Comparative study of global invariant descriptors for object recognition

Anant Choksuriwong, Bruno Emile, Christophe Rosenberger, Hélène Laurent

Comparative study of global invariant descriptors for object recognition

A. Choksuriwong, B. Emile, C. Rosenberger, H. Laurent

Laboratoire Vision et Robotique - UPRES EA 2078

ENSI de Bourges - Université d'Orléans

88 boulevard Lahitolle, 18020 Bourges cedex, France

phone: (+33) 254082550, fax: (+33) 254077800

bruno.emile@univ-orleans.fr

Abstract: Even if lots of object invariant descriptors have been proposed in the literature, putting them into practice in order to obtain a robust recognition system face to several perturbations is still a studied problem. After presenting the most commonly used global invariant descriptors, a comparative study permits to show their ability to discriminate between objects with few training. The COIL-100 image database that presents a same object translated, rotated and scaled is used to test the invariance face to geometrical transforms. Partial object occultation or presence of complex background are examples of used images in order to test the robustness of the studied descriptors. The article compares them in a global and in a local context (computed on the neighborhood of a pixel). The SIFT descriptor is used as reference for local invariant descriptors. This study shows the relative performance of invariant descriptors used in a global and in a local context and identifies the different situations they are best suited.

Keywords : Comparative study, invariant features, object recognition, support vector machine, SIFT.

1 Introduction

The interpretation of images is still a complex problem [4], [15] and is primordial in lots of applications (military, industrial and medical). Lots of research works have been achieved for the conception of vision systems in order to recognize object in an image [17]. Object recognition methods can be classified into two classes: methods using invariant features [16], [19] or non-invariant features [21],[23]. Our study concerns the first class. The object recognition principle can then be decomposed in two steps: the first one consists in learning the object description in a database by a set of invariant features (learning phase). The second one recognizes an object by using a selected attributes vector which is extracted from the image and used by a supervised classification method (classification phase).

Appearing in the two phases (learning and classification), the choice of features that characterize an object and discriminate it from the others is a fundamental step in object recognition. An object can appear at different locations and sizes in an image. Consequently, features must be invariant by translation, rotation and scale. Lots of works have been achieved to solve the problem of object recognition using different descriptors calculated on all the image.

There exists four classes of features:

- by transformation: invariants are obtained from image transform [14],
- algebraic: moments are calculated from images [3], [12],
- by learning: a learning phase (by neural networks) can permit to obtain a recognition scheme which will be invariant to geometric transformations. The learning database is then composed on multiple images presenting the same object at different scales and rotations [22],

- combined approaches [16].

More than invariance property, the features robustness is an important point. Object can appear truncated, with different luminances, colors, or on a complex background (noisy or textured for example).

In this paper, different global invariant features are compared in order to identify the best one in different contexts. The first part presents different features from the literature. The second part presents the experimental protocol, and the comparison results for the global and local (global features applied locally) approaches. The SIFT descriptor is used as reference for local invariant descriptors [16]. We finally conclude this paper and give some perspectives of this work.

2 Pattern recognition

In order to recognize an object in an image, we need to make several choices. The first one concerns the selection of characteristic features for each object. These features must have in general some properties as invariance by rotation, scale and translation of the object. They can be directly computed on the global image or locally in the neighborhood of some detected key-points. The second choice concerns the decision criterion for the object recognition among one of the known objects of the database using the previous features. We present in the next section several features from the state of art.

2.1 Features invariant by translation, rotation and scale

Let $I(x, y)$ be a grey-level pixel of the image I.

2.1.1 Invariants by transformation

Fourier-Mellin transform (FMT)

The standard Fourier-Mellin transform [5] of an image I described in the radial polar form is given by ($(k, v) \in \mathbb{Z} \times \mathbb{R}$):

$$M_I(k, v) = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} I(r, \theta) r^{-iv} e^{-ik\theta} \frac{dr}{r} d\theta \tag{1}$$

In [9], the author suggests computing the standard FMT of $I_\sigma(r, \theta) = r^\sigma I(r, \theta)$ instead of $I(r, \theta)$, where $\sigma > 0$, defining the analytical Fourier-Mellin transform (AFMT) of I:

$$M_{I_\sigma}(k, v) = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} I_\sigma(r, \theta) r^{-iv} e^{-ik\theta} \frac{dr}{r} d\theta \tag{2}$$

The complete family of similarity invariant features based on the AFMT and suggested in [9] can then be easily rewritten and applied to any strictly positive σ value [5]:

$$F_{I_\sigma}(k, v) = M_{I_\sigma}(0, 0)^{\frac{-\sigma+iv}{\sigma}} e^{ikArg(M_{I_\sigma}(1,0))} M_{I_\sigma}(k, v) \tag{3}$$

Each feature $M_{I_\sigma}(k, v)$ is constructed in order to compensate for the rotation and size changes of an object. These features yield invariance with respect to translation, scale and rotation. For the following experiments, σ was set to 1 and 33 paramaters are computed.

2.1.2 Algebric invariants

Algebraic invariants are obtained from composition of moments (quotients and powers). A moment is the sum of all the pixels of the image weighted by polynomials depending on the pixel position.

Hu's moments (Hu)

The $(p + q)th$ order moments $(p > 0, q > 0)$ of I are given by:

$$m_{p,q} = \int_{R^2} x^p y^q I(x, y) dx dy \tag{4}$$

The centroid (x_0, y_0) of I is defined by: $x_0 = \frac{m_{1,0}}{m_{0,0}}$ and $y_0 = \frac{m_{0,1}}{m_{0,0}}$. The centered image I_T, with $I_T(x,y) = I(x + x_0, y + y_0)$ is translation-invariant.

The central moments, which are also translation-invariant, are defined by:

$$v_{p,q} = \int_{R^2} x^p y^q I_T(x,y) dxdy. \tag{5}$$

Finally, the normalized moments, described below, are scale-invariant:

$$\mu_{p,q} = \frac{v_{p,q}}{v_{0,0}^{(1+(p+q)/2)}} \tag{6}$$

The seven Hu moments defined [12] and derived from these functions, are translation, scale and rotation-invariant.

Zernike's moments (Zernike)

Zernike's moments use a set of Zernike polynomials. This set is complete and orthonormal in the interior of the unit circle. These 2D image moments allow to overcome the major drawbacks of regular geometrical moments regarding noise effects and presence of image quantization error. Their orthogonality property helps in achieving a near zero value of redundancy measure in a set of moments functions [3], [13]. The Zernike moments formulation is given below:

$$A_{mn} = \frac{m+1}{\pi} \sum_x \sum_y I(x,y)[V_{mn}(x,y)] \tag{7}$$

with $x^2+y^2 < 1$. The values of m and n define the moment order. Zernike polynomials $V_{mn}(x,y)$ are expressed in the radial-polar form:

$$V_{mn}(r,\theta) = R_{mn}(r)e^{jn\theta} \tag{8}$$

where $R_{mn}(r)$ is the radial polynomial given by:

$$R_{mn}(r) = \sum_{s=0}^{\frac{m-|n|}{2}} \frac{(-1)^s (m-s)! r^{m-2s}}{s!(\frac{m+|n|}{2} - s)!(\frac{m-|n|}{2} - s)!} \tag{9}$$

These moments yield invariance with respect to translation, scale and rotation. For this study m and n were set to 15 and resulting for each image in the computation of 72 features.

2.1.3 Local features

Local features on a neighborhood

We proposed in [2] a scheme for the computation of invariant descriptor in a local context. The first step of the computation of local features is to detect some key-points in the image.

Lots of key-points detectors have been proposed in the literature [20]. They are either based on a preliminary contour detection or directly computed on grey-level images. The Harris detector [10] that is used in this article belongs to the second class; it is consequently not dependant of a prior success of the contour extraction step. This detector is based on statistics of the image and rests on the detection of average changes of the auto-correlation function. Figure 1 presents the interest points obtained for one object extracted from the COIL-100 basis and presented under geometric transformations. We can observe that not all points are systematically detected. However, this example shows the good repeatability of the obtained key-points.

The second step concerns the local computation of descriptors. Invariant features (Fourier-Mellin, Hu and Zernike) are then computed on a neighborhood of each detected key-point (see Figure 2). The neighborhood is a window of size $w \times w$ pixels centered on the central pixel (with w odd value).

SIFT algorithm

The invariant descriptor developed in the SIFT algorithm, described in [16], is applied locally on key-points and is based upon the image gradients in a local neighborhood. This descriptor is used for the comparison of global descriptors applied locally. A recent work [18] shows that the SIFT algorithm is the most efficient local descriptor. The descriptor is created by sampling the magnitudes and orientations of the image gradients in a neighborhood of each key-point and building smoothed orientation histograms that contain the important aspect of the neighborhood. Each local descriptor is composed on a 4x4 array (histogram) composed each with 8 orientation components. A 128-elements vector is then built for each key-point. We used in this article the implementation provided by Lowe [16].

3 Comparative study

The goal of the proposed comparative study is to evaluate the invariant descriptors performances on a large image database for object recognition. The three invariant descriptors (Hu's moments, Zernike's moments and Fourier-Mellin transform features) and the local descriptor SIFT are applied in this section on a database of 100 objects extracted from the Columbia Object Image Library (COIL-100). After presenting the images data set, we analyze some experimental results. In order to estimate the features capabilities, we use a support vector machine (SVM) as supervised classification method. The recognition performances of the different descriptors regarding invariance to geometrical transform, discrimination capability and robustness face to several perturbations are compared.

3.1 Experimental protocol

3.1.1 Images database

In order to evaluate the invariance properties of the descriptors, the database is composed, for each object, of 72 grey-level images presenting rotation and scale changes (see Figure 3). 75 images presenting different perturbations (see Figure 4) are used for robustness evaluation:

- 10 views with a uniform grey-level background,
- 10 views with noised grey-level background,
- 10 views with a textured background,

- 10 views with a black occlusion,
- 10 views with a grey-level occlusion,
- 10 views with a luminance modification,
- 15 views with an additive gaussian noise (standard deviation of 5, 10 or 20).

We use so in this paper a database composed of 14.700 images.

3.1.2 Supervised classification using Support Vector Machine (SVM)

The goal of this section is to present the decision criterion using the previous features for the object recognition among one of the objects contained in the knowledge database. Some known classification methods are for example the fuzzy classifier based on the k-nearest neighbors [1] or the minimization distance from a class [6] but the most currently effective supervised classification method is the SVM. The principle of this method is now presented.

For two classes problems, $y_i \in \{-1, 1\}$, $i = 1 : l$, the Support Vector Machine implements the following algorithm. First of all, the training points $\mathbf{x}_i$ (the features) are projected in a space $\mathcal{H}$ (of possibly infinite dimension) thanks to a function $\Phi(\cdot)$. Then, the goal is to find in this space, an optimal decision hyperplane, in the sense of a criterion that we are going to define [24]. Note that for a same training set, different transformations $\Phi(\cdot)$ lead to different decision function. The transformation is achieved in an implicit manner thanks to a kernel $K(\cdot, \cdot)$ and the decision function is defined as:

$$f(\mathbf{x}) = \sum_{i=1}^{\ell} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{10}$$

with $\alpha_i \in \mathrm{I\!R}$ and b are the parameters of the hyperplane. In SVMs, the optimality criterion to maximize is the margin, that is the distance between the hyperplane and the nearest point

$\Phi(\mathbf{x}_i)$ of the training set (see Figure 5 in the case where $\Phi(\cdot)$ is the identity function). The α_i allowing to optimize this criterion are defined by solving the following problem:

$$\begin{cases} \max_{\alpha_i} \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i K(\mathbf{x}_i, \mathbf{x}_j y_j) \\ \text{with constraints,} \\ 0 \ \leq \ \alpha_i \ \leq \ C \ , \\ \sum_{i=1}^{\ell} \alpha_i y_i \ = \ 0 \ . \end{cases} \tag{11}$$

where C is a coefficient penalizing examples located in or beyond the margin and providing to reach a compromise between their numbers and the width of the margin.

Originally, SVMs have essentially been developed for the two classes problems, however several approaches can be used for extending to multi-class problems [11]. The method we use in this communication, is called *one against one*. Instead of learning N decision functions, each class is here discriminated from another one. Thus, $\frac{N(N-1)}{2}$ decision functions are learned and each of them makes a vote for the affectation of a new point $\mathbf{x}$.

3.2 Comparison of global features

Several aspects are taken in consideration to characterize the performances of a feature:

- its invariance property face to the geometric transformations,
- its discriminating power,
- its robustness regarding the perturbations due to the acquirement or the extraction phase.

The performances of the different global features are now compared.

3.2.1 Preliminary study

We present first some preliminary results obtained for four objects (see Figure 6) as illustration. Two of these four objects are visually relatively similar.

If we want to discriminate one object from the other independently of its size and its position, the feature values must be similar for the same object in every images. Let E_j be a vector of values $\sigma_{1,j}, ..., \sigma_{i,j}, ..., \sigma_{100,j}$ where $\sigma_{i,j}$ is the standard deviation of j^{th} feature on the first 72 views (without alteration) of the i^{th} object. Each value of E_j must be as minimal as possible to keep near values of the feature for a same object (especially for totally symmetric objects).

The feature must also discriminate different objects. Let M_j be a vector of values $m_{1,j}, ..., m_{i,j}$, $..., m_{100,j}$ where $m_{i,j}$ is the mean value of the j^{th} feature of the i^{th} object. Each value M_j must be as different as possible from the others for differentiating the objects.

The Figure 7 presents as illustration, the evolution of a feature of the three families (Hu, Zernike and Fourier-Mellin) that tends to expected behavior. The X-axis corresponds to the features values obtained for the 72 views of the four objects (without any alterations).

If we consider the object number 34, the values of the fourth Zernike moment are relatively similar for different views of the object and leads to low value of the standard deviation. That means this descriptor has good properties of invariance. The 18th Fourier-Mellin descriptor has a similar behavior for this object. On the contrary, the third Hu's moment does not have this behavior. If we consider now the four curves, they are much different for the Fourier-Mellin descriptor and the Zernike moment than for the Hu's moment.

Figure 8 presents the same features but the database is now composed of the 75 images presenting different perturbations. If we look at the different curves, it is hard to put into obviousness a correct behavior of these three invariant descriptors. We can notice nevertheless that the Hu's moment has the same value for different objects, this shows an incorrect behavior.

One notes first in this preliminary study that the Hu's moment seems to be less discriminative and invariant than the two others. The Fourier-Mellin and Zernike' descriptors are more relevant to distinguish two different objects.

3.2.2 Recognition results

The performances of the different invariant descriptors are analyzed with respect to the recognition rate given a learning set. Hence, for a given ratio, the learning and testing sets have been built by splitting randomly all examples. Then, due to the randomness of this procedure, multiple trials have been performed with different random draws of the learning and testing set.

We list in the following paragraph, all the parameters of our experiment:

1. *The learning set* x_i: corresponding to the values of an invariant descriptor computed on an image from the database.

2. *The classes* $y_i \in \{1, 100\}$ corresponding to the object class.

3. *Algorithm performance*: the efficiency is given through a percentage of the well recognized objects composing the testing set.

4. *Number of trials*: fixed to 100, in order to compensate the random drawing.

5. *Kernel* K: a gaussian kernel of bandwidth σ is chosen

$$K(x, y) = e^{\frac{-\|x-y\|^2}{2\sigma^2}} \tag{12}$$

x and y correspond to the descriptors vectors of objects.

6. *bandwidth* σ: set to 2 (data are normalized) after optimization [7].

7. *penalization coefficient* C: set to 100 after optimization [7].

In order to complete this preliminary study, a recognition rate comparison of the three descriptors is made on 72 views of each object without alteration for different sizes of the learning database using the SVM classification method. The learning database is composed of some views of each object (N among 72) and the recognition tests are realized on the remaining database (72-N views for each object). Note that 100 objects are in this database, we have so 7.200 images. Experimental results are presented in Figure 9. We first note that for all features the recognition rate grows with the size of the learning database. Second, the comparison result is clear, the best method is Zernike followed by Fourier-Mellin and Hu moments. By using 54 views of each object, the global Zernike moments allow to recognize correctly in 92% cases the 18 left views.

3.2.3 Robustness study

In a second step, the learning database is composed, for all the 100 objects, of the 72 views without perturbation and we compute the recognition rate on perturbed images composed of 75 altered views for each object (see table 1). Note that 100 objects are in this database, we have so 14.700 images. The results show that Zernike moments are the best descriptors in perturbing context except for background changes and luminance variations where Fourrier-Mellin features are more relevant.

3.2.4 Conclusion of global features comparison

The comparison of the global features shows that the Zernike moments give the best results in many different configurations when computed on the whole image. The recognition rate however remains weak when very few objects of the database are learned. It can penalize, for example some robotic applications where the object training is extracted from only few images. We can also note that the recognition rate remains weak face to complex backgrounds.

3.3 Comparison of local features

A comparison study between the invariants applied locally (Fourier-Mellin, Hu and Zernike) and the classical local feature SIFT (used here as reference) will permit to identify the potential benefit of invariants in a local context. For the local features based on a neighborhood, the size of the window around the key-points and the choice of the Harris algorithm parameters can be adjusted.

The class of each key-point is obtained by the SVM that takes into account the invariant descriptors computed within its neighborhood. The performances of Local Zernike (LZ), Local Hu (LH) and Local Fourier-Mellin (LFM) correspond then to a percentage of well recognized key-points. The class of an unknown object can correspond to the majority class after the voting (MV) of each key-point [8]. We then computed Local Zernike Majority Voting (LZMV). This approach permits the recognition of an object even if for example it is occulted by another one.

3.3.1 Recognition results

We made a comparative study of the local features by considering the recognition rate. As for the global study, the learning database is first composed of some views (without alteration) of each object (N among 72) and the recognition tests are realized on the remaining database (72-N views for each object). The local features on a neighborhood are calculated on each key-point so the learning phase is longer than the one for the global approach.

As shown in table 2, the recognition rate increases with the database size for unaltered images. The local Zernike approach with majority voting gives a little poorer results than the SIFT algorithm except for a training database with few images. On the other hand, the local method gives better results than the global one.

Table 3 shows the evolution of the recognition rate for the local Zernike approach face to the size of the neighborhood (use of 50% of the database in the learning ans test databases). We notice that we obtain similar performances for different sizes of the neighborhood even if the size 15×15 pixels is in this case the optimal choice.

3.3.2 Robustness results

As for the comparison in a global context, the learning database is now composed for all the 100 objects of the 72 views without alteration and the recognition rate is computed for perturbed images (75 for each object). The results are presented in table 4.

The results depend on the applied perturbation. For the uniform or textured backgrounds, SIFT gives the best recognition rate. On the other hand, for noised images or noised backgrounds and luminance perturbations, the global or local Zernike methods give the best recognition rate.

The local Zernike with Majority Voting approach (LZMV) gives the most uniform results with a recognition rate over 89% for all alterations.

3.4 Comparison of global and local descriptors

Without perturbations (see Table 2 and Figure 9), it is clear that the recognition rate increases when local descriptors are used. If we compare the robustness face to different alterations, the contribution of the local descriptors depends on the perturbations. If noise is added on the image, the use of global descriptors is better than the use of local descriptors. The lowest global descriptors is better than the best local one.

For luminance changes, global Zernike and Fourier-Mellin are better than local descriptors. The main reason is that the key-points detector used in the local approach produce in these cases many key-points that are not relevant for the object recognition. For all the other tested perturbations (noise background, textured background, occlusion), it is favorable to use local descriptors.

4 Conclusion and perspectives

We present in this paper a comparative study of several invariant features in a global or local context for object recognition. Without perturbations, the SIFT algorithm gives better recognition rate except for a training database with few images. With only some views of an object, the LZMV method is therefore more adapted.

With perturbed images the best method depends on the alterations. The SIFT algorithm and Zernike's approaches give the best results. It could be therefore interesting to analyze the image nature to apply the best recognition algorithm. We would have then more than 94.9% of recognition rate (except for textured background). In general, the local Zernike approach with majority voting gives a recognition rate over 89% for all alterations.

Perspectives of this study concern first the fusion of different invariant descriptors such as SIFT and global ones. We would like to implement a multi-descriptors kernel in the SVM method to improve the recognition results. These invariant descriptors will be applied for the navigation of mobile robots in a known context.

References

[1] B. Charroux, "Analyse d'images : coopération d'opérateurs de segmentation guidée par l'interprétation." PhD thesis, university of Paris XI Orsay, 1996.

[2] A. Choksuriwong, H. Laurent, C. Rosenberger, C. Maaoui : "Object recognition using local characterisation and Zernike moments" Lecture Notes in Computer Science, ISBN: 3-540-29032 ,International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS), pages 108-115, 2005.

[3] C.W. Chong, P. Raveendran and R. Mukundan, "A comparative analysis of algorithms for fast computation of Zernike moment", In Pattern Recognition, 36:731-742,2003.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 886-893, 2005.

[5] S. Derrode and F. Ghorbel, "Robust and efficient Fourier-Mellin transform approximations for invariant grey-level image description and reconstruction", In Computer Vision and Image Understanding, 83(1):57-78, 2001.

[6] R. Duda, P. Hart and D. Stork, "Pattern classification", Wiley interscience, 2001.

[7] R.-E. Fan, P.-H. Chen, and C.-J. Lin. Working set selection using the second order information for training SVM. Journal of Machine Learning Research 6, 1889-1918, 2005

[8] J. Friedman, "Another approach to polychotomous classification", Technical report, Department of Statistics, Standford University, 1996.

[9] F. Ghorbel, "A complete invariant description for gray-level images by the harmonic analysis approach", In Pattern Recognition Letters, 15:1043-1051, 1994.

[10] C. Harris and M.J. Stephens, "A combined corner and edge detector", In Proc. ALVEY, 147-152, 1988.

[11] C.W. Hsu and C.J. Lin, "A comparison of methods for multi-class support vector machines", In Proc. IEEE Transactions on Neural Networks, 13:415-425, 2002.

[12] M.K. Hu, "Visual pattern recognition by moment invariants", In Proc. IEEE Transactions on Information Theory, 8:179-187, 1962.

[13] A. Khotanzad and Y. Hua Hong, "Invariant image recognition by Zernike moments", In Proc. IEEE Transaction on Pattern Analysis and Machine Intelligence, 12(5):489-497,1990.

[14] V. Kyrki and J.K. Kamarainen, "Simple gabor feature space for invariant object recognition", Pattern recognition letters, 25(3), 2004.

[15] B. Leibe, E. Seemann and B. Schiele, "Pedestrian detection in crowded scenes", In Proceedings of the EEE International Conference on Computer Vision and Pattern Recognition, 2005.

[16] D. Lowe, "Distinctive image features from scale-invariant key-points", In International Journal fo Computer Vision, 60(2):91-110, 2004.

[17] D. Marr, "Vision", W.H. Freeman, 1982.

[18] K. Mikolajczyk, C. Schmid, "A performance evaluation of local descriptors". In Proc. IEEE Transaction on Pattern Analysis and Machine Intelligence, 27(10):1615-1630, 2005.

[19] B. Park, K. Lee, S. Lee and J. Lee, "Recognition of partially occluded objects using probabilistic arg (attributed relational graph) - based matching", In Proc. Computer Vision and Image Understanding, 90(3):217-241, June 2003.

[20] C. Schmid, R. Mohr and C. Bauckhage, "Evaluation of interest point detectors", In International Journal on Computer Vision, 37(2):151-172, 2000.

[21] D. Sim, O. Kwon and R. Park, ”Object matching algorithm using robust Hausdorff distance measures”, In Proc. IEEE Transactions on Image Processing, 8(3):425-429, March 1999.

[22] M.Stoecker and H.J. Reitboeck, ”A neural network for position invariant pattern recognition combinig spiking neurons with the fourier-tranform”, International journal of neural systems, 7(6), 1997.

[23] M. Turk and A Pentland, ”Eigenfaces for recognition”, In Journal of cognitive neuroscience, 3(1):71-86, 1991.

[24] V. Vapnik, ”Statistical learning theory”, Wiley, 1998.

	GZ	GFM	GH
uniform background	31.8%	**44.1**%	10.3%
noise background	34.9%	**56.5**%	9.9%
textured background	7.5%	**13.3**%	4.2%
black occlusion	**74.7**%	60.4%	43.5%
gray-level occlusion	**71.2**%	66.7%	42.6%
luminance	95.9%	**97**%	13.8%
noise (std=5)	**100**%	98.6%	95.2%
noise (std=10)	**100**%	95.2%	95.2%
noise (std=20)	**100**%	90.2%	91.4%

Table 1: Recognition rate for each descriptor face to alterations. Global Zernike (GZ), Global fourier-Mellin (GFM), Global Hu (GH)

	GZ	LZ	LH	LFM	LZMV	SIFT
10%	60.21%	80.94%	55.85%	62.18%	**86.57%**	53.2%
20%	62.58%	87.31%	60.44%	65.58%	91.75%	**96.5%**
25%	68.9%	94%	70.56%	81.71%	95.9%	**100%**
50%	84.6%	94.1%	74.23%	85.70%	98.2%	**100%**
75%	91.9%	97.7%	80.24%	89.56%	99.1%	**100%**

Table 2: Recognition rate for different sizes of the learning database. Global Zernike (GZ), Local Zernike (LZ), Local Hu (LH), Local Fourier-Mellin (LFM), Local Zernike Majority Voting (LZMV), SIFT

Neighborhood size	7*7	11*11	15*15	19*19
Recognition rate	91.2%	94.1%	98.2%	97.3%

Table 3: Influence of the neighborhood size on the recognition rate for LZMV (50% of the database is used for learning and test)

	GZ	LZ	LH	LFM	LZMV	SIFT
uniform background	31.8%	83.1%	72.13%	81.50%	95.66%	**98.12%**
noise background	34.9%	62.5%	55.10%	58.20%	**94.9%**	91.65%
textured background	7.5%	54.3%	48.22%	51.30%	89.2%	**90.42%**
black occlusion	74.7%	77.9%	65.47%	70.45%	90.4 %	**95.13%**
gray-level occlusion	71.2 %	79.37%	62.41%	69.48%	91.6 %	**94.89%**
luminance	**95.9**%	87.74%	73.14%	81.56%	93.4 %	91.17%
noise (std=5)	**100**%	70.5%	54.11%	56.41%	91.7 %	89.27
noise (std=10)	**100**%	68.3%	48.92%	55.17%	90.6 %	88.89%
noise (std=20)	**100**%	62.2%	45.74%	53.82%	90.2 %	85.46%

Table 4: Robustness face to alterations. Global Zernike (GZ), Local Zernike (LZ), Local Hu (LH), Local Fourier-Mellin (LFM), Local Zernike Majority Voting (LZMV), SIFT

Anant Choksuriwong is a Phd student at ENSI of Bourges (France). He belongs to the laboratory of Vision and Robotics of Bourges in the Signal, Image and Vision research unit. His research interests concern object detection for robotic applications.

Bruno Emile is an assistant professor at IUT of Châteauroux (France). He obtained his Phd from the university of Nice in 1996. He belongs to the laboratory of Vision and Robotics of Bourges in the Signal, Image and Vision research unit. His research interests concern object detection.

Christophe Rosenberger is an assistant professor at ENSI of Bourges (France). He obtained his Phd from the university of Rennes I in 1999. He belongs to the laboratory of Vision and Robotics of Bourges in the Signal, Image and Vision research unit. His research interests concern evaluation of image processing and image understanding.

Helene Laurent is an assistant professor at ENSI of Bourges (France). She obtained his Phd from the university of Nantes in 1998. She belongs to the laboratory of Vision and Robotics of Bourges in the Signal, Image and Vision research unit. Her research interests concern segmentation evaluation.

List of Figures

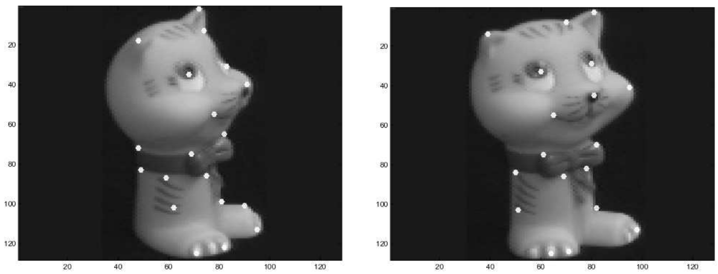


Figure 1: Keypoints detection (using the Harris algorithm) for the same object under different geometric transformations

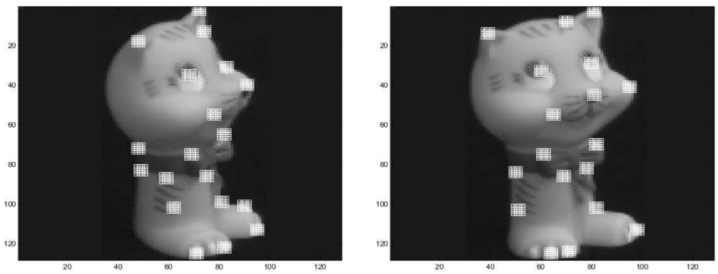


Figure 2: Detected keypoints (using the Harris algorithm) and associated neighborhoods

Figure 3: Three objects in the COIL-100 database presented with different orientations and scales

Figure 4: Alterations examples

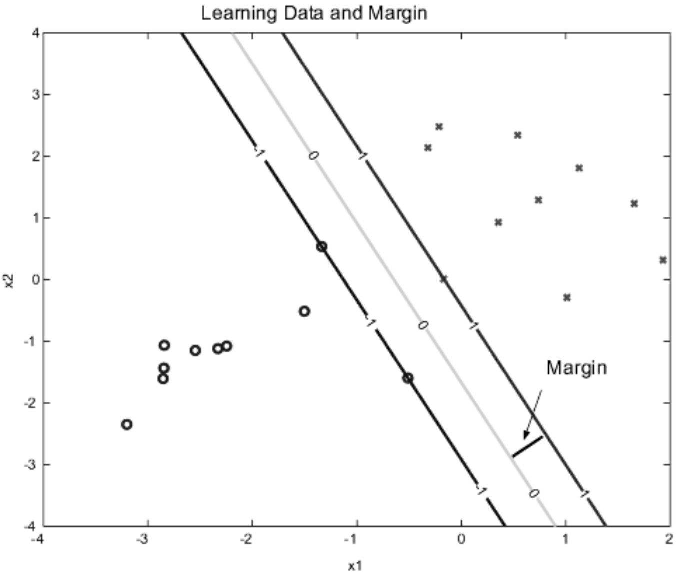
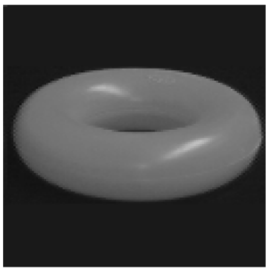


Figure 5: Illustration of the SVM discrimination for linearly separable data.

Object number 24 Object number 34



Object number 47 Object number 50

Figure 6: Four tested objects (24,34,47,50)

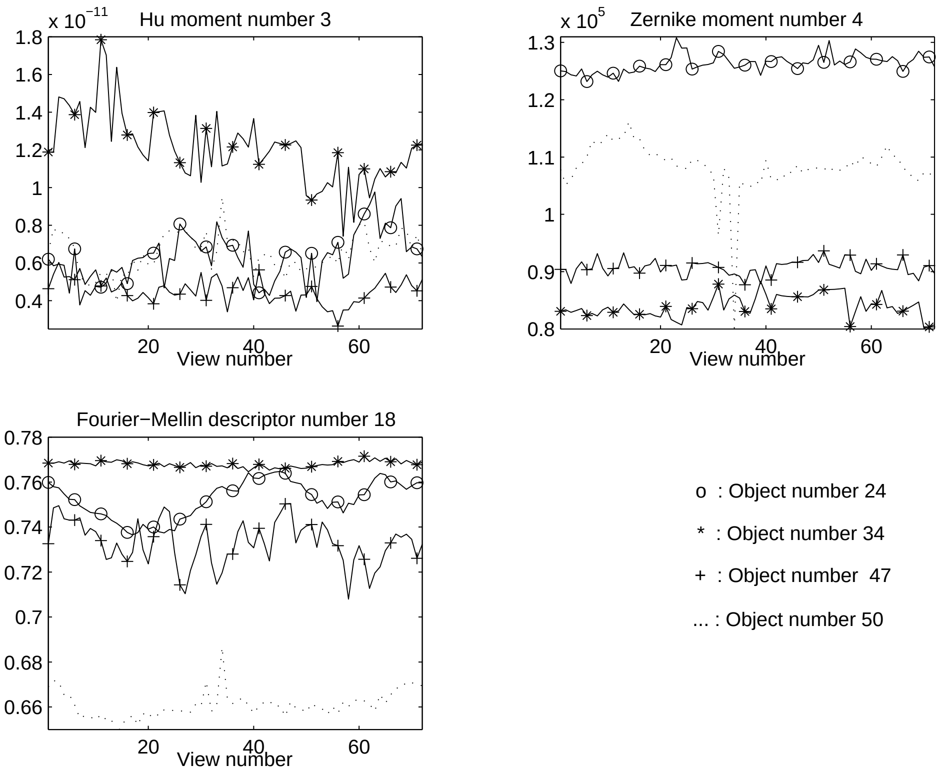


Figure 7: Evolution of three features for the 72 views of four objects of the database without alteration

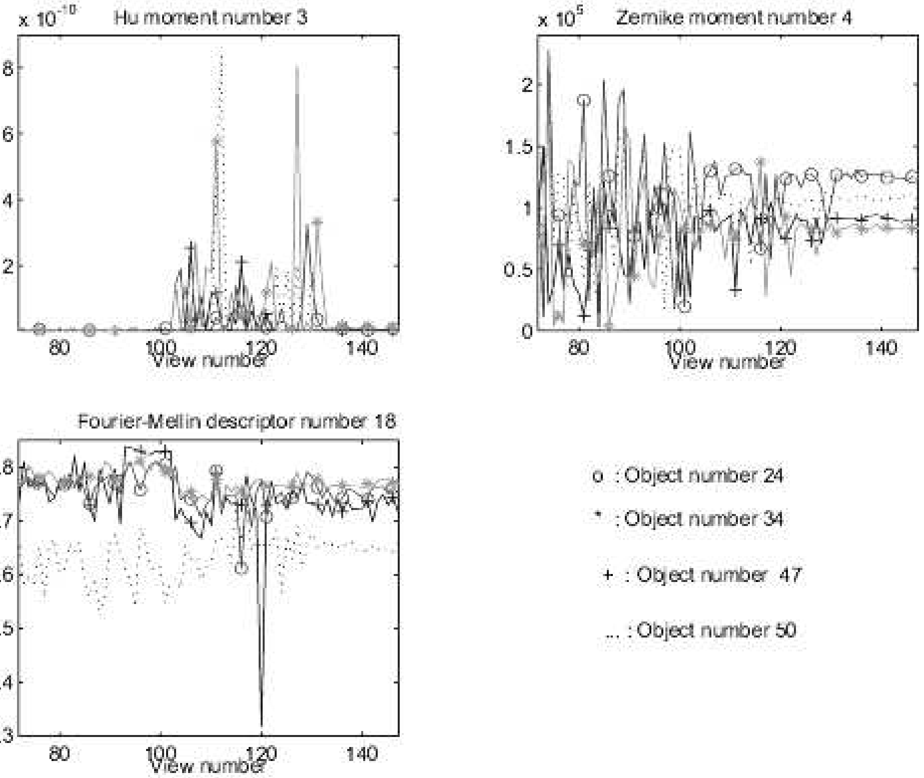


Figure 8: Evolution of three features for four objects of the database considering 75 alterations

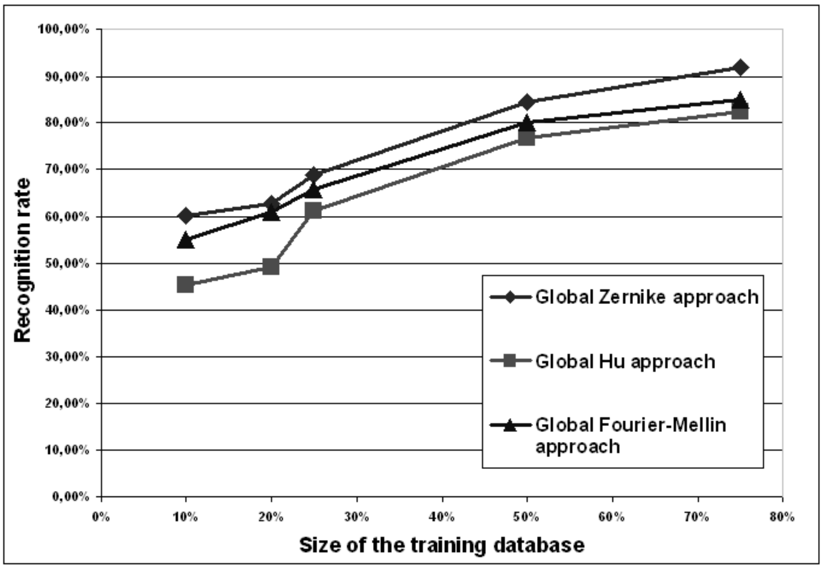


Figure 9: Correct classification rate for different size of the training database.