the literature.

Table I presents a list of some representative works among those using questionable classification criteria to decide whether a datum matches a multivariate normal distribution. The first four treatises listed in the table (before the double line) explicitly provide an erroneous confidence probability  $P$ , while the rest do not specify such a value (presumably  $P \geq 90\%$ ) and therefore use incomplete or inadequate criteria. Most of the references correspond to the specific area of moving object detection and take the GMM approach proposed in [13] as a starting point; this is indicated in the second column of the table. The third column shows the criteria used in each work (i.e., the shape of the decision/confidence regions): testing each dimension independently against a threshold (hyper-rectangular regions) or testing the Mahalanobis distance (7) (ellipsoidal regions). The works that are ambiguous about the criterion used are marked with a '?' entry. The fourth, fifth and sixth columns contain the following information extracted from the references: the dimension  $n$  of the normal model, the distance  $d$  considered to build the abovementioned rectangular or ellipsoidal regions, and the target confidence probability  $P$ . Observe that most cited works have dimension  $n = 2, 3$  or  $5$  since they correspond to models that use spatial coordinates and/or the color vector of a pixel as components to apply a distance criterion. The last two columns report the actual confidence value ( $\bar{P}$ ) obtained by us resulting from the parameters of the model (information in columns 3-5). If the criterion (column 3) is clear, only the corresponding column (Rect. or Ellip.) is given. Otherwise, two confidence values are provided depending on both possible disambiguations of the criterion.

Table I is further explained in section III-C, but let us discuss now the two deceiving situations reported therein: *i*) the case of choosing hyper-rectangular confidence regions instead of ellipsoidal ones (section III-A), and, *ii*) the case of choosing an ellipsoidal confidence region of incorrect size (section III-B). Both usual situations arise from misguided generalizations of the familiar unidimensional theory to multiple dimensions.

#### A. Case I: Treating each dimension independently

Consider an uncorrelated normal distribution  $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$  with diagonal covariance matrix  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ . Let us compare two classification criteria: a datum  $\mathbf{x} = (x_i) \in \mathbb{R}^n$  is classified as a match of the previous distribution if

$$\text{Criterion 1: } d_{\text{Mah}}(\mathbf{x}, \boldsymbol{\mu}) \leq d.$$



TABLE I  
REFERENCES USING INCOMPLETE OR INADEQUATE CLASSIFICATION CRITERIA.

References	Based on [13]	Criteria	$n$	$d$	$P(\%)$	$\bar{P}(\%)$	
						Rect.	Ellip.
2007 Wang et al. [11]		Ellip.	3	2.5	98		90
2009 Li & Prince [10]		Ellip.	2	2.5	99		95.6
2009 Nieto et al. [9]		Rect.	2	2.5	99	97.5	
2012 Pedro et al. [33]	✓	Rect.	3	2.5	95	96.3	
<hr/>							
2000 Stauffer & Grimson [13]	✓	?	3	2.5	-	96.3	90
2003 Hayman & Eklundh [14]	✓	Rect.	3	2 - 3	-	86.9 - 99.2	
2003 Zang & Klette [20]	✓	?	3	2.5	-	96.3	90
2004 Zivkovic [17]	✓	Ellip.	3	3	-		97.1
2005 Lee [21]	✓	?	3	3	-	99.2	97.1
2005 Jin & Mokhtarian [7]		Rect.	3	3	-	99.2	
2006 Luo et al. [8]		Rect.	5	3	-	98.7	
2009 Ming et al. [22]	✓	?	3	2.5	-	96.3	90
2009 Ying et al. [23]	✓	?	2	2.5	-	97.5	95.6
2011 Camplani & Salgado [19]	✓	Ellip.	3	2.5	-		90
2011 Suhr et al. [15]	✓	Rect.	3	2.5	-	96.3	
2012 Mirabi & Javadi [24]	✓	?	3	2.5	-	96.3	90
2012 Gallego et al. [16]	✓	Rect.	3	2.5	-	96.3	

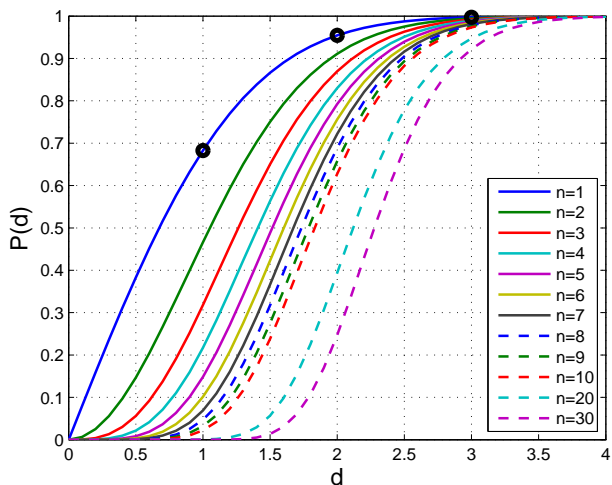


Fig. 4. Curves representing Eq. (14) for  $n = \{1, 2, \dots, 9, 10, 20, 30\}$ . The three circular markers illustrate the 68-95-99.7 rule on the curve representing Eq. (1). Note the different horizontal axis range with respect to Fig. 3.

**Criterion 2:** The magnitude of the z-score of each coordinate  $x_i$  of  $\mathbf{x}$  is smaller than  $d$ , i.e.,  $|x_i - \mu_i|/\sigma_i \leq d$  for  $i = 1, \dots, n$ , where  $\boldsymbol{\mu} = (\mu_i) \in \mathbb{R}^n$ .

First of all, criterion 2 is meaningless unless the covariance matrix  $\Sigma$  of the normal distribution is diagonal (i.e., the equidensity ellipsoids of the PDF are aligned with the coordinate axes), whereas criterion 1 is meaningful for any admissible covariance  $\Sigma$ . That is why, to compare both criteria, a diagonal  $\Sigma$  was chosen. Also observe that criterion 1 tests a mixture of the coordinates  $x_i$  against  $d$ , whereas criterion 2 does not.

Let us compare the confidence regions generated by both criteria. For criterion 1,

$$R_1(d) = \{\mathbf{x} \in \mathbb{R}^n \text{ such that } d_{\text{Mah}}(\mathbf{x}, \boldsymbol{\mu}) \leq d\}$$

defines a confidence region with ellipsoidal shape bounded by an equidensity contour since  $f_X(\partial R_1(d)) = f_0$ . For criterion 2, if we define the intervals  $I_i(d) = \{y \in \mathbb{R} \text{ such that } |y -$

$\mu_i|/\sigma_i \leq d\}$  for  $i = 1, \dots, n$ , then

$$R_2(d) = \{(x_i) = \mathbf{x} \in \mathbb{R}^n \text{ such that } x_i \in I_i(d)\}$$

is the Cartesian product of the intervals  $I_i(d)$ , i.e., a hyper-rectangle in  $\mathbb{R}^n$ , which explains the label ‘Rect.’ in Table I. By virtue of Result 1, confidence regions  $R_2$  are not optimal for normal distributions with  $n > 1$  because they are not the interior of equidensity contours. Even without Result 1, we may justify the superiority of criterion 1 over criterion 2 simply by examining the relationship between both criteria and the PDF. The value of the normal PDF (8) is a function of the single parameter  $d_{\text{Mah}}$ . Therefore, any decision that is not a function of  $d_{\text{Mah}}$  (such as criterion 2) is not taking into account the joint PDF; it is artificially dissociating the decision from the probability density, which is definitely not the desired goal.

Even assuming that criterion 2 is valid, although not optimal for normal distributions, many works in the literature set the classification threshold  $d$  not taking into account the dimensionality  $n$  of their models. Let us show how to set such a threshold and quantify the classification difference with respect to criterion 1 by computing the probabilities associated with the previous events. For criterion 1,  $P_1 := P(R_1(d))$  is given by (10). For criterion 2,

$$P_2 := P(R_2(d)) = \prod_{i=1}^n P(I_i(d)) = \left( \text{erf} \left( \frac{d}{\sqrt{2}} \right) \right)^n, \quad (14)$$

which is illustrated in Fig. 4. Obviously, if  $n = 1$ , then  $P_2(d) = P_1(d)$ . Comparing Figs. 3 and 4 for  $n > 1$ , observe that  $P_1(d) < P_2(d)$  because  $R_1(d) \subset R_2(d)$ . For  $n > 1$  and a target probability  $P_0 = P_1(d_1) = P_2(d_2)$ , we have  $d_2 < d_1$ , but Result 1 guarantees that  $\text{Vol}(R_1(d_1)) < \text{Vol}(R_2(d_2))$ , i.e., the volume of the ellipsoid of ‘radius’  $d_1$  is smaller than the volume of the hyper-rectangle of sides  $2d_2\sigma_i$ .

Fig. 5 shows the difference  $P_2 - P_1 \geq 0$  between classification criteria. It is stunning that, as the dimension  $n$  increases, the difference can be as large as possible for some values of  $d$ . Typically, based on the z-score for the

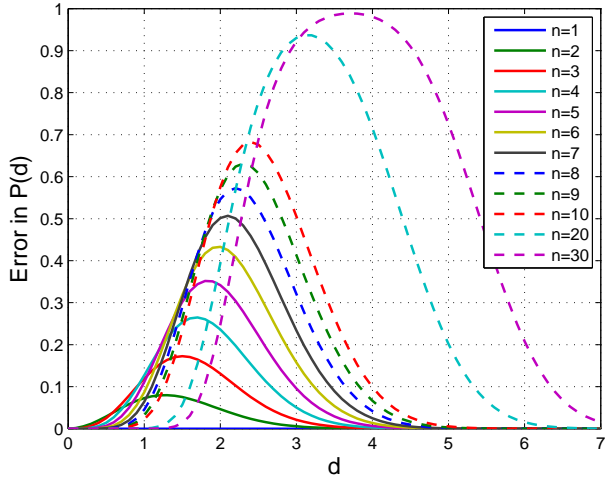


Fig. 5. *Case I.* Error curves corresponding to difference of probabilities  $P_2 - P_1$  for  $n = \{1, 2, \dots, 9, 10, 20, 30\}$ .  $P_1$  and  $P_2$  are represented in Figs. 3 and 4, respectively.

unidimensional distribution, researchers select  $d \in [2, 3]$ . For a small dimensionality such as  $n = 3$ , this yields moderate errors of 3% – 12%. However, such deviations can become dramatic as  $n$  increases, for example errors of 15% – 45% if  $n = 6$ , 50% – 68% if  $n = 10$ , or even larger, 40% – 95% if  $n = 20$ .

#### B. Case II: Setting the threshold without taking into account the dimension of the model

We concluded in previous sections that the optimal confidence region for a normal distribution depends on the Mahalanobis distance of the datum  $\mathbf{x}$  from the mean. This section analyzes methods based on this distance, so the covariance matrix does not need to be diagonal in the following discussion.

Another common error, more subtle to detect than Case I, is that of setting an incorrect threshold on the Mahalanobis distance, typically, without taking into account the dimension  $n$  of the model. Instead, values drawn from the familiar unidimensional theory (1), e.g., thresholds  $w \equiv d = \{1, 2, 3\}$  corresponding to the 68-95-99.7 rule, are used. The confidence regions are still ellipsoids, but their Mahalanobis “radius” is smaller than it should be for a target probability. The selected classification regions are over-confident [27]: one might think that they accumulate enough probability to reach a desired confidence level when in fact they do not because they are smaller than required.

Fig. 6 shows the errors caused by such a deceiving choice of classification threshold. For small dimensionality  $n$ , the errors might be small, hence difficult to detect. However, as  $n$  increases (for more complicated models), the errors can be arbitrarily large. This is definitely a situation to avoid. For instance, a threshold  $d = 2.5$  causes an error of  $\approx 10\%$  if the model has  $n = 3$ , whereas the error grows up to 40% if  $n = 6$ , and  $> 95\%$  if  $n \geq 10$ . In addition, in the typical interval  $d \in [2, 3]$  and for some values of  $n$ , there is a wild error variation. For example, errors of 3% – 23% if  $n = 3$ , 17% – 64% if  $n = 6$ , or 50% – 90% if  $n = 10$ . Therefore,

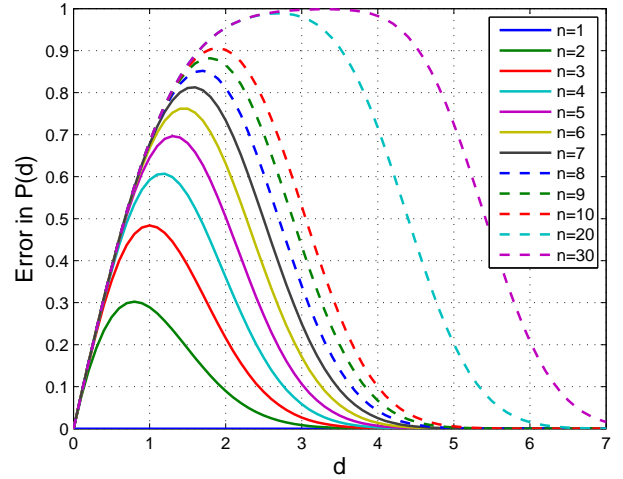


Fig. 6. *Case II.* Error curves corresponding to difference of probabilities  $P(d; 1) - P(d; n)$  for  $n = \{1, 2, \dots, 9, 10, 20, 30\}$ .  $P(d; n)$  is represented in Fig. 3.

caution is paramount for the selection of the right classification threshold depending on the dimensionality of the model  $n$ , both to avoid over-confidence (smaller threshold than required) and under-confidence (larger threshold than required).

#### C. Discussion of confidence regions used in the literature

The works in Table I admit several classifications. According to the sixth column, two categories may be distinguished: works that provide an incorrect confidence level [9][10][11][33], and works that are incomplete (marked with ‘-’) because they do not specify such confidence values. According to the third column, many works [20][21][22][23][24] are ambiguous about the classification criterion used. In most cases, this is due to the ambiguity inherited from [13]. If the ambiguity is resolved, two interpretations are possible: some treatises [14][16][33] opt for hyper-rectangular regions, while others [17][19] choose ellipsoidal ones. In other references it is not possible to infer the criteria used, and the choice between both criteria causes confidence variations, e.g., a 6% variation for the  $n = 3$ ,  $d = 2.5$  case [13][20][22][24]. The last two columns of Table I are given taking into account both possible disambiguations.

Within classification scenarios, the confidence level that determines the decision regions is usually designed based on a target false detection rate (type I error) of the system, and it therefore also affects the misdetection rate (type II error). Specifically, in moving object detection applications through background modeling ([13] and subsequent works), the two situations that differ from a nominal confidence design are as follows. On the one hand, under-confident decision regions can cause cropped detections and the miss of moving objects with similar features to those of the background. On the other hand, over-confident regions improve the detection rate but they also increase the false alarm rate due to a very restrictive background classification.

In [13] and related works, the models are described for arbitrary dimension  $n$ , but they are usually exemplified with  $n = 3$  (color components), and, as we have analyzed, the

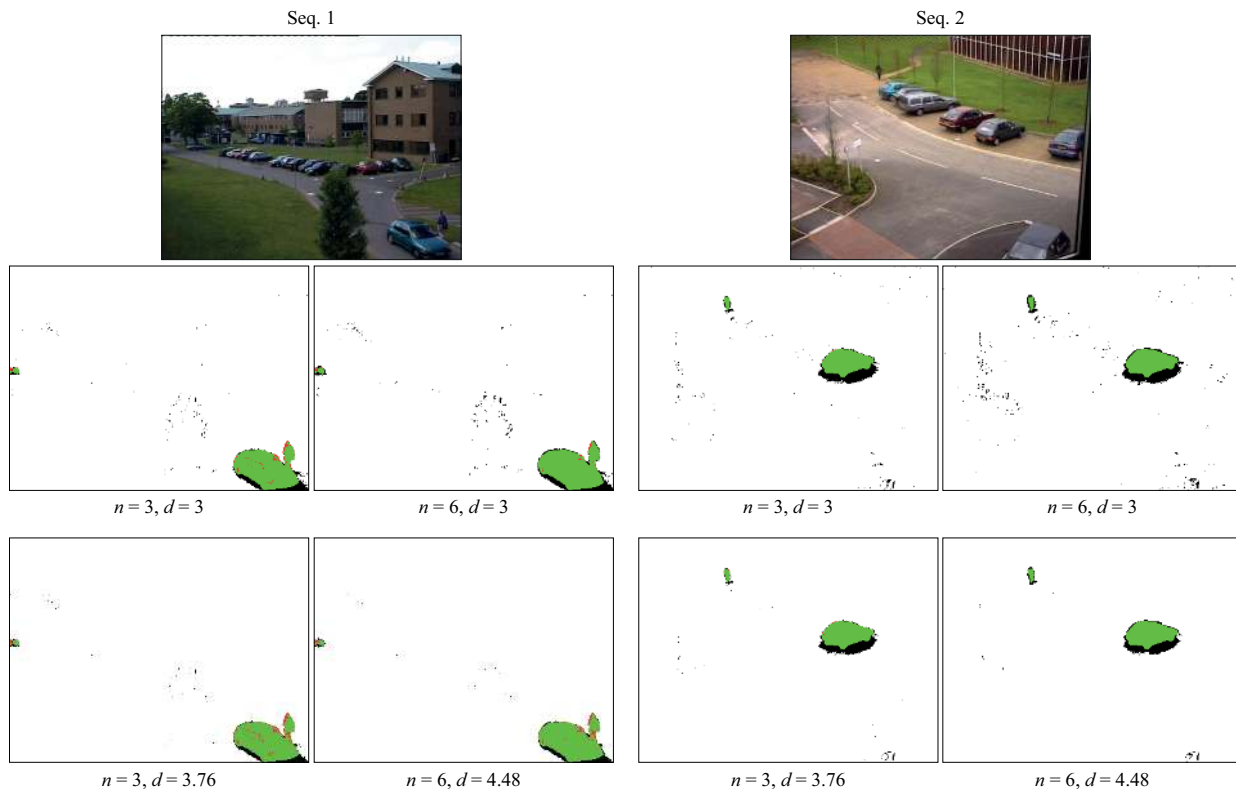


Fig. 7. Results obtained from the application of the classical GMM method to detect moving objects in two different outdoor sequences. Color notation: correct detections (green), misdetections (red), and false detections (black).

TABLE II  
MAHALANOBIS DISTANCE  $d(P)$  OF GENERIC  $n$ -DIMENSIONAL NORMAL DISTRIBUTIONS TO OBTAIN TYPICAL CONFIDENCE LEVELS.

Dimension $n$	Confidence level $P$		
	$\approx 68\%$	$\approx 95\%$	$\approx 99.7\%$
1	1	2	3
2	1.52	2.49	3.44
3	1.88	2.83	3.76
4	2.17	3.12	4.03
5	2.43	3.36	4.27
6	2.65	3.58	4.48
7	2.86	3.79	4.67
8	3.05	3.97	4.86
9	3.23	4.15	5.03
10	3.40	4.31	5.19

deviations in confidence values are moderate for such small dimensionality (cf. the last columns of Table I,  $P$  vs  $\bar{P}$ ). Nevertheless, it is expected that future works consider higher dimensional models such as [18] ( $n = 6$ ) and [34] ( $n = 10$ ), and, since many of the references listed in Table I set the decision regions without taking into account  $n$ , they might lead to severe errors if the dimension used is large.

As a final recommendation, Table II summarizes, based on Fig. 3, the correct Mahalanobis distance thresholds for general  $n$ -dimensional normal distributions (from  $n = 1$  to  $n = 10$ ) to obtain the accumulated probabilities considered as a reference in the one-dimensional case. It can be observed that the required threshold increases with the dimension for each fixed probability.

TABLE III  
COMPARISON OF DIFFERENT CLASSIFICATION THRESHOLDS FOR BACKGROUND MODELING: QUANTITATIVE RESULTS

$n$	$d$	Seq. 1			Seq. 2		
		Recall	Precision	$F$ -score	Recall	Precision	$F$ -score
3	3	88.37	66.63	75.98	94.55	59.49	73.03
	3.76	83.40	75.47	79.24	91.69	66.49	77.08
6	3	95.07	54.31	69.12	98.80	49.49	65.95
	4.48	88.25	73.84	80.40	96.05	65.14	77.63

#### D. Numerical experiments

To evaluate the disadvantages resulting from the use of inadequate classification criteria in real applications, we have analyzed their effect on the quality of the results provided by the popular GMM-based moving object detection strategy [13]. Since the RGB color space is probably the most commonly used in the literature [19][20][22], initially, this analysis has been carried out using it ( $n = 3$ ). Additionally, a combination of RGB components and their gradients ( $n = 6$ ) have been used, as proposed by several authors [34][35] to improve classification if background and foreground have similar colors.

All tests performed use ellipsoidal decision regions, but two different distance thresholds are compared: one borrowed from the unidimensional case ( $d = 3$ , used by many authors and corresponding to a 99.7% confidence level), and the correct one taking into account the dimensionality  $n$  of the model (for the same confidence level). The quality of the detections



provided by the analyzed method is measured by means of the conventional recall, precision and  $F$ -score evaluation parameters [36].

Figure 7 illustrates some of the results obtained for two outdoor sequences from the PETS database [37], and Table III reports the corresponding quality measurements. It can be observed that the background variations are better captured using the adequate classification thresholds ( $d = 3.76$  if  $n = 3$  or  $d = 4.48$  if  $n = 6$ ) than in case of the fixed threshold taken from the unidimensional case ( $d = 3$ , over-confident region): the amount of false detections is significantly reduced in all the analyzed cases, and consequently, the overall quality of the detections (given by the  $F$ -score) is clearly improved. This quality improvement is more noticeable as model dimensionality increases.

#### IV. CONCLUSION

In this work, we have provided a detailed comparative analysis between the correct confidence regions corresponding to general multidimensional normal distributions and certain incomplete and/or erroneous but frequent practices observed in the literature.

To this end, we have studied general  $n$ -dimensional probability distributions and have proved that, for a broad class of them, the optimal confidence region accumulating a target confidence level is the interior of an equidensity contour of the PDF. For multivariate normal distributions, the Mahalanobis distance provides the right dimensionless quantity (fully characterizing the equidensity contours of the PDF) for decision making, i.e., for establishing confidence regions. Hence, decisions based on individual  $z$ -scores (one per dimension), which lead to hyper-rectangular confidence regions, are discouraged.

We have derived the formula that relates the size of an optimal confidence region (specified by its largest Mahalanobis distance) and its cumulated probability. Concisely, the squared Mahalanobis distance follows a Chi-squared distribution with  $n$  degrees of freedom. We used this formula to quantify the errors harvested by many works in the literature, mostly caused by a wrong extrapolation of unidimensional results without taking into account the dimensionality of the model. Using several plots, we showed that such errors can be significantly large depending on the dimension  $n$ . In particular, we discussed the deviations in the typical threshold range  $d \in [2, 3]$ .

To prevent errors from propagating to future works, we have specified the correct threshold values that describe the optimal confidence regions for some common target confidence levels. Finally, we have demonstrated the theoretical analysis with a concrete example in the context of moving object detection.

#### REFERENCES

- [1] P. Larrañaga, H. Karshenas, C. Bielza, and R. Santana, "A review on probabilistic graphical models in evolutionary computation," *Journal of Heuristics*, pp. 1–25, 2012.
- [2] E. Chen, O. Haik, and Y. Yitzhaky, "Classification of moving objects in atmospherically degraded video," *Optical Engineering*, vol. 51, no. 10, pp. 101 710–1, 2012.
- [3] M. Kristan, A. Leonardis, and D. Skočaj, "Multivariate online kernel density estimation with Gaussian kernels," *Pattern Recognition*, vol. 44, no. 10, pp. 2630–2642, 2011.
- [4] A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.
- [5] C. Cuevas, D. Berjón, F. Morán, and N. García, "Moving object detection for real-time augmented reality applications in a GPGPU," *IEEE Trans. Consumer Electronics*, vol. 58, no. 1, pp. 117–125, 2012.
- [6] T. Huang, X. Fang, J. Qiu, and T. Ikenaga, "Adaptively adjusted Gaussian mixture models for surveillance applications," *Advances in Multimedia Modeling*, pp. 689–694, 2010.
- [7] Y. Jin and F. Mokhtarian, "Towards robust head tracking by particles," in *IEEE Int. Conf. Image Processing*, vol. 3, 2005, pp. 864–867.
- [8] X. Luo, S. Bhandarkar, W. Hua, and H. Gu, "Nonparametric background modeling using the CONDENSATION algorithm," in *IEEE Int. Conf. Video and Signal Based Surveillance*, 2006, pp. 1–6.
- [9] M. Nieto, C. Cuevas, and L. Salgado, "Measurement-based reclustering for multiple object tracking with particle filters," in *IEEE Int. Conf. Image Processing*, 2009, pp. 4097–4100.
- [10] P. Li and S. Prince, "Joint and implicit registration for face recognition," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2009, pp. 1510–1517.
- [11] H. Wang, D. Suter, K. Schindler, and C. Shen, "Adaptive object tracking based on an effective appearance filter," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1661–1667, 2007.
- [12] H. Hassanpour, M. Sedighi, and A. Manashty, "Video frame's background modeling: Reviewing the techniques," *Journal of Signal and Information Processing*, vol. 2, no. 2, pp. 72–78, 2011.
- [13] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.
- [14] E. Hayman and J.-O. Eklundh, "Statistical background subtraction for a mobile observer," in *Int. Conf. Computer Vision*, 2003, pp. 67–74 vol.1.
- [15] J. Suhr, H. Jung, G. Li, and J. Kim, "Mixture of Gaussians-based background subtraction for bayer-pattern image sequences," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 21, no. 3, pp. 365–370, 2011.
- [16] J. Gallego, M. Pardás, and G. Haro, "Enhanced foreground segmentation and tracking combining Bayesian background, shadow and foreground modeling," *Pattern Recognition Letters*, vol. 33, no. 12, pp. 1558–1568, 2012.
- [17] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *IEEE Int. Conf. Pattern Recognition*, vol. 2, 2004, pp. 28–31.
- [18] Q. Wan and Y. Wang, "Background subtraction based on adaptive non-parametric model," in *World Congress on Intelligent Control and Automation*, 2008, pp. 5960–5965.
- [19] M. Camplani and L. Salgado, "Adaptive background modeling in multicamera system for real-time object detection," *Optical Engineering*, vol. 50, no. 12, pp. 127 206:1–17, 2011.
- [20] Q. Zang and R. Klette, "Evaluation of an adaptive composite Gaussian model in video surveillance," in *Computer Analysis of Images and Patterns*. Springer, 2003, pp. 165–172.
- [21] D. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 827–832, 2005.
- [22] Y. Ming, C. Guodong, and Q. Lichao, "Player detection algorithm based on Gaussian mixture models background modeling," in *IEEE Int. Conf. Intelligent Networks and Intelligent Systems*, 2009, pp. 323–326.
- [23] L. Ying-hong, X. Chang-zhen, Y. Yi-xin, and L. Ya-li, "Moving object detection based on edged mixture Gaussian models," in *IEEE Int. Workshop on Intelligent Systems and Applications*, 2009, pp. 1–5.
- [24] M. Mirabi and S. Javadi, "People tracking in outdoor environment using Kalman filter," in *IEEE Int. Conf. Intelligent Systems, Modelling and Simulation*, 2012, pp. 303–307.
- [25] A. Papoulis and U. S. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2002.
- [26] A. C. Rencher, *Methods of Multivariate Analysis*, ser. Wiley Series in Probability and Statistics. Wiley, 2003.
- [27] R. Blanc, E. Syrkinina, and G. Székely, "Estimating the Confidence of Statistical Model Based Shape Prediction," in *Information Processing in Medical Imaging*. Springer, 2009, vol. 5636, LNCS, pp. 602–613.
- [28] C. Schmid and R. Mohr, "Combining greyvalue invariants with local constraints for object recognition," in *IEEE Int. Conf. Computer Vision and Pattern Recognition*, 1996, pp. 872–877.
- [29] K. Arras and S. Vestli, "Hybrid, high-precision localisation for the mail distributing mobile robot system MOPS," in *IEEE Int. Conf. Robotics and Automation*, vol. 4, 1998, pp. 3129–3134.

- [30] M. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Trans. Robotics and Automation*, vol. 17, no. 3, pp. 229–241, 2001.
- [31] Z. Guo, G. Jiang, H. Chen, and K. Yoshihira, "Tracking probabilistic correlation of monitoring data for fault detection in complex systems," in *IEEE Int. Conf. Dependable Systems and Networks*, 2006, pp. 259–268.
- [32] D. Won, H. Oh, S. Huh, D. Shim, and M. Tahk, "Multiple UAVs tracking algorithm with a multi-camera system," in *IEEE Int. Conf. Control Automation and Systems (ICCAS)*, 2010, pp. 2357–2360.
- [33] L. Pedro, G. A. Caurin, V. Belini, R. Pechoneri, A. Gonzaga, I. Neto, F. Nazareno, and M. Stücheli, "Hand gesture recognition for robot hand teleoperation," in *ABCMS Symposium Series in Mechatronics*, vol. 5, 2012, pp. 1065–1074.
- [34] C. Cuevas and N. García, "Tracking-based non-parametric background-foreground classification in a chromaticity-gradient space," in *IEEE Int. Conf. Image Processing*, 2010, pp. 845–848.
- [35] S. Atev, O. Masoud, and N. Papanikolopoulos, "Practical mixtures of gaussians with brightness monitoring," in *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, IEEE, 2004, pp. 423–428.
- [36] F.-C. Cheng and S.-J. Ruan, "Accurate motion detection using a self-adaptive background matching framework," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, no. 2, pp. 671–679, 2012.
- [37] Computational Vision Group, "PETS: Performance Evaluation of Tracking and Surveillance," <http://www.cvg.rdg.ac.uk/>, University of Reading.



**Guillermo Gallego** received the Ingeniero de Telecomunicación degree (five years engineering program) from the Universidad Politécnica de Madrid, Spain, in 2004, M.S. degree in mathematical engineering (Magíster en Ingeniería Matemática) from the Universidad Complutense de Madrid, Madrid, in 2005, and the M.S. in electrical and computer engineering, M.S. in mathematics, and Ph.D. in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, in 2007, 2009 and 2011, respectively.

He was a recipient of the Fulbright Scholarship to pursue graduate studies at the Georgia Institute of Technology in 2005. Since 2011, he has been a Marie-Curie COFUND post-doctoral researcher with the Universidad Politécnica de Madrid. His research interests fall within the areas of signal processing, geometry, optimization, computer vision and ocean engineering.



**Carlos Cuevas** received the Ingeniero de Telecomunicación degree (integrated BSc-MSc accredited by ABET) in 2006 and the Doctor Ingeniero de Telecomunicación degree (Ph.D. in Communications), in 2011, both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain.

Since 2006 he has been a member of the Grupo de Tratamiento de Imágenes (Image Processing Group) of the UPM. His research interests include signal and image processing, computer vision, pattern recognition and automatic target recognition.



**Raúl Mohedano** received the Ingeniero de Telecomunicación degree (five years engineering program) in 2006 from the Universidad Politécnica de Madrid (UPM), Madrid, Spain.

Since 2007 he has been a member of the Grupo de Tratamiento de Imágenes (Image Processing Group) of the UPM, where he holds a personal research grant from the Comunidad de Madrid. His research interests are in the area of computer vision.



**Narciso García** received the Ingeniero de Telecomunicación degree (five years engineering program) in 1976 (Spanish National Graduation Award) and the Doctor Ingeniero de Telecomunicación degree (Ph.D. in Communications) in 1983 (Doctoral Graduation Award), both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain.

Since 1977 he is a faculty member with the UPM where he is currently a Professor of Signal Theory and Communications. He leads the Grupo de Tratamiento de Imágenes (Image Processing Group) of the UPM. He has been actively involved in Spanish and European research projects, serving also as evaluator, reviewer, auditor, and observer of several research and development programmes of the European Union. He was a co-writer of the EBU proposal, basis of the ITU standard for digital transmission of TV at 34–45 Mb/s (ITU-T J.81). His professional and research interests are in the areas of digital image processing, video compression and computer vision.