



Dimension Reduction and Classification Methods for Object Recognition in Vision

Charles Bouveyron · Stéphane Girard · Cordelia Schmid

email: charles.bouveyron@inrialpes.fr

1 Introduction

This paper addresses the challenging task of recognizing and locating objects in natural images. In computer vision, many successful approaches to object recognition use local image descriptors. Such descriptors do not require segmentation, in addition they are robust to partial occlusion and invariant to image transformations (particularly scale changes). Among the existing descriptors, a recent comparison [4] showed that the SIFT descriptor [2] was particularly robust. However, the SIFT descriptor is high-dimensional (typically 128-dimensional) and this penalizes classification. In this paper, we propose to use statistical dimension reduction techniques to obtain a more discriminant representation of data, in order to increase recognition results.

We will first describe the two stages of the recognition process (See Fig. 1), learning and recognition, then we will present experimental results obtained on motorbikes images.

2 Learning the model

Our object model consists of a set of image parts. Learning here is supervised, *i.e.* we manually select our parts from a set of automatically extracted features (*cf.* section 2.1). We then determine a classifier for each part (*cf.* section 2.2).

2.1 Feature extraction

The first step is to detect all characteristic points of the image structure, called interest points, for each training image using the Harris-Laplace operator [3]. This operator is based on the auto-correlation matrix that describes the local structure of the image. Points are therefore distinctive and scale-invariant. From these points we manually select the ones corresponding to our object parts and to the background. We then characterize each point with the SIFT descriptor [2]. Each descriptor is computed by sampling the magnitudes and orientations of the image gradient in an area around the interest point and then building smoothed orientation histograms. For this, the area is divided in 4×4 sub-areas, each of them containing 8 orientations, which leads to a descriptor in dimension $128 = 4 \times 4 \times 8$.

Charles Bouveyron, Stéphane Girard: LMC-IMAG, BP 53, 38041 Grenoble Cedex 9, France
Cordelia Schmid: Inria Rhône-Alpes, 655 av. de l'Europe, 38334 Saint Ismier Cedex, France

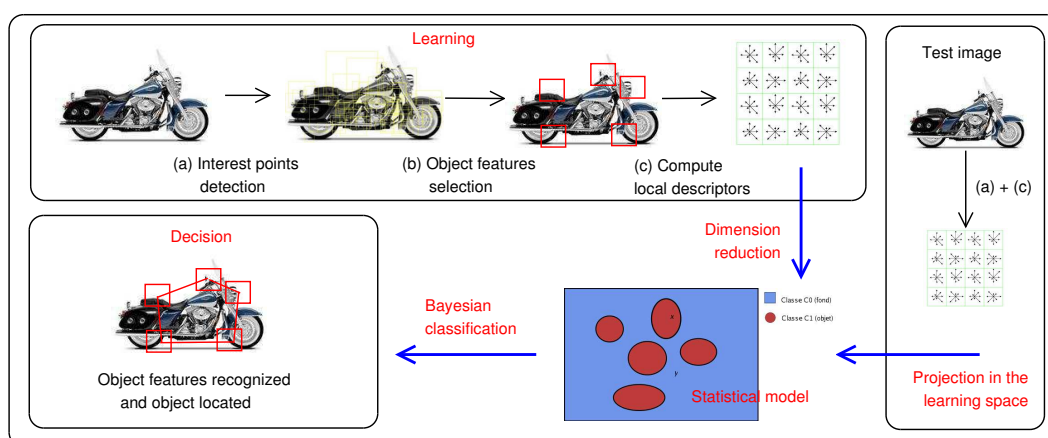


Figure 1: The object recognition process.

2.2 Dimension reduction and classification

The classification is based on a Bayesian classifier. In order to improve performance, we add before a dimension reduction step.

Dimension reduction – A recent technique called PCA-SIFT [1] showed that dimension reduction using Principal Component Analysis (PCA) increases recognition results in an image retrieval application. The authors demonstrate that PCA-SIFT based descriptors are more robust, discriminant and more compact than standard SIFT descriptor. They also show that matching based on this modified descriptor increases results and speed.

We use dimension reduction to obtain a more discriminant representation of the descriptors and show that this improves recognition results. More specifically, we compare three methods: two linear methods (PCA and LDA) and one nonlinear method (LLE). Our approach is in a supervised learning context so that we can use Discriminant Analysis method; LDA gives the $(k - 1)^1$ discriminant axes maximizing the inter-class variance and minimizing the intra-class variance. The LLE method, introduced recently by Roweis *et al.* [5], finds an embedding that preserves the local geometry in a linear neighborhood of each data point in the original space. We will also show that the choice of a dimension reduction technique has to tie up with the choice of a classification method.

Statistical model – We consider the statistical model made of 6 classes, the background and 5 motorbike parts. For the background class, the density of the feature x is supposed to be Gaussian with mean m_0 and covariance matrix Σ_0 . For each of the 5 motorbike parts, the density f_i of x is also supposed to be Gaussian with mean m_i and covariance matrix Σ_i , $i = 1, \dots, 5$.

We learn on the learning set the parameters m_i and Σ_i for each classes $i = 0, \dots, 5$. Therefore, during the recognition, we can classify a test descriptor x using a Bayesian classification. However, the Bayesian classification in this full version (R_3) requires the estimation of the full covariance matrix Σ_i for each class, which is difficult in high dimension. Therefore, we have also used two approximations: (R_2) $\forall i$, $\Sigma_i = \Sigma$ and (R_1) $\forall i$, $\Sigma_i = s_i Id$, where Σ is a common covariance matrix.

¹ k is the number of classes.

3 Recognition

Recognition consists in deciding whether test images (different from the training set) contain the object of interest or not. Like in the training stage, we compute the SIFT descriptors for the test images. We obtain an image representation which we can compare with the ones obtained in the learning step. Then, test descriptors are projected in the learning space. We have to decide now if a given descriptor belongs to the object class or not. Many classification methods exist, but we have chosen to use the Bayesian classification methods because they are statistically based. The Bayesian rule consists in affecting the point x to the class C_i with the *maximal a posteriori probability* $P(C_i|x) = p_i f_i(x) / \sum_{j=0}^5 p_j f_j(x)$, where p_i is the proportion of the class $i = 1, \dots, 5$ in the learning set.

4 Experimental results

4.1 Dataset and protocol

For our experimentation, we chose to work with motorbikes images. We computed the descriptors for a set of 200 images and we selected the ones corresponding to 5 motorbike features: headlight, front wheel, rear wheel, seat and handlebars (See Fig. 1-b). We obtained the two following datasets: *learning set* and *test set*. The *learning set* is constituted by 150 descriptors of each of the 5 characteristic elements and 750 descriptors of the background. In the same manner, the *test set* is constituted by 50 descriptors of each characteristic element and 250 descriptors of the background.

For each dimension $d = 1, \dots, 128$, we reduced the dimension of the learning set using PCA, LDA² and LLE, then we learned the density of each class following the statistical model. Next, we projected the test descriptors in the learning space and computed the *a posteriori* probability for each class using the three decision rules (R_1), (R_2) and (R_3). Finally, we affected each test descriptor to one class according to (R_1), (R_2) and (R_3).

4.2 Results and discussion

The classification results are presented in Figure 2; the two graphs show the classification rate (y axis) with respect to the dimension reduction (x axis). On one hand, they show that reduction dimension by PCA increases recognition results, particularly with the full bayesian rule (R_3). The better results are obtained in 40 dimensions and computing times are large. On the other hand, LDA increases also recognition results according to all classification rules. LDA gives a very discriminant representation in $(k - 1)$ dimensions (here $k = 6$) which allows to use efficiently the (R_2) rule. Consequently, classification results are good and computing times are small.

By lack of space, we do not present results obtained using the LLE method. Results are worse than *maximum* of the other methods (quantitatively, the classification rate is 4% less). LLE is a non-linear method for dimension reduction, but it is penalized by a difficult parametrization. In addition, computing times are very long. In this paper, in order to carry out realistic experiments, we had to consider a large background class which is the main reason for not so satisfying classification rates.

²For LDA, we take first the $(k - 1)$ discriminant axes, then we complete with PCA axes.

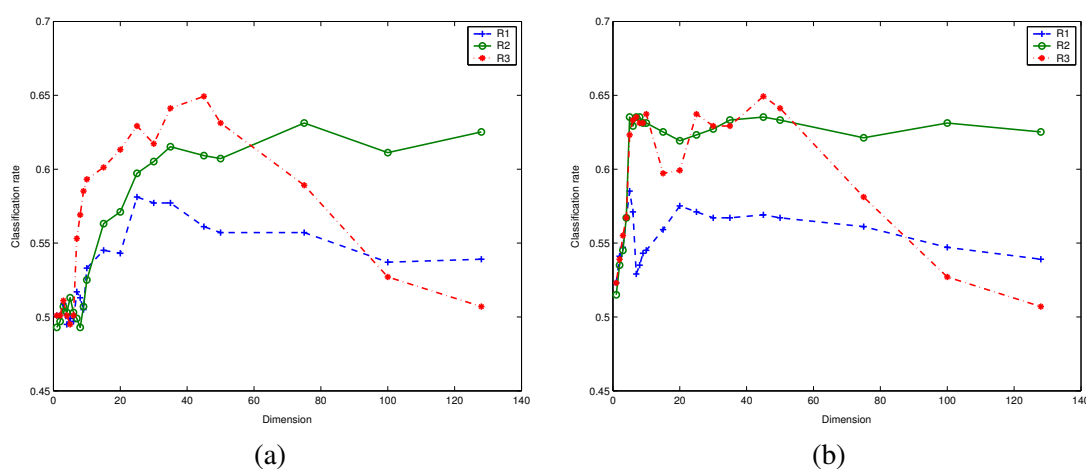


Figure 2: Classification results with (a) PCA and (b) LDA using three classification rules.

5 Further work

In order to increase further the recognition results, we could choose a different statistical model based on a Gaussian mixture. In addition, we look for a dimension reduction for each class using PCA or LDA. Another extension to the present work is to include some spatial information or neighborhood relationship, which could be modeled using bayesian networks or hidden Markov models.

6 Acknowledgments

This work was supported by the French Department of Research through the *ACI Masse de données*.

Bibliography

- [1] Y. Ke and R. Sukthankar (2004), "PCA-SIFT: A more distinctive representation for a local image descriptors", Techn. Rep., *INTEL Research*.
- [2] D. Lowe (2004), "Distinctive image features from scale-invariant keypoints", to appear in the *International Journal of Computer Vision*.
- [3] K. Mikolajczyk and C. Schmid (2001), "Indexing based on scale invariant interest points", *International Conference on Computer Vision*, pp. 525-531.
- [4] K. Mikolajczk and C. Schmid (2003), "A performance evaluation of local descriptors", *IEEE Conference on Computer Vision and Pattern Recognition*.
- [5] S. Roweis and L. Saul (2000), "Nonlinear reduction dimension by Locally Linear Embedding", *Science*, 290, pp. 2323-2326.