Dimension Reduction and Classification Methods for Object Recognition in Vision

Charles BOUVEYRON

LMC-IMAG & INRIA Rhône-Alpes
Framework

Goal: approach the capability of training and visual recognition of the human eye.

- locate an object on natural images with complex background,
- recognize the object class (require a learning stage).
Framework

Goal: approach the capability of training and visual recognition of the human eye.
- locate an object on natural images with complex background,
- recognize the object class (require a learning stage).

Tools:
- local descriptors of the image,
- statistical methods:
  - dimension reduction,
  - Bayesian classification.
Some applications

- Object recognition and localisation in images or videos,
- Automatic localisation of tumors in MRI images,
- Identification of peoples for access control,
- Interpretation of aerial images (applications in biology),
- Content-based image retrieval (like Google Image Search).
Outline of the talk

1 – Object recognition:
   - Learning the model
   - Recognition
2 – Why to reduce the dimension?
3 – Previous works
4 – Our approach
5 – Experimental results
6 – Conclusion and further work
1 – Object recognition

**Learning:**
- detection and description,
- dimension reduction,
- learning the densities.

**Recognition:**
- detection and description,
- projection in learning space,
- classification using Bayesian rules.
The Harris-Laplace operator allows to detect characteristic points of the image structure with their characteristic scale.

Fig. 1 – Interest points detection using the Harris-Laplace operator.
1 – Learning: detection

The Harris-Laplace operator allows to detect characteristic points of the image structure with their characteristic scale.

Harris filter:

- Based on the autocorrelation matrix $A$ computed on a neighborhood of the keypoint $x$:

$$A = \begin{pmatrix}
\sum I_x I_x & \sum I_x I_y \\
\sum I_y I_x & \sum I_y I_y
\end{pmatrix}.$$ 

- Eigenvalues of $A$:
  - 2 strong eigenvalues $\rightarrow x$ is an interest point,
  - 1 strong eigenvalue $\rightarrow x$ is a contour,
  - 0 strong eigenvalue $\rightarrow x$ is in an homogeneous region.
1 – Learning: detection

- The Harris-Laplace operator allows to detect characteristic points of the image structure with their characteristic scale.

- Laplace operator: detection of the characteristic scale

- computation of the Laplacian for several scales around the keypoints,

- characteristic scale: $s^* = \arg\max_s \{\Delta_{x,y}(s)\}$
1 – Learning: description

- **SIFT descriptor** [Low04] : a robust descriptor [MS03b]
  - based on the gradient values in 8 directions of $4 \times 4$ patches around the interest point
  - descriptor 128-dimensional.
- invariant with luminosity changes (and geometric transformations).
We consider a statistical model made of 6 classes:

1 background class modeled by a multidimensional Gaussian density \( f_0 \),

5 motorbike parts. Each of them is also modeled by a Gaussian density \( f_i, i = 1, ..., 5 \).

Fig. 2 – Statistical model for 4 object classes and 1 background class.
1 – Recognition: classification

We used the three following classification methods:

- \((R_3)\) Bayesian classification (require to estimate the full covariance matrix \(\Sigma_i\) for each class \(i\)),
- \((R_2)\) Naive Bayesian classification \((\forall i = 1, \ldots, 5, \Sigma_i = \Sigma)\),
- \((R_1)\) Independence assumption \((\forall i = 1, \ldots, 5, \Sigma_i\) diagonal matrix).
1 – Recognition: classification

We used the three following classification methods:

- \((R_3)\) Bayesian classification (require to estimate the full covariance matrix \(\Sigma_i\) for each class \(i\)),
- \((R_2)\) Naive Bayesian classification \((\forall i = 1, \ldots, 5, \Sigma_i = \Sigma)\),
- \((R_1)\) Independence assumption \((\forall i = 1, \ldots, 5, \Sigma_i\) diagonal matrix).

The Bayesian classification rule consists in affecting the point \(x\) to the class \(C_i\) with the maximal a posteriori probability (MAP) \(P(C_i|x)\):

\[
x \in C_i, \text{ if } i = \arg\max_{j=0,\ldots,5} \{p_j f_j(x)\}
\]
2 – Why to reduce the dimension?

- Dimension reduction:
  - find a new feature space,
  - with a dimension significantly smaller,
  - which contains a large part of the original information.
2 – Why to reduce the dimension?

- Dimension reduction:
  - find a new feature space,
  - with a dimension significantly smaller,
  - which contains a large part of the original information.

- Why?
  - find a more discriminant representation of the data.
  - overcome the *curse of the dimensionality* (particularly for classification methods).
  - this also allows to denoise the data.
3 – Previous works

Eigen-Images \textsuperscript{[TP91]}:
- Descriptor based on grayscale histogram.
- Dimension reduction using PCA for face recognition.
3 – Previous works

- **Eigen-Images [TP91]**:
  - Descriptor based on grayscale histogram.
  - Dimension reduction using PCA for face recognition.

- **PCA-SIFT [KS03]**: show that dimension reduction increases recognition results.
  - SIFT in dimension 3042 then apply PCA to obtain a modified descriptor in dimension 20 (empirically computed)
  - Results (in image retrieval application):
    - using SIFT in dimension 3042: 40% corrects,
    - using PCA-SIFT in dimension 20: 90% corrects.
4 – Our approach: goals

We will compare the following dimension reduction methods:

- Principal Component Analysis (PCA),
- Discriminant Analysis (LDA),
- Nonlinear method: Locally Linear Embedding (LLE).
4 – Our approach: goals

- We will compare the following dimension reduction methods:
  - Principal Component Analysis (PCA),
  - Discriminant Analysis (LDA),
  - Nonlinear method: Locally Linear Embedding (LLE).

- Main goals:
  - find a more discriminant data representation in order to increase recognition results,
  - accelerate the computing times (in order to be used in real-time applications),
4 – Our approach: PCA, LDA and LLE

PCA: Principal Component Analysis

find a new data representation in a subspace of smaller size and maximizing the variance.
4 – Our approach: PCA, LDA and LLE

- **PCA: Principal Component Analysis**
  - find a new data representation in a subspace of smaller size and maximizing the variance.

- **LDA: Linear Discriminant Analysis**
  - find the $(k - 1)$ discriminant axes maximizing the *interclass* variance and minimizing the *intraclass* variance.
4 – Our approach: PCA, LDA and LLE

- PCA: Principal Component Analysis
  - find a new data representation in a subspace of smaller size and maximizing the variance.

- LDA: Linear Discriminant Analysis
  - find the \((k - 1)\) discriminant axes maximizing the interclass variance and minimizing the intraclass variance.

- LLE: Locally Linear Embedding [SR01]
  - find an embedding which preserves the local geometry in a linear neighborhood of each point in the original space.
4 – Our approach: data

- We chose to work with a set of 200 motorbike images,
- We are interested by 5 features of the object.

(a) Examples of image database:

(b) The 5 object features:
5 – Results: projection using LDA

Fig. 3 – Supervised dimension reduction using LDA without background.
5 – Results: protocol

For each dimension \( d = 1, \ldots, 128 \):

- We reduced the dimension of the learning set using PCA, LDA and LLE,
- then, we learned the densities of each class following the model.
- Next, we projected test descriptors in the learning space,
- and computed the \textit{a posteriori} probability for each class according to \((R_1),(R_2)\) and \((R_3)\).
- Finally, affected each test descriptor to the class with the maximum \textit{a posteriori} probability.
5 – Results: PCA vs LDA

Fig. 4 – Classification results with (a) PCA, (b) LDA using the three classifiers.
5 – Results: discussion

- Dimension reduction by PCA:
  - increases recognition results, particularly with the rule $(R_3)$,
  - the better results are obtained in 45 dimensions.
5 – Results: discussion

- Dimension reduction by PCA:
  - increases recognition results, particularly with the rule \( R_3 \),
  - the better results are obtained in 45 dimensions.

- Dimension reduction by LDA:
  - also increases recognition results, particularly with the rule \( R_2 \) in low dimension,
  - very good results are obtained in fixed dimension: \( (k - 1) \) dimensions (here \( k = 6 \)),
  - computing times are smaller.
5 – Results: discussion

- Dimension reduction by PCA:
  - increases recognition results, particularly with the rule \((R_3)\),
  - the better results are obtained in 45 dimensions.

- Dimension reduction by LDA:
  - also increases recognition results, particularly with the rule \((R_2)\) in low dimension,
  - very good results are obtained in fixed dimension: \((k - 1)\) dimensions (here \(k = 6\)),
  - computing times are smaller.

- Dimension reduction by LLE:
  - recognition results are worse than other methods.
6 – Conclusion

Our work highlights that:

- dimension reduction increases recognition results,
- which allows to use efficiently complex classification methods
- particularly, LDA gives a very discriminant representation in low and fixed dimension,
- this dimension is known in advance: \( d = (k - 1) \).
6 – Further work

- Statistical model:
  - use a model based on Gaussian mixture (object and background),
  - find a method for automatically detecting object features.
6 – Further work

- Statistical model:
  - use a model based on Gaussian mixture (object and background),
  - find a method for automatically detecting object features.

- Dimension reduction:
  - use kernel based methods (KPCA and KLDA),
  - use dimension reduction for each class.
6 – Further work

- Statistical model:
  - use a model based on Gaussian mixture (object and background),
  - find a method for automatically detecting object features.

- Dimension reduction:
  - use kernel based methods (KPCA and KLDA),
  - use dimension reduction for each class.

- Include other informations:
  - model spatial information or neighborhood relationship using Bayesian network or hidden Markov models.
6 – Dimension reduction for each class

The idea:

- The idea is to reduce the dimension for each class independently,
- and, for a test descriptor $x$, compute the \textit{a posteriori probability} $P(C_i|x)$ in each space,
- finally, affect $x$ to the class with the MAP.
6 – Dimension reduction for each class

The idea:

- The idea is to reduce the dimension for each class independently,
- and, for a test descriptor \( x \), compute the *a posteriori probability* \( P(C_i|x) \) in each space,
- finally, affect \( x \) to the class with the MAP.

Using PCA or LLE:

- reduce the dimension in \( k \) different spaces,
The idea:

The idea is to reduce the dimension for each class independently,
and, for a test descriptor $x$, compute the *a posteriori probability* $P(C_i | x)$ in each space,
finally, affect $x$ to the class with the MAP.

Using PCA or LLE:
reduce the dimension in $k$ different spaces,

Using LDA:
reduce the dimension in $(k - 1)$ different spaces,
in each of $(k - 1)$ spaces, compare the MAP of one object feature class to the background class.
6 – Dimension reduction for each class

Reduce the dimension in $(5 - 1)$ different spaces,

Projection of test data of the class $n^1$ in different spaces,

Compute the MAP for each test descriptor and affect to the best class.
References


