

Rozenn Dahyot · Pierre Charbonnier · Fabrice Heitz

## A Bayesian approach to object detection using probabilistic appearance-based models

Received: 18 July 2003 / Accepted: 17 September 2004 / Published online: 5 November 2004  
© Springer-Verlag London Limited 2004

**Abstract** In this paper, we introduce a Bayesian approach, inspired by probabilistic principal component analysis (PPCA) (Tipping and Bishop in *J Royal Stat Soc Ser B* 61(3):611–622, 1999), to detect objects in complex scenes using appearance-based models. The originality of the proposed framework is to explicitly take into account general forms of the underlying distributions, both for the in-eigenspace distribution and for the observation model. The approach combines linear data reduction techniques (to preserve computational efficiency), non-linear constraints on the in-eigenspace distribution (to model complex variabilities) and non-linear (robust) observation models (to cope with clutter, outliers and occlusions). The resulting statistical representation generalises most existing PCA-based models (Tipping and Bishop in *J Royal Stat Soc Ser B* 61(3):611–622, 1999; Black and Jepson in *Int J Comput Vis* 26(1):63–84, 1998; Moghaddam and Pentland in *IEEE Trans Pattern Anal Machine Intell* 19(7):696–710, 1997) and leads to the definition of a new family of non-linear probabilistic detectors. The performance of the approach is assessed using receiver operating characteristic (ROC) analysis on several representative databases, showing a major improvement in detection performances with respect to the standard methods that have been the references up to now.

**Keywords** Eigenspace representation · Probabilistic PCA · Bayesian approach · Non-Gaussian models · M-estimators · Half-quadratic algorithms

### 1 Introduction

A reliable detection of objects in complex scenes exhibiting non-Gaussian noise, clutter and occlusions remains an open issue in many instances. Since the early 1990s, *appearance-based* representations have met unquestionable success in this field [4, 5]. Global appearance models represent objects using raw 2D brightness images (intensity surfaces), without any feature extraction or construction of complex 3D models. Appearance models have the ability to efficiently encode shape, pose and illumination in a single, compact representation, using data reduction techniques. The early success of global appearance models, in particular, in face recognition [5], has given rise to a very active research field, which has resulted in the ability of recognising 3D objects in databases of more than 100 objects [4] or, more recently, the recognition, with good reliability, of comprehensive object classes (cars, faces) [6] in complex, unstructured environments.

Linear, as well as non-linear models, associated to various data reduction techniques have been proposed to obtain parsimonious representations of large image databases. Standard linear techniques include principal component analysis (PCA) and independent component analysis (ICA). Non-linear extensions have been proposed such as non-linear-PCA, principal curves and surfaces, non-linear manifolds, kernel-based methods and neural networks [7–9]. In this paper, we are interested in a particular class of global appearance models, namely, probabilistic appearance models, which were introduced by Moghaddam and Pentland [3, 10] in 1995. These models offer several advantages:

This revised version was published online in November 2004 with corrections to the section numbers.

R. Dahyot · P. Charbonnier (✉)  
ERA 27 LCPC, Laboratoire des Ponts et Chaussées,  
11 rue Jean Mentelin, B.P. 9, 67035 Strasbourg, France  
E-mail: Pierre.Charbonnier@equipement.gouv.fr  
E-mail: dahyot@mee.tcd.ie

F. Heitz  
LSIIT UMR CNRS 7005,  
Université Louis Pasteur—Pôle API,  
Bd Sébastien Brant, 67400 Illkirch, France  
E-mail: heitz@lsiit.u-strasbg.fr

*Present address:* R. Dahyot  
Department of Statistics, Trinity College,  
Dublin, 2, Ireland

- They are probabilistic: they make it possible to represent a *class* of images and make available all the traditional methods of statistical estimation (maximum likelihood, Bayesian approaches)
- They are linear and, thus, are suited to efficient implementation [11]
- Although linear, they have outperformed, in terms of detection and recognition, not only the traditional linear approaches (PCA, ICA), but also non-linear approaches (such as neural networks or non-linear kernel PCA), in a recent comparison carried out by Moghaddam [8]

In the Bayesian approach proposed in the present paper, linear (i.e. PCA-based) data reduction techniques are associated to non-linear noise models and non-Gaussian prior models to derive robust and efficient image detectors. The proposed framework unifies different PCA-based models previously proposed in the literature [2, 3]. Our approach straightforwardly integrates non-linear statistical constraints on the distribution of the images in the eigenspace. We show experimentally the importance of an appropriate model for this distribution and its impact on the performance of the detection process. Moreover, the approach enables, when necessary, to introduce robust hypotheses on the distribution of noise, allowing to cope with clutter, *outliers* and occlusions. This leads to the definition of a novel family of general-purpose detectors that experimentally outperform the existing PCA-based methods [2, 3].

The paper is organised as follows. Section 2 briefly reviews existing PCA-based detection methods. Section 3 describes the different constituents of the proposed Bayesian approach: eigenspace representation, non-linear noise models and non-Gaussian priors. Detection algorithms and implementations are detailed in Sect. 4. Section 5 presents a comparison between the proposed Bayesian detector and several state-of-the-art approaches. Three databases have been used to illustrate the contributions of the various components of the model. An objective assessment is proposed using receiver operating characteristic (ROC) analysis, showing the benefits of the approach.

---

## 2 PCA-based statistical detection

Detection, classification and recognition algorithms that use PCA-based subspace representations [3, 5] first relied on the computation of simple, Euclidean distances between the observed image and the training samples. The quadratic distance to the centre of the training class (sum of squared differences (SSD)) or the orthogonal distance to the eigenspace (distance from feature space (DFFS)) have, thus, first been used [3, 5]. None of these distances, however, is satisfying: the first one assigns the same distance to all images belonging to a hyper-sphere, while the second gives the same measure for all obser-

vations distributed on spaces which are parallel to the eigenspace. It is, therefore, easy to generate examples that would make these methods fail. A significant improvement has been obtained by recasting the problem in a probabilistic framework [1, 3, 10, 12]. Moghaddam and Pentland [10] proposed a statistical formulation of PCA based on multivariate Gaussian (or mixture-of-Gaussians) distributions. The resulting probabilistic model embeds distance information both in the eigenspace and in its orthogonal. Experimental results by Moghaddam and Pentland [3] and Moghaddam [8] have shown the major contribution of this approach, not only by comparison with SSD and DFFS, but also comparatively to methods based on non-linear representations, such as non-linear PCA or kernel PCA [8]. Tipping and Bishop [1, 12] and Roweis [13] have recently proposed, in independent but similar works, other probabilistic interpretations of PCA, probabilistic PCA and sensible PCA, respectively. These rely on a latent variable model which, assuming Gaussian distributions, yields the same representation as Moghaddam and Pentland [3, 10].

Most of the time in eigenspace methods, the noise distributions have been considered as Gaussian. As is well known, such a hypothesis is seldom verified in practice. Therefore, standard detection and recognition methods based on Gaussian noise models are sensitive to gross errors or *outliers* stemming, for instance, from partial occlusions or clutter. M-estimators, introduced by Huber [14] in robust statistical estimation, are forgiving about such artifacts. They have, in particular, been used to develop PCA-based robust recognition methods [2]. More recently, an alternative to M-estimation, based on random sampling and hypotheses testing, was proposed to address the problem of robust recognition [15].

Another important limitation of standard methods concerns the a priori modelling of the distribution of the learning images in the eigenspace. In standard PCA-based approaches, these densities are generally considered as Gaussian [1–3] or uniform, although they are often non-Gaussian (see Sect. 3.5 and [4]). This strongly biases the detection towards the mean image. Thus, modelling complex *in-eigenspace* distributions remains a key issue for the practical application of eigenspace methods. The first approach proposed to address this problem in the field of visual recognition was described by Murase and Nayar [4]. Murase and Nayar have used an ad hoc B-spline representation of the non-linear manifold corresponding to the distribution of training images in the eigenspace. Locally linear embedding [9], mixtures of Gaussians [3, 12] and other more computationally involved non-linear models [8] have also been considered more recently. Non-linear generalisations of PCA have been developed using auto-associative neural networks [16] or self-organising maps [17]. Neural networks, however, are prone to over-fitting and require the definition of a proper architecture and learning scheme. The notion of a non-linear, low-dimensional manifold

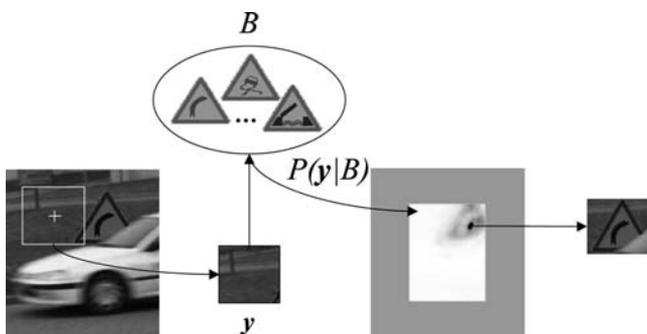
passing “through the middle of the data set” was formalised in [18] for two dimensions. Extensions to a larger dimension (which is far from being trivial) have been recently proposed, such as non-linear PCA [19] or probabilistic principal surfaces [20], but their implementation remains involved. Another approach that has become popular in the late 1990s is kernel PCA [21], where an implicit non-linear mapping is applied to the input by means of a Mercer kernel function, before a linear PCA is applied in the mapped feature space. The approach is appealing because its implementation is simpler, since it uses only linear algebra, but the optimal choice of the kernel function is still an open issue.

In this paper, we propose an alternative to these non-linear representations that preserves the linearity of the underlying latent variable model (thus, preserving computational simplicity). The proposed Bayesian framework generalises most PCA-based models previously proposed in the literature. Our approach combines linear data reduction techniques (to preserve computational efficiency), non-Gaussian models on the in-eigenspace distribution (to represent complex variabilities) and robust hypotheses on the distribution of noise (to cope with clutter, outliers and occlusions). The proposed representation is described in the next section.

### 3 Detection: a Bayesian approach

#### 3.1 Principle of detection

Figure 1 illustrates the general principle of the detection method [3]. The image is analysed in a raster scan manner: at each position  $(i, j)$ , an observation vector,  $\mathbf{y}$ , is extracted from a window. The localisation of the modelled pattern is obtained by computing, for each position  $(i, j)$  of the window, the likelihood  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  of the observation vector, according to a learned model,  $\mathcal{B}$ . The likelihood computed for the window centred at  $(i, j)$  is stored in a likelihood map at the same location  $(i, j)$ . When the scene has been completely scanned, a simple



**Fig. 1** Detection process: extraction of the observation vector at each pixel location (*left*). Computation of the log-likelihood (*centre*). Thresholding of the log-likelihood map to locate the object (*right*)

thresholding of the likelihood map is used to determine whether or not one or several objects are present in the scene and to find out their location.

#### 3.2 A Bayesian framework for the detection

The observation vector  $\mathbf{y}$  can be decomposed using two independent random vectors as follows:

$$\mathbf{y} = f(\mathbf{c}) + \mathbf{w} \quad (1)$$

We consider here that the relation  $f$  between the observation  $\mathbf{y}$  and a latent vector  $\mathbf{c}$  is known (see Sect. 3.3). We assume that  $\mathbf{c}$  captures most of the information of the class  $\mathcal{B}$ . Vector  $\mathbf{w}$  represents the (modelling, observation) noise on the reconstruction  $f(\mathbf{c})$ .

*Likelihood* The likelihood of the observation  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  is computed by integrating the joint distribution of  $(\mathbf{c}, \mathbf{y})$  with respect to  $\mathbf{c}$ :

$$\begin{aligned} \mathcal{P}(\mathbf{y}|\mathcal{B}) &= \int \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})d\mathbf{c} \\ &= \int \mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B})\mathcal{P}(\mathbf{c}|\mathcal{B})d\mathbf{c} \end{aligned} \quad (2)$$

In this expression, the first term  $\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B}) = \mathcal{P}(\mathbf{w}|\mathcal{B})$  is the noise distribution. The second term  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  is the prior distribution of the latent variables (distribution in the eigenspace). For computational efficiency, it is desirable to have a simple analytical expression of the likelihood (Eq. 2). Unfortunately, except for some particular cases, e.g. Gaussian noise and Gaussian a priori [1], the analytical expression of  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  is generally not tractable.

*Approximated likelihood* Another solution, which we adopt in this paper, consists of approximating the distribution  $\mathcal{P}(\mathbf{y}|\mathcal{B})$ . Several approximations have been proposed in the Bayesian framework to compute such distributions [22, 23]. The approximation we use, proposed in [23], is based on the hypothesis that  $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$  peaks sharply where the latent variables  $\mathbf{c}$  are the most probable:

$$\hat{\mathbf{c}} = \arg \max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$$

The likelihood of the observation  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  may then be approximated by the height of the peak multiplied by its “width”  $\sigma_{\text{peak}}$ :

$$\begin{aligned} \mathcal{P}(\mathbf{y}|\mathcal{B}) &\cong \sigma_{\text{peak}} \max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) \\ &\propto \mathcal{P}(\mathbf{y}|\hat{\mathbf{c}}, \mathcal{B})\mathcal{P}(\hat{\mathbf{c}}|\mathcal{B}) \end{aligned} \quad (3)$$

This approximation, which confuses the distribution and its mode, is linked to other standard methods used in Bayesian inference, such as Laplace’s method [22] (see also [7], p. 92). It is usually a good approximation, in particular, in the Gaussian case, and leads to good detection results, even in non-Gaussian cases. It is, moreover, justified a posteriori by our experimental results.

*Computational complexity* Assuming the relation shown in Eq. 1, then Eq. 2 or Eq. 3 provide a generic way to compute the likelihood of the observations  $\mathbf{y}$  for the class of interest  $\mathcal{B}$ . Whatever the function  $f$  and the assumptions for  $\mathcal{P}(\mathbf{w}|\mathcal{B})$  and  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ , computing  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  with Eq. 2 or Eq. 3 can be performed using simulation methods [22]. However, as explained in Sect. 3.1, the likelihood  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  has to be computed for each observation window extracted from the image. Due to their computation costs, those simulation methods are, therefore, ill-suited in practice.

In order to show the potential of our approach, we limit hereafter the application of this method as follows. First, the relation  $f$  in Eq. 1 is chosen to be linear (see Sect. 3.3). Second, the distributions of the noise  $\mathbf{w}$  and the informative vector  $\mathbf{c}$  are limited to several hypotheses that we present in detail in Sects. 3.4 and 3.5, respectively. These restrictions allow us to define efficient algorithms for the computation of the likelihood  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  in Sect. 4.

### 3.3 Linear projection model $f$

#### 3.3.1 Eigenspace decomposition

In a preliminary (training) phase, a representative set (database)  $\mathcal{B}$  of grey-scale images  $\mathbf{x}_k$ ,  $k=1, \dots, K$ , of dimension  $N$  pixels, is collected by selecting views corresponding to the different appearances of the objects to be modelled. Data are collected in vector form by considering the lexicographic ordering of the picture elements. The sample mean  $\boldsymbol{\mu}$  of the training set is defined as:

$$\boldsymbol{\mu} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \quad (4)$$

The covariance matrix of the training database is estimated by:

$$\Sigma = \frac{1}{K} \sum_{k=1}^K (\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^T \quad (5)$$

It is symmetric, positive semi-definite and may be diagonalised:  $\Sigma = \mathbf{U}_N \Lambda_N \mathbf{U}_N^T$ . In this expression,  $\mathbf{U}_N$  is the matrix collecting the  $N$  orthonormal eigenvectors of  $\Sigma$ .  $\Lambda_N$  is the diagonal matrix of the corresponding  $N$  eigenvalues. Each training sample  $\mathbf{x}$  can, thus, be written as the sum of the sample mean  $\boldsymbol{\mu}$  and a linear combination of  $J$  eigenvectors (ordered as columns in matrix  $\mathbf{U}$ ), with a reconstruction error  $\mathbf{w}^r$ :

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{U}\mathbf{c} + \mathbf{w}^r = \boldsymbol{\mu} + \sum_{j=1}^J c_j \mathbf{u}_j + \mathbf{w}^r \quad (6)$$

The eigenvectors  $\mathbf{u}_j$  associated to the  $J$  largest eigenvalues are selected in Eq. 6, yielding the standard PCA

representation. Equation 6 is the basis of the *non-standard* statistical model we develop in the next section.

#### 3.3.2 Observation model

Our observation model corresponds to a non-standard statistical interpretation of PCA, inspired by the latent variable representation proposed by Tipping and Bishop in the Gaussian case [1]. More precisely, we consider that the observation  $\mathbf{y}$  can be reconstructed from the eigenvectors as follows:

$$\mathbf{y} = \boldsymbol{\mu} + \sum_{j=1}^J c_j \mathbf{u}_j + \mathbf{w}^r \quad (7)$$

Here,  $\mathbf{w}$  is the sum of the classical reconstruction error  $\mathbf{w}^r$  (due to the truncation of the representation to  $J$  eigenvectors) and of an observation noise  $\mathbf{w}^o$  produced by the recording system, possible occlusions or textured background (clutter). In standard probabilistic PCA [1, 3], the latent variables  $c_j$  are uncorrelated and follow a Gaussian prior. The same holds for the noise term  $\mathbf{w}$ . In the non-standard model proposed here, we relax these assumptions by allowing non-Gaussian prior models as well as non-Gaussian (robust) noise models. The proposed representation, thus, no longer corresponds to standard probabilistic PCA and, in particular, the parameters of the model (corresponding to the eigenvectors  $\mathbf{u}_j$ ) are no longer maximum likelihood estimates, as in [1]. On the other hand, the benefit of the proposed approach is its ability to better represent the complex distributions that may occur in real cases.

### 3.4 Noise models $\mathcal{P}(\mathbf{w}|\mathcal{B})$

The classical noise distribution model is the Gaussian model with a diagonal covariance matrix [1]:

$$\begin{aligned} \mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B}) &\propto \exp \left[ -\frac{\|\mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\mathbf{c}\|^2}{2\sigma_g^2} \right] \\ &= \exp \left[ -\frac{1}{2} \sum_{n=1}^N \left( \frac{w_n}{\sigma_g} \right)^2 \right] \end{aligned} \quad (8)$$

The Gaussian hypothesis is not satisfactory when observations are corrupted by non-linear artifacts, occlusions or clutter. In these cases, large residual values  $w_n$  (i.e. outliers) are generated, which are highly improbable under the Gaussian assumption. In order to take into account the possible occurrence of outliers, Black and Jepson [2] proposed the use of robust, M-estimators in eigenspace models [14]. These models take into account outliers by replacing the Gaussian distribution with:

$$\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B}) \propto \exp \left[ -\frac{1}{2} \sum_{n=1}^N \rho \left( \frac{w_n}{\sigma_g} \right)^2 \right] \quad (9)$$

where  $\rho$  is a non-quadratic function, which may, moreover, be non-convex [2, 14, 24].

### 3.5 Prior models in eigenspace $\mathcal{P}(\mathbf{c}|\mathcal{B})$

*Uniform distribution* The simplest model considers a uniform distribution of the latent variables in eigenspace:

$$\mathcal{P}(\mathbf{c}|\mathcal{B}) = \text{constant}$$

This corresponds to the absence of prior knowledge on the eigenspace distribution.

*Gaussian distribution* Another standard hypothesis consists of assuming that the distribution of the latent variables is Gaussian:

$$\mathcal{P}(\mathbf{c}|\mathcal{B}) = \frac{\exp\left[-\frac{1}{2}\mathbf{c}^T\Lambda^{-1}\mathbf{c}\right]}{\prod_{j=1}^J\sqrt{2\pi\lambda_j}}$$

where the  $J$  variances  $\lambda_j$  in the diagonal matrix  $\Lambda$  are the eigenvalues computed during the training phase. This is equivalent to associating a Gaussian model with PCA, as in [1, 3].

*Other prior distributions* Although classical, the Gaussian assumption may be inappropriate for modeling real distributions. As an illustration, we consider three examples of training databases. The first database is an excerpt from the Columbia Object Image Library (COIL) database, the other two contain European road sign images (used in our application). Figure 2 shows a few sample images from the COIL data set, which is made of 72 grey-scale images of the object “duck” [25] from different viewing angles.

The second database considers a single object rotating in the image plane (one image every  $2^\circ$ , see Fig. 3). Since the object under concern is the mean image of the white triangular European road signs, the database is called AVG.

The third database is composed of colour images of 43 (yellow and white) triangular road signs, rotating in the image plane (one image every  $10^\circ$ , see Fig. 4), and is called A43.

Figure 5 presents the distribution of the training images projected onto a 3D eigenspace for the COIL and A43 databases. We can notice that the distributions

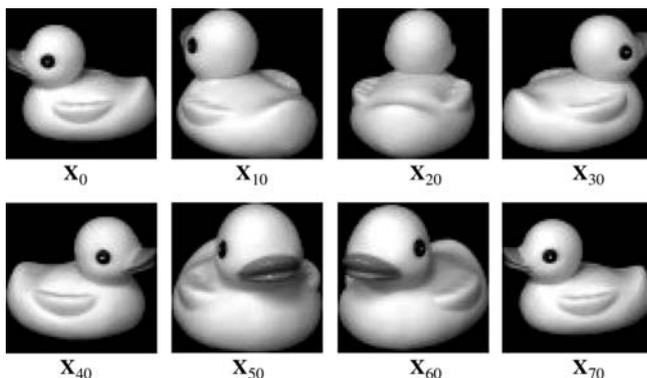


Fig. 2 Sample training images from the COIL database [25]

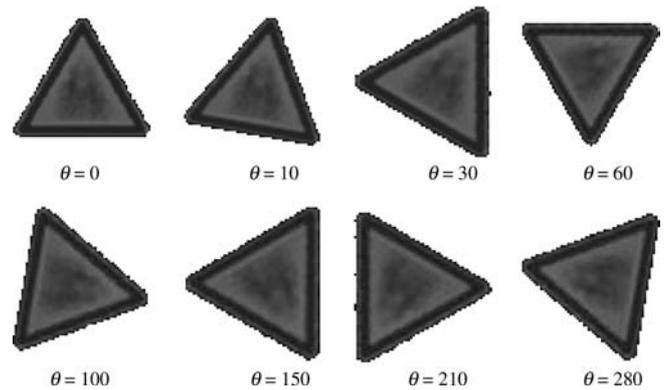


Fig. 3 AVG database: the mean image of the white traffic signs is learned with its rotation in the image plane ( $\theta$  denotes the rotation angle)

in the eigenspace are not Gaussian in either case. In the case of the COIL database, we obtain a low-dimensional non-linear manifold that can be parameterised by object pose [4]. In the case of the A43 database, we can observe two distinct clouds that correspond to the yellow road signs on one side and to the white signs on the other side. This variability is the main one and is captured by the first principal component,  $\mathbf{u}_1$ . The circles that appear in the plane defined by  $(\mathbf{u}_2, \mathbf{u}_3)$  are typical in the case of image plane rotation variability [26] (see also Fig. 6 for the case of the single-object database AVG and Sect. 5.2).

Although extensions to Gaussian mixture models [3, 12] can bring some flexibility to the standard representation, parametric models rely on the knowledge of the form of the underlying densities and might fail to fit the distributions actually encountered in practice [7]. In contrast, non-parametric density estimation methods are an efficient tool for modelling arbitrary distributions without making assumptions about the forms of the densities. We apply the Parzen window method, with Gaussian kernels, to the projections of the training images in the eigenspace:

$$\mathcal{P}(\mathbf{c}|\mathcal{B}) = \frac{1}{K} \sum_{k=1}^K \left( \frac{1}{\sqrt{2\pi}\sigma_p} \right)^J \exp \left[ \frac{-\|\mathbf{c} - \mathbf{c}_k\|^2}{2(\sigma_p)^2} \right] \quad (10)$$

We recall that  $K$  is the number of training images in the learning set  $\mathcal{B}$  and that  $J$  is the dimension of the



Fig. 4 Some of the A43 database training images

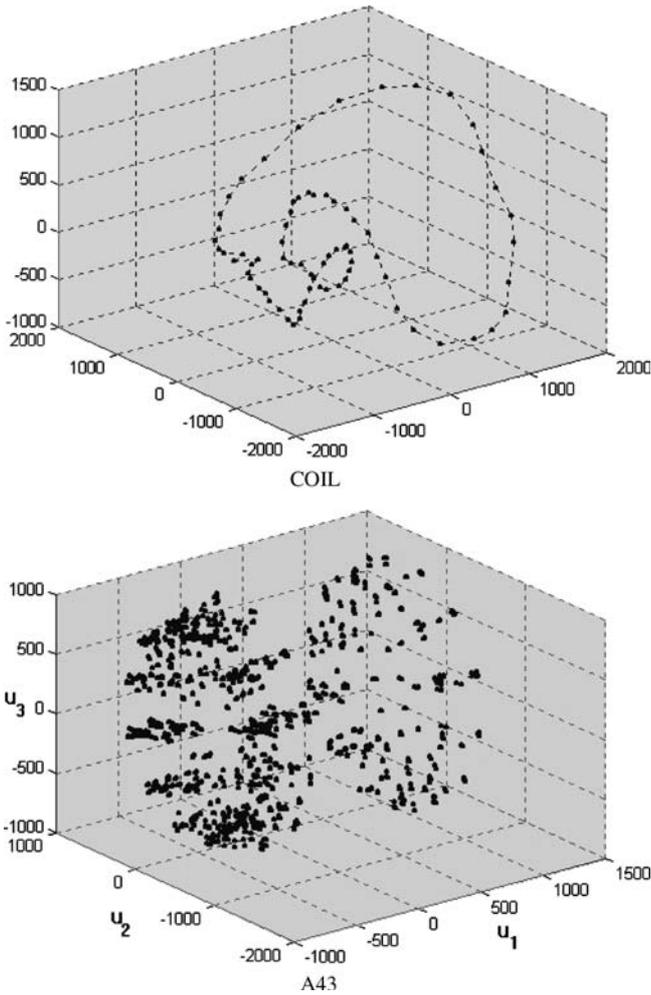


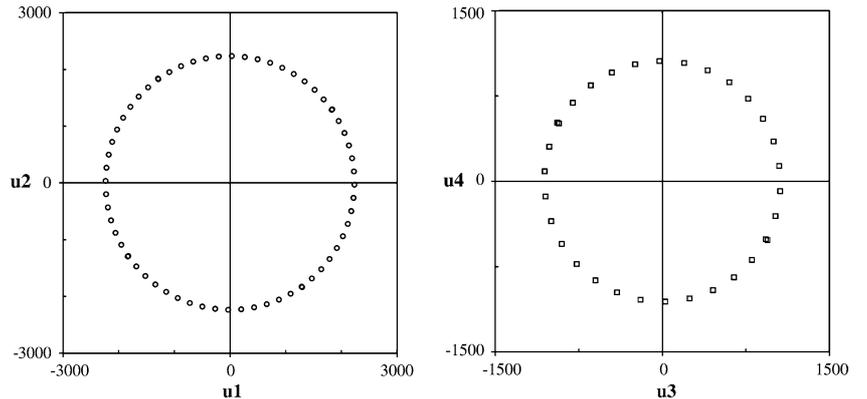
Fig. 5 Distribution of the latent variables  $\mathbf{c}$  of the training images in a 3D eigenspace

eigenspace. The parameter  $\sigma_P$ , called bandwidth, which controls the resolution of the pdf model, is set experimentally.

#### 4 Detection algorithms

Using approximation shown in Eq. 3 implies maximising the distribution  $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$  with respect to  $\mathbf{c}$ .

Fig. 6 Distribution of the latent variables  $\mathbf{c}$  in the first two planes of the eigenspace for the AVG database. The circular pattern is typical when learning image plane rotation variability [30]



This can be seen as a problem of reconstruction onto the eigenspace, in the sense of maximum a posteriori (MAP) estimation:

$$\begin{aligned}\hat{\mathbf{c}} &= \arg \max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) \\ &= \arg \max_{\mathbf{c}} \{\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B}) \cdot \mathcal{P}(\mathbf{c}|\mathcal{B})\}\end{aligned}\quad (11)$$

According to Eq. 3, the likelihood of the observation is approximated by:

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) \propto \mathcal{P}(\mathbf{y}|\hat{\mathbf{c}}, \mathcal{B}) \cdot \mathcal{P}(\hat{\mathbf{c}}|\mathcal{B})$$

We shall now consider different assumptions for the noise distribution  $\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B})$  and for the prior  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ . We derive an expression or algorithms for the estimation of  $\hat{\mathbf{c}}$  and the computation of  $-\ln \mathcal{P}(\mathbf{y}|\mathcal{B})$ . We begin by the Gaussian noise assumption, making a link with standard detectors associated to probabilistic PCA. Then, we present non-Gaussian noise models and come up with some robust detectors. Finally, the most general case obtained by considering non-Gaussian models both for the noise distribution and for the prior is presented in Sect. 4.3.

#### 4.1 Standard detection methods

In this paragraph, we consider the classical Gaussian hypothesis for the noise distribution.

##### 4.1.2 Gaussian noise, uniform prior

Let us first make the assumption of a uniform distribution of the eigenspace components,  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ . The joint likelihood is, hence, reduced to the noise distribution (as shown in Eq. 8) and the likelihood of the observation can be expressed as:

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) \propto \exp \left[ -\frac{\|\mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\hat{\mathbf{c}}\|^2}{2\sigma_g^2} \right] \quad (12)$$

The corresponding estimate  $\hat{\mathbf{c}}$  is the least squares solution, i.e. the projection of the observation onto the eigenspace:

$$\hat{\mathbf{c}} = \mathbf{U}^T(\mathbf{y} - \boldsymbol{\mu})$$

Taking the cologarithm of Eq. 12, we obtain the usual similarity measure called DFFS [10], corresponding to the orthogonal Euclidean distance between the observation and the eigenspace. This detector is renamed as GU, to recall the hypotheses: Gaussian noise distribution and uniform prior distribution in the eigenspace:

$$\text{GU}(\mathbf{y}) = \|\mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\hat{\mathbf{c}}\|^2 \quad (13)$$

#### 4.1.2 Gaussian noise, Gaussian prior

The latent variables are now assumed to follow a Gaussian prior: the random  $J$ -dimensional vector  $\mathbf{c}$ , is normally distributed with zero mean and covariance matrix  $\Lambda$ . This matrix is diagonal and collects the  $J$  principal eigenvalues  $\lambda_j$  computed during the learning phase by PCA. The likelihood is expressed as:

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) \propto \exp\left[-\frac{\|\mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\hat{\mathbf{c}}\|^2}{2\sigma_g^2}\right] \cdot \exp\left[-\frac{\hat{\mathbf{c}}^T \Lambda^{-1} \hat{\mathbf{c}}}{2}\right] \quad (14)$$

where  $\sigma_g$  is estimated as described in [3]. The MAP estimate is given by:

$$\hat{\mathbf{c}} = \left(\mathbf{I}_J + \sigma_g^2 \Lambda^{-1}\right)^{-1} \mathbf{U}^T \cdot (\mathbf{y} - \boldsymbol{\mu})$$

where  $\mathbf{I}_J$  is the  $J \times J$  identity matrix. Since  $\mathbf{I}_J$  and  $\Lambda$  are diagonal, the computation of the estimate is straightforward. Taking the cologarithm of Eq. 14 yields the following detector, called GG (Gaussian–Gaussian):

$$\text{GG}(\mathbf{y}) = \frac{\|\mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\hat{\mathbf{c}}\|^2}{\sigma_g^2} + \sum_{j=1}^J \frac{(\hat{c}_j)^2}{\lambda_j}$$

This expression is similar to the one proposed by Moghaddam [3] and Tipping [1].

## 4.2 Robust detection methods

In the following paragraphs,  $\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B})$  is no longer assumed to be Gaussian in order to take into account outliers.

#### 4.2.1 Robust noise model, uniform prior

For an easier presentation of the algorithms, the distribution of  $\mathbf{c}$  is first assumed to be uniform:  $\mathcal{P}(\mathbf{c}|\mathcal{B}) = \text{constant}$ . According to Eq. 9, the MAP energy can be written, up to an additive constant, as:

$$\mathcal{J}(\mathbf{c}) = \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right) \quad (15)$$

Using the half-quadratic theory [27], we introduce an *augmented* energy, denoted  $\mathcal{J}^*$ , depending on an additional variable  $\mathbf{b}$  and having the same minimum as  $\mathcal{J}$ :

$$\mathcal{J}(\mathbf{c}) = \min_{\mathbf{b}} \left\{ \mathcal{J}^*(\mathbf{c}, \mathbf{b}) = \sum_{n=1}^N \rho^*\left(\frac{w_n}{\sigma_\rho}, b_n\right) \right\}$$

The augmented energy is minimised alternately w.r.t.  $\mathbf{c}$  and  $\mathbf{b}$ . The minimum w.r.t.  $\mathbf{b}$  for a fixed value of  $\mathbf{c}$  is given by an analytic expression. The minimum w.r.t.  $\mathbf{c}$  for a fixed value of  $\mathbf{b}$  is computed using linear techniques [27]. Two expressions of  $\mathcal{J}^*$  have been proposed, leading to two different algorithms, which will now be presented.

*ARTUR or location step with modified weights* The first form of augmented energy is:

$$\mathcal{J}^A(\mathbf{c}, \mathbf{b}^A) = \sum_{n=1}^N b_n^A \left(\frac{w_n}{\sigma_\rho}\right)^2 + \psi(b_n^A)$$

This expression leads to the so-called *iterative reweighted least squares* algorithm, whose step ( $m$ ) can be written as:

$$\begin{cases} \mathbf{w}^{(m)} = \mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\mathbf{c}^{(m)} \\ \forall n \in \{1, \dots, N\}, \quad b_n^{A(m+1)} = \frac{\rho'\left(\frac{w_n}{\sigma_\rho}\right)}{2\frac{w_n}{\sigma_\rho}} \\ \mathbf{c}^{(m+1)} = (\mathbf{U}^T \mathbf{B}^{A(m+1)} \mathbf{U})^{-1} \mathbf{U}^T \mathbf{B}^{A(m+1)} (\mathbf{y} - \boldsymbol{\mu}) \end{cases} \quad (16)$$

where  $\mathbf{B}^{A(m+1)}$  is the diagonal matrix that collects the weights  $b_n^{A(m+1)}$ . This algorithm is widely used [28], in particular, in robust recognition [2, 24]. However, the matrix products and matrix inverse in Eq. 16 involve costly computations.

*LEGEND or the location step with modified residuals* [29] The second form of augmented energy is:

$$\mathcal{J}^L(\mathbf{c}, \mathbf{b}^L) = \sum_{n=1}^N \left(\frac{w_n}{\sigma_\rho} - b_n^L\right)^2 + \xi(b_n^L) \quad (17)$$

This expression leads to the so-called *iterative least squares algorithm with modified residuals*:

$$\begin{cases} \mathbf{w}^{(m)} = \mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\mathbf{c}^{(m)} \\ \forall n \in \{1, \dots, N\}, \quad b_n^{L(m+1)} = w_n \left(1 - \frac{\rho'\left(\frac{w_n}{\sigma_\rho}\right)}{2\frac{w_n}{\sigma_\rho}}\right) \\ \mathbf{c}^{(m+1)} = \mathbf{U}^T (\mathbf{y} - \boldsymbol{\mu} - \sigma_\rho \mathbf{b}^{L(m+1)}) \end{cases} \quad (18)$$

Both algorithms are equivalent to the ones proposed by Huber [14]. The latter is, as far as we know, seldom used in practice, although it has some attractive properties when reconstructing on an orthogonal basis. It can be shown that one step of LEGEND produces a smaller energy decrease than one step of ARTUR, which implies

a slower convergence rate. On the other hand, each step of LEGEND involves much less computation since it is tantamount to a simple projection. In our application, the LEGEND algorithm turned out to be faster than ARTUR in terms of computation time [29].

The estimate  $\hat{\mathbf{c}}$  is considered in both cases after convergence of the algorithm. It is computed with the Geman-McClure's function  $\rho$  (GM, see Table 1). Since this function is non-convex, we use a continuation strategy: the non-convexity is gradually introduced by successively considering the following functions: HS (hyper-surfaces, convex), HL (Hebert and Leahy) and finally GM, as proposed in [24] (see Table 1, for the expression of the different  $\rho$  functions). The scale parameter,  $\sigma_\rho$ , is estimated in a preliminary off-line step using the training images [24]. Hence, the method does not require any user interaction for parameter tuning.

The cologarithm of the likelihood may be written in this case as:

$$\text{RU}(\mathbf{y}) = \min_{\mathbf{c}} \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right) = \mathcal{J}(\hat{\mathbf{c}}) \quad (19)$$

It can be interpreted as a robust version of the DFFS and will be referred to as RU (for robust uniform).

#### 4.2.2 Robust noise model, Gaussian prior

Assuming a Gaussian a priori distribution in the eigenspace, the previous energy becomes:

$$\mathcal{J}(\mathbf{c}) = \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right) + \sum_{j=1}^J \frac{(c_j)^2}{\lambda_j} \quad (20)$$

The minimisation algorithms are similar to Eqs. 16 and 18, except for the estimation of  $\mathbf{c}^{(m+1)}$ , which becomes for ARTUR:

$$\mathbf{c}^{(m+1)} = \left(\mathbf{U}^T \mathbf{B}^{A(m+1)} \mathbf{U} + \sigma_\rho^2 \Lambda^{-1}\right)^{-1} \mathbf{U}^T \mathbf{B}^{A(m+1)} (\mathbf{y} - \boldsymbol{\mu})$$

and for LEGEND:

$$\mathbf{c}^{(m+1)} = \left(\mathbf{I}_J + \sigma_\rho^2 \Lambda^{-1}\right)^{-1} \mathbf{U}^T (\mathbf{y} - \boldsymbol{\mu} - \sigma_\rho b^{\mathbf{L}(m+1)})$$

Once again, the estimate is the value obtained after convergence. The cologarithm of the likelihood may be expressed as:

$$\text{RG}(\mathbf{y}) = \min_{\mathbf{c}} \left\{ \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right) + \sum_{j=1}^J \frac{(c_j)^2}{\lambda_j} \right\} \quad (21)$$

**Table 1** Robust  $\rho$  functions used in continuation [24]

Acronym	$\rho(x)$	Convexity
HS	$2\sqrt{1+x^2} - 2$	Convex
HL	$\log(1+x^2)$	Non-convex
GM	$x^2/(1+x^2)$	Non-convex

### 4.3 Detection using non-Gaussian distributions

We now consider the most general case where neither the noise distribution nor the prior is considered as Gaussian. In this case,  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  is estimated from the training database, using Parzen window estimates, as described in Sect. 3.5. However, the MAP estimation of  $\mathbf{c}$  according to Eq. 3:

$$\hat{\mathbf{c}}_{\text{MAP}} = \arg \max_{\mathbf{c}} \{\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B}) \mathcal{P}(\mathbf{c}|\mathcal{B})\} \quad (22)$$

is more involved, due to the more complex shape of the prior distribution. Therefore, we resort to a second approximation: we first compute the maximum likelihood estimate of  $\mathbf{c}$ :

$$\hat{\mathbf{c}}_{\text{ML}} = \arg \max_{\mathbf{c}} \{\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B})\} \quad (23)$$

and then we approximate the likelihood of the observation by:

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) \propto \mathcal{P}(\mathbf{y}|\hat{\mathbf{c}}_{\text{ML}}, \mathcal{B}) \mathcal{P}(\hat{\mathbf{c}}_{\text{ML}}, \mathcal{B}) \quad (24)$$

Such an approximation is, of course, only valid when the peaks of  $\mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{\text{ML}}|\mathcal{B})$  and  $\mathcal{P}(\hat{\mathbf{c}}|\mathcal{B})$  coincide, which is clearly not the case in general. However, we believe that it is justified in our case since the maximum likelihood reconstructions are already satisfactory, especially with a robust noise model [2]. Let us emphasise that the introduction of a non-Gaussian prior term in Eq. 24 implements a very useful constraint for the computation of  $\mathcal{P}(\mathbf{y}|\mathcal{B})$ , which is not taken into account by standard methods. This significantly improves the performance of the detection process, as shown by experimental results (see Sect. 5).

This latter detector will be referred to as robust non-Gaussian (RNG). In the most general case, when the prior distribution is modelled by Parzen windows with Gaussian kernels, the negative log-likelihood can be expressed as:

$$\begin{aligned} \text{RNG}(\mathbf{y}) = \min_{\mathbf{c}} \left\{ \frac{1}{2} \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right) \right\} \\ - \log \left[ \sum_{k=1}^K \exp \left[ -\frac{\|\hat{\mathbf{c}}_{\text{ML}} - \mathbf{c}_k\|^2}{2(\sigma_\rho)^2} \right] \right] \end{aligned} \quad (25)$$

The variance  $\sigma_\rho$  weights the influence of the a priori in the eigenspace with respect to the robust likelihood term.

## 5 Experimental results

This section is devoted to the assessment of the different detectors described previously by using the three databases presented in Sect. 3.5: COIL, AVG and A43. Test images have been created from occurrences of the objects of interest by embedding the objects in various textured backgrounds, with large occlusions (see for

instance Figs. 7 and 11). ROC curves enable an objective comparison of the different detectors. These are plots of the true positive rate against the false alarm rate. In our case, the former is defined as the ratio of correct decisions to the total number of occurrences of the objects, while the latter is the ratio of the number of incorrect decisions to the total number of possible false alarms (i.e. the locations where no object is present in the images—roughly, the size of the images  $\times$  the number of images). The correctness of detection is assessed using the following rule: since the exact position of the object of interest is known, the detection is considered to be correct if it occurs in the 8-neighbourhood of the true solution (i.e. a 1-pixel tolerance in accuracy). Note that the detection is performed by simple thresholding of the likelihood map, without any kind of post-processing. For better visualisation, the ROC curves presented hereafter are plotted on a semi-logarithmic scale.

Table 2 reviews the different proposed detectors. Their acronyms recall the underlying hypotheses about the noise model and the distribution of latent variables (i.e. in-eigenspace distribution). The detector proposed by Moghaddam and Pentland in [3], based on a Gaussian model, has also been implemented and tested.

### 5.1 Importance of a robust noise model

This first experiment compares the detectors based on Gaussian and non-Gaussian noise assumptions. Let us recall that the latter allows the presence of *outliers* in the observations.

We use the COIL database with  $J=5$ . The test set collects 21 scenes (300 $\times$ 200 pixels each) containing 57 occurrences of the objects of interest, with partial occlusions and cluttered background (see Fig. 7).

Figure 7 presents the likelihood maps computed with the GU, GG, RU and RNG detectors on two test scenes. For information, the mean computation time for a single likelihood map is about 1 min for the GG detector, 5 min for RU and RNG and 11 min for RG on an AMD 750 MHz PC using a non-optimised C program. In Fig. 7, the pixels outside the rectangular frame correspond to positions where complete observations could not be extracted. Visually, the results of GU and GG are quite similar. The RU detector leads to stronger peaks, allowing a better localisation of the objects of interest. The RNG detector, which integrates a non-Gaussian prior on the latent variables  $\mathbf{c}$ , leads to a likelihood map that is visually similar to the RU map.

**Table 2** Proposed detectors and their underlying assumptions

		$\mathcal{P}(\mathbf{c} \mathcal{B})$		
		Uniform	Gaussian	Non-Gaussian
$\mathcal{P}(\mathbf{w} \mathcal{B})$	Gaussian	GU <sub>(DFFS)</sub>	GG	–
	Robust	RU	RG	RNG

This visual impression is confirmed by the corresponding ROC curves, shown in Fig. 8, from which all detectors may be compared. Let us first notice that the GG detector and the Gaussian detector proposed by Moghaddam and Pentland in [3] have led to the same results in all our experiments, so we will only display one curve for both detectors in the rest of the paper. As expected, the RU detector exhibits significantly higher true positive rates than the standard Gaussian noise GU detector. Besides, Gaussian and uniform prior models on the latent variables  $\mathbf{c}$  lead to very similar results. This is hardly surprising since the distribution in the eigenspace is neither uniform nor Gaussian, as already illustrated in Fig. 5.

A non-Gaussian prior distribution is taken into account in the RNG detector (see Eq. 25), which is based on a non-Gaussian noise model and on a Parzen window estimate of the prior density. As can be seen, the introduction of an appropriate prior in the RNG detector slightly improves the results of RU (see Figs. 7 and 8). Overall, the RNG model leads to the best results—significantly better than detectors based on Gaussian assumptions only.

*Remark* As it can be expected, sample size and feature dimensionality have a significant impact on the proposed technique. Their influence on statistical pattern recognition methods based on learning is indeed now well documented. Diminishing the sample size or the feature dimensionality generally degrades the recognition performance. The ROC curves in Fig. 9 illustrate the influence of  $J$  for the RU detector. Degradation of the observation, i.e. noise, affects the results in the same way. We already noticed this in the case of object recognition [24]. This is illustrated in Fig. 10 where a zero-mean Gaussian noise with a standard deviation of 20 has been added to the analysed scenes, resulting in a 22-dB signal-to-noise ratio (SNR). For a 80% true positive rate, the false alarm rate is about 0.15% for the noiseless test set, while it is about 0.26% in the noisy case.

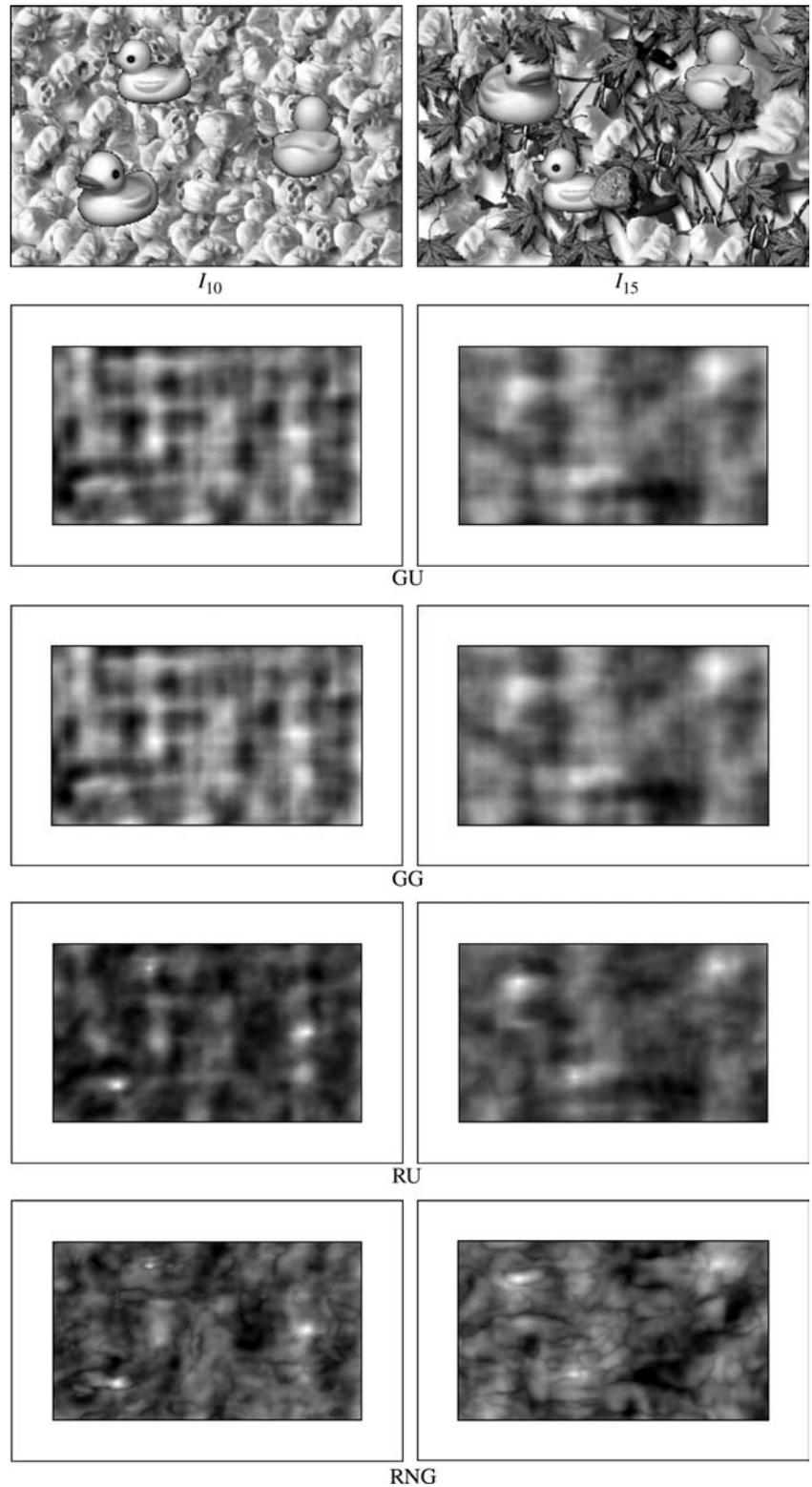
### 5.2 Importance of the prior model (I)

This second experiment shows that a careful modelling of the prior distribution in the eigenspace may be of great benefit to detection performances.

The tests have been conducted using the AVG training database, with  $J=20$ , considering 18 colour scenes (300 $\times$ 200 pixels each) containing 29 occurrences of triangular traffic signs. Two of them are presented in Fig. 11. A specificity of this test is that a trap, corresponding to the mean of the AVG training database, has been introduced in  $I_{17}$  (circular pattern on the left).

The mean image does not look at all like a triangular traffic sign. Nevertheless, it is a trap for the GU, GG and RU detectors, which do not take into account the

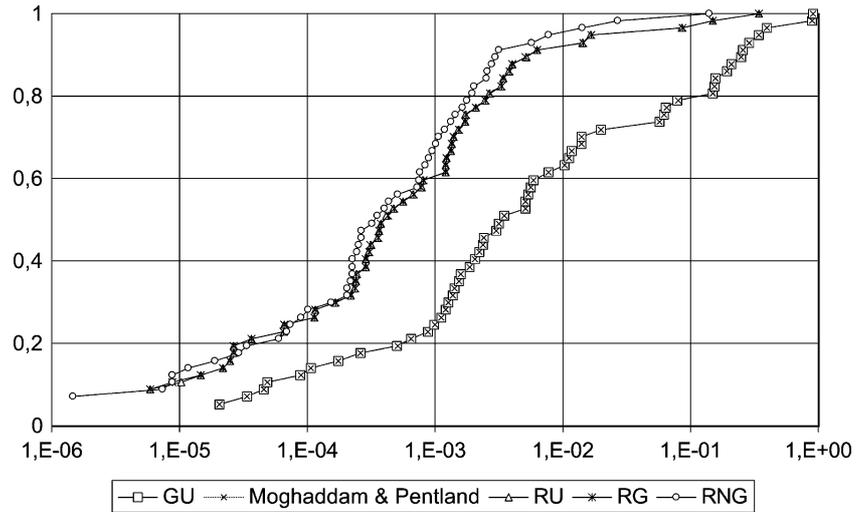
**Fig. 7** Examples of test scenes and their log-likelihood maps computed using the GU, GG, RU detectors and the complete model RNG. Bright intensities correspond to a high likelihood value. COIL database,  $J=5$



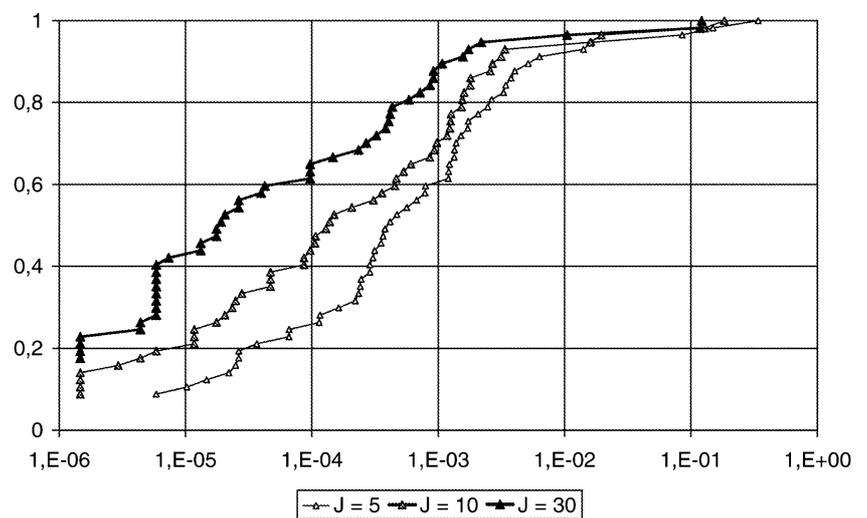
particular form of the underlying distribution in eigenspace. This is due to the fact that the mean image belongs to the eigenspace and, hence, minimises the reconstruction error, whatever the noise model, Gaussian or robust. As can be seen, the mean is detected in

scene  $I_{17}$  by the RU detector with an even higher likelihood value than the two other true targets (see Fig. 11). The robust RU detector is also misled by the grass areas in  $I_2$ . The performance of the different detectors are summarised Fig. 12: the Gaussian detec-

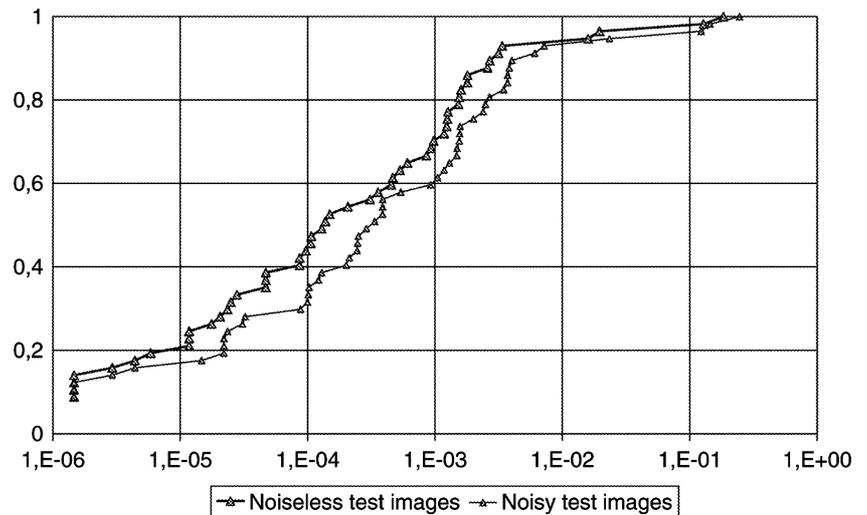
**Fig. 8** Receiver operating characteristic (ROC) curves for standard Gaussian detectors (GU, GG or Moghaddam and Pentland [3, 10]); robust noise model (RU and RG, which can hardly be distinguished); complete RNG model. COIL database,  $J=5$



**Fig. 9** Receiver operating characteristic (ROC) curves for the robust detector RU for different values of  $J$ , COIL database



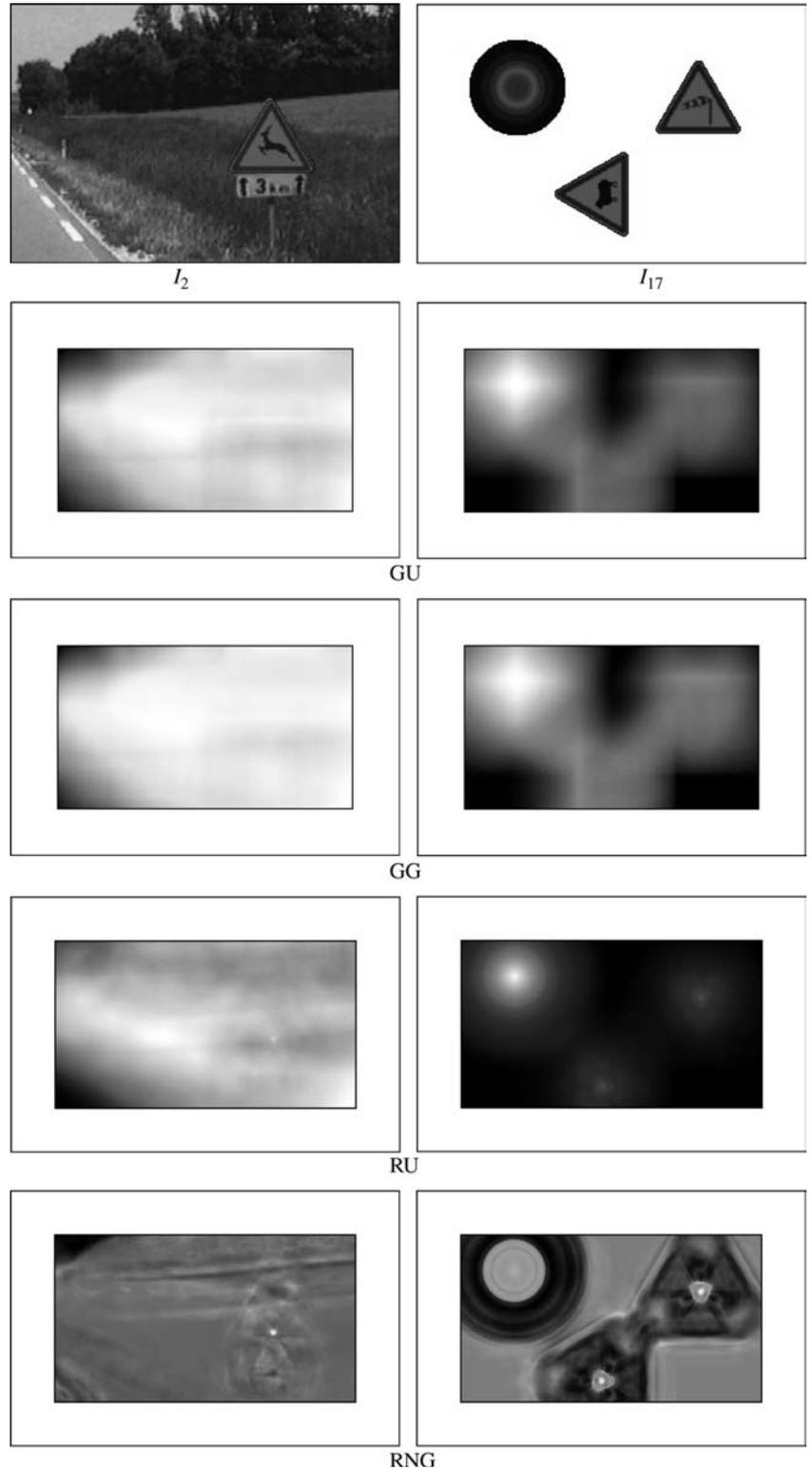
**Fig. 10** Influence of observation noise (SNR = 22 dB) on the robust detector RU, COIL database,  $J=10$



tors (GU, GG) yield similar poor results, in general. The robust RU detector only brings a slight improvement, with the ROC curves remaining unsatisfactory.

A significant improvement is obtained by introducing an adequate (prior) model of the eigenspace distribution, associated to a robust noise model. Since

**Fig. 11** Log-likelihood maps for scenes  $I_2$  and  $I_{17}$ . Bright intensities correspond to a high likelihood value. AVG database,  $J=20$

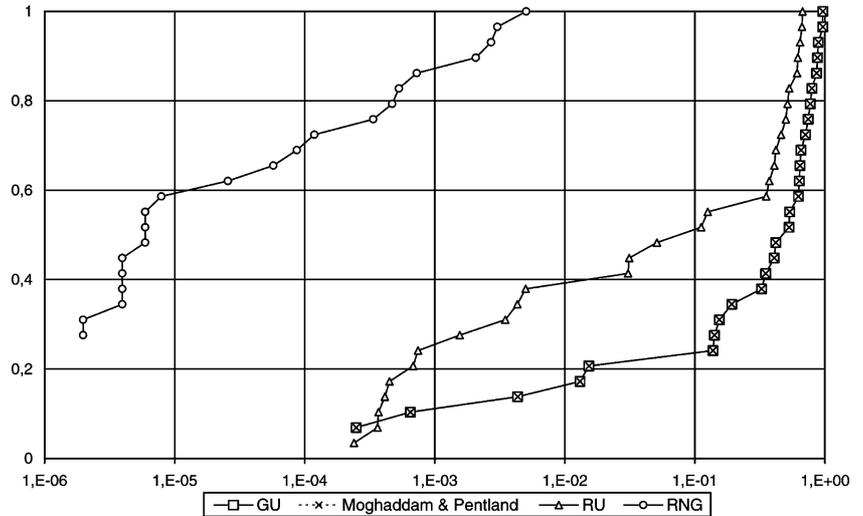


this particular database is composed of a single object rotating in the image plane, it is not necessary to resort to Parzen windows estimation for the prior density. Indeed, the analytic expression of  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  is known in this case [26]: the eigenvalues are double ( $\lambda_{2j} = \lambda_{2j-1}$ ) and, in each plane associated to a pair of eigenvectors, the coordinates of the training images in

the eigenspace are circularly distributed, with radius  $R^2 = \lambda_{2j} + \lambda_{2j-1}$  [29]. More precisely:

$$\mathcal{P}(\mathbf{c}|\mathcal{B}) \propto \prod_{j=1}^{\frac{J}{2}} \exp \left[ -\frac{1}{2\gamma} \left| (c_{2j-1})^2 + (c_{2j})^2 - \lambda_{2j} - \lambda_{2j-1} \right| \right] \quad (26)$$

**Fig. 12** Receiver operating characteristic (ROC) curves for standard Gaussian detectors (GU, GG or Moghaddam and Pentland [3, 10]); robust noise model (RU); complete RNG model. AVG database,  $J=20$



This is illustrated in Fig. 6. This remark also explains the circular shapes observed on the right side of Fig. 5. In this particular case, the detector RNG is defined by:

$$\text{RNG}(\mathbf{y}) = \min_{\mathbf{c}} \left\{ \sum_{n=1}^N \rho \left( \frac{w_n}{\sigma_\rho} \right) \right\} + \frac{1}{\gamma} \sum_{j=1}^{\frac{J}{2}} \left| (\hat{c}_{2j-1})^2 + (\hat{c}_{2j})^2 - \lambda_{2j} - \lambda_{2j-1} \right| \quad (27)$$

where  $\gamma$  plays the same role as  $\sigma_\rho$  in the general case. As expected, the performance of the complete RNG model are, by far, the best in this experiment (see Figs. 11 and 12).

### 5.3 Importance of the prior model (II)

This last experiment is another illustration of the importance of an accurate modelling of the eigenspace distribution, this time in the case of a more general form of the underlying pdf.

The A43 training database is used, with  $J=30$ . The detection test is performed over 27 colour scenes (300×200 pixels each) containing 58 occurrences of traffic signs (samples are shown Fig. 13).

The likelihood maps computed with RU for scenes  $I_2$  and  $I_{17}$  show the poor performance of this detector in this case (visually, it is only slightly better than the non-robust detectors, GU and GG). The positions of the objects of interest cannot be distinguished easily. This impression is confirmed by the ROC curves presented Fig. 14 for detectors GU and RU. The results using a Gaussian prior (detectors GG and RG) are identical, and are, therefore, not depicted here. Once again, the uniform or Gaussian assumptions on the prior are obviously not adapted to the true distribution in the eigenspace (see Fig. 5), which explains the poor results given by these

detectors. Using a robust noise model slightly improves the results, but they remain mediocre: detecting 80% of the objects yields a 68% false alarm rate!

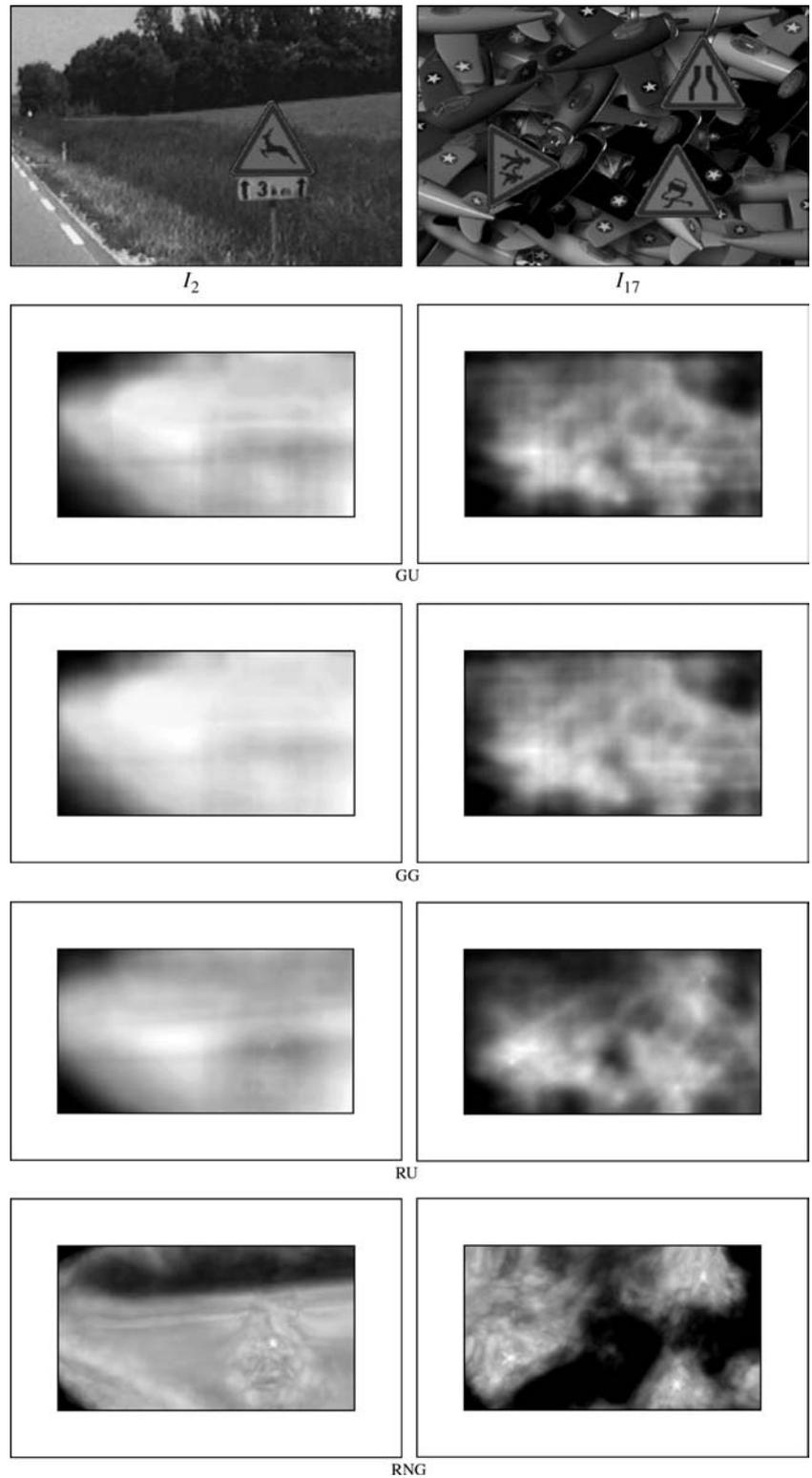
For the RNG detector, as already explained, the distribution in the eigenspace is modelled using Parzen windows with Gaussian kernels (see Sect. 4.3). We use the approximation in Eq. 23: the likelihood is computed using a robust ML estimate of the latent variables  $\mathbf{c}$ . Besides, a high weight is given to the prior term. The corresponding detection maps are presented Fig. 13. They allow an easy localisation of the target objects, which appear as bright spots on the likelihood maps. Figure 14 shows the corresponding ROC curve. One can readily see the improvement brought by the complete RNG model: more than 70% of the objects of interest are detected before the first false alarms appear. An accurate model for the distribution in the eigenspace is, therefore, essential.

## 6 Conclusion

In this paper, we have presented a novel Bayesian approach for object detection using global appearance-based representations. The proposed framework combines non-Gaussian noise models with general, non-linear assumptions about the distribution of latent variables in the eigenspace. Non-Gaussian noise models yield robust estimators, which can deal with severely degraded occurrences of objects. A key feature of the proposed approach is its ability to embed non-linear priors on the eigenspace in a linear latent variable representation. This significantly improves the performances of the detector in critical situations.

This work finally unifies several standard detection methods proposed in the literature and leads to the definition of a new family of probabilistic detectors able to cope with complex object distributions and adverse situations, such as cluttered backgrounds, partial occlusions or corrupted observations.

**Fig. 13** Examples of likelihood maps (bright intensities correspond to a high likelihood value). A43 database,  $J=30$

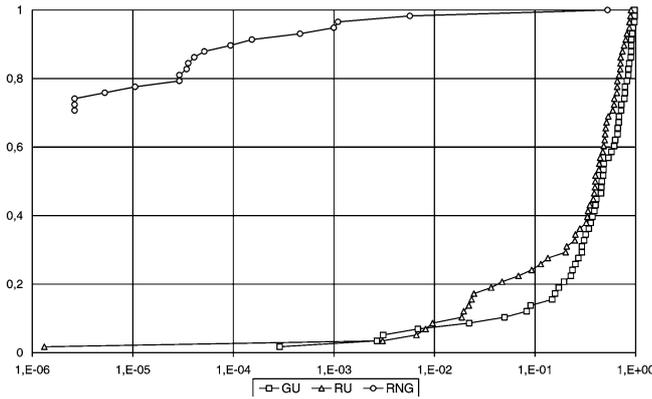


## 7 Originality and contribution

In this paper, we are interested in a particular class of appearance-based representations [4, 5], namely probabilistic appearance models [3, 10]. They can represent

large classes of images and make available all the traditional methods of statistical estimation.

Our Bayesian model is inspired by the latent variable representation proposed by Tipping and Bishop [1] in the Gaussian case (namely, probabilistic PCA, or PPCA for short). The originality of our approach is that it



**Fig. 14** Receiver operating characteristic (ROC) curves for the GU and RU detectors and for the complete model (RNG). A43 database,  $J=30$

explicitly takes into account general, non-Gaussian forms of the underlying distributions, both for the prior and for the observation model. In particular, it straightforwardly integrates non-linear models for the distribution of the images in the eigenspace. Thus, it deviates from standard PPCA, and, in particular, the parameters of the models are no longer maximum likelihood estimates. The benefit of our approach is its ability to better represent the complex distributions that may occur in practical applications. The proposed framework also unifies the main PCA-based models mentioned in the literature [2, 3].

The performance of the approach has been assessed using receiver operating characteristic (ROC) curve analysis on several representative databases. The experimental results clearly show the impact of an appropriate model for the in-eigenspace distribution on the performances of the detection process. Moreover, the approach also enables, when necessary, to introduce robust hypotheses on the distribution of noise, allowing to cope with clutter, outliers and occlusions, which is also illustrated by the experimental results.

The main contribution of the paper is, thus, the definition of a novel family of general purpose detectors that experimentally compare favourably with several state-of-the-art PCA-based detectors recently described in the literature.

## 8 About the authors

Rozenn Dahyot received the diploma of the general engineering school ENSPS in France and an MSc (DEA) in computer vision from the University of Strasbourg in 1998. She gained her Ph.D. in image processing from the University of Strasbourg, France in 2001. She is currently a research associate in the Department of Statistics at Trinity College, Dublin, Ireland. Her research interests concern multimedia understanding, object or event detection and recognition and statistical learning, amongst others.

Pierre Charbonnier obtained his engineering degree (1991) and his Ph.D. qualification (1994) from the University of Nice-Sophia Antipolis, France. He is currently a senior researcher (“Chargé de Recherche”) for the French Ministry of Equipment, Transport and Housing at the Laboratoire Régional des Ponts et Chaussées in Strasbourg (ERA 27 LCPC), France. His interests include statistical models and deformable models applied to image analysis.

Fabrice Heitz received his engineering degree in electrical engineering and telecommunications from Telecom Bretagne, France, in 1984 and his Ph.D. degree from Telecom Paris, France, in 1988. From 1988 until 1994, he was with INRIA Rennes as a senior researcher in image processing and computer vision. He is now a professor at Ecole Nationale Supérieure de Physique, Strasbourg, (Image Science, Computer Science and Remote Sensing Laboratory LSIT UMR CNRS 7005), France. His research interests include statistical image modelling, image sequence analysis and medical image analysis. Professor Heitz was an associate editor for the IEEE Transactions on Image Processing journal from 1996 to 1999. He is currently the assistant director of LSIT.

**Acknowledgements** This work was supported by a Ph.D. grant awarded by the Laboratoire Central des Ponts-et-Chaussées, France.

## References

1. Tipping ME, Bishop CM (1999) Probabilistic principal component analysis. *J Roy Stat Soc B* 61(3):611–622
2. Black MJ, Jepson AD (1998) Eigentracking: robust matching and tracking of articulated objects using a view-based representation. *Int J Comput Vis* 26(1):63–84
3. Moghaddam B, Pentland A (1997) Probabilistic visual learning for object representation. *IEEE Trans Pattern Anal Machine Intell* 19(7):696–710
4. Murase H, Nayar SK (1995) Visual learning and recognition of 3-D objects from appearance. *Int J Comput Vis* 14(1):5–24
5. Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3(1):71–86
6. Schneiderman H (2000) A statistical approach to 3D object detection applied to faces and cars. PhD thesis, Carnegie Mellon University, Pittsburg, Pennsylvania
7. Duda RO, Hart PE, Stork DG (2001) *Pattern classification*, 2nd edn. Wiley, New York
8. Moghaddam B (2002) Principal manifolds and Bayesian subspaces for visual recognition. *IEEE Trans Pattern Anal Machine Intell* 24(6):780–788
9. Saul LK, Roweis ST (2003) Think globally, fit locally: unsupervised learning of low dimensional manifolds. *J Machine Learn Res* 4:119–155
10. Moghaddam B, Pentland A (1995) Probabilistic visual learning for object detection. In: *Proceedings of the 5th international conference on computer vision*, Cambridge, Massachusetts, June 1995, pp 786–793
11. Hamdan R, Heitz F, Thoraval L (2003) A low complexity approximation of probabilistic appearance models. *Pattern Recogn* 36(5):1107–1118
12. Tipping ME, Bishop CM (1999) Mixtures of probabilistic principal component analysers. *Neural Comput* 11(2):443–482
13. Roweis ST (1998) EM algorithms for PCA and SPCA. In: Jordan, MI, Kearns MJ, Solla SA (eds) *Advances in neural infor-*

- mation processing systems, vol 10. MIT Press, Cambridge, Massachusetts, pp 626–632
14. Huber PJ (1981) Robust statistics. Wiley, New York
  15. Leonardis A, Bischof H (2000) Robust recognition using eigenimages. *Comput Vis Image Und, CVIU* 7(1):99–118
  16. Kramer MA (1991) Nonlinear principal component analysis using autoassociative neural networks. *AiChE J* 32(2):233–243
  17. Kohonen T (2001) Self-organizing maps, vol 30, 3rd edn. Springer, Berlin Heidelberg New York
  18. Hastie T, Stuetzle W (1989) Principal curves. *J Am Stat Assoc* 84(406):502–516
  19. Chalmond B, Girard S (1999) Nonlinear modeling of scattered multivariate data and its application to shape change. *IEEE Trans Pattern Anal Machine Intell* 21(5):422–432
  20. Chang K, Ghosh J (2001) A unified model for probabilistic principal surfaces. *IEEE Trans Pattern Anal Machine Intell* 23(1):22–41
  21. Scholkopf B, Smola A, Muller K (1998) Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput* 10(5):1299–1319
  22. Bernardo JM, Smith AF (2000) Bayesian theory. Wiley, New York
  23. MacKay DJC (1995) Probable network and plausible predictions—a review of practical Bayesian methods for supervised neural networks. *Network—Comput Neural* 6(3):469–505
  24. Dahyot R, Charbonnier P, Heitz F (2000) Robust visual recognition of colour images. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR 2000), Hilton Head Island, South Carolina, June 2000, vol 1, pp 685–690
  25. Nene SA, Nayar SK, Murase H (1996) Columbia object image library (COIL-20). Technical report CUCS-005-96, Department of Computer Science, Columbia University
  26. Park RH (2002) Comments on “optimal approximation of uniformly rotated images: relationship between Karhunen-Loeve expansion and discrete cosine transform.” *IEEE Trans Image Processing* 11(3):332–334
  27. Charbonnier P, Blanc-Féraud L, Aubert G, Barlaud M (1994) Two deterministic half-quadratic regularization algorithms for computed imaging. In: Proceedings IEEE international conference on image processing (ICIP'94), Austin, Texas, November 1994, pp 168–172
  28. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1995) Numerical recipes in C: the art of scientific computing. Cambridge University Press, Cambridge, UK
  29. Dahyot R (2001) Appearance-based road scene video analysis for the management of the road network (in French). PhD thesis, Université Louis Pasteur Strasbourg, France
  30. Jogan M, Leonardis A (2001) Parametric eigenspace representations of panoramic images. In: Proceedings of the 10th international conference on advanced robotics (ICAR 2001), 2nd workshop on omnidirectional vision applied to robotic orientation and nondestructive testing (NDT), Budapest, Hungary, August 2001, pp 31–36