Robust Object Detection in Complex Backgrounds using ICA Compression

Vicente M.A., Fernández C., Reinoso O., Pérez C.
Department of Industrial System Engineering
Miguel Hernandez University, Avda. del Ferrocarril, s/n
03202 Elche, Spain
suni@umh.es

ABSTRACT

This paper describes a method for recognizing real three-dimensional objects in non-controlled backgrounds using independent component analysis to eliminate redundant image information present in each object image. The proposed method is a two-step process that allows a coarse color-based detection and an exact localization using shape information. The paper describes an efficient implementation, making this approach suitable for real-time applications.

Keywords

object recognition, eigenspace, color histograms, lateral histograms, independent component analysis, principal component analysis.

1. INTRODUCTION

Appearance-based recognition systems have appeared as a powerful alternative to model-based approaches that use 3D geometry when it is difficult to obtain geometrical models of the objects [Bie95a] [Vic02b]. The appearance of a 3D object in a 2D image depends on its shape, its color, its pose in the global scene, its reflectance properties and the sensor and illumination characteristics. The earlier appearance-based recognition systems were focused on the holistic approach: an object image may be considered as a vector of pixels where the value of each entry in the vector is the grayscale (or color) value of the corresponding pixel. For example, a NxN image may be unwrapped and treated as a vector of length N^2.

WSCG POSTERS proceedings
WSCG'2004, February 2-6, 2004, Plzen, Czech Republic.

The image is said to sit in N-dimensional space, which is considered to be the original space of the image.

In these systems the whole object image is projected to a lower dimensional space using different techniques or subspaces, the most frequently used is the subspace created by the eigenvectors of the covariance matrix of the training images (Principal Component Analysis) ([Mur95a], [Turk91a] [Nay96a]), another common subspace is the created by the basis vectors obtained using Linear Discriminant Analysis or the subspaces computed by Independent Component Analysis [Hyv01a]. These methods have been successfully used in different applications (face recognition, robot positioning, bin-picking, etc.) but there are some difficulties still unresolved, like partial object occlusions or complex backgrounds.

This paper proposes a two-step process that allows robust object recognition in complex backgrounds based on ICA representation. The system basically consists of a coarse detection (color-based) and a refinement phase (shape-based), each containing a selection of ICA features and the use of well-known pattern recognition techniques.

In Section 2, we review eigenspace analysis versus the independent component analysis, and justify the choice of ICA. Section 3 proposes the two-step recognition approach. Section 4 shows some experimental results and evaluates the approach, and finally Section 5 concludes this paper.

2. EIGENSPACE VS ICA REPRESENTATION

2.1.1 Eigenspace approach

Eigenspace (PCA) is calculated by finding the eigenvectors of the covariance matrix created from the set of training images. The eigenvectors corresponding to non-zero eigenvalues of the covariance matrix represent an orthonormal basis that projects the original images in the N-dimensional space.

This technique of principal component analysis enables us to create and use a reduced set of variables. A reduced set (the classes obtained from the training images) is much easier to analyze and interpret than the original variables (the training images themselves).

In the original subspace method one image is chosen from each object to create the data matrix, and the object recognition system is represented by just one subspace. In advanced methods, one subspace is obtained for each object, so there are as many subspaces as objects to recognize. Then, for each subspace the data matrix is created from different images taken from the same object [Vic02a].

2.1.2 ICA representation

The independent component analysis of an N dimensional random vector is a linear transform that minimizes the statistical dependence between its components. This analysis has a great number of applications such as data analysis and compression, blind source separation, blind deconvolution, denoising, etc.

If the random vector we wish to represent through ICA has no noise and is zero-centered, the ICA model can be expressed as:

$$\vec{x} = A \cdot \vec{s} \tag{6}$$

where $\vec{x}$ is the random vector representing our data, $\vec{s}$ is the random vector of independent components with dimension $M \leq N$, and **A** is the mixture matrix. The pseudoinverse of **A**, represented by **W**, is called the projection matrix and it provides an alternative representation of the ICA model:

$$W \cdot \vec{x} = \vec{s} \tag{7}$$

Various objective functions have been proposed for the estimation of the projection matrix such as nongaussianity, likelihood, mutual information, and tensorial methods [Hyv01a]. In our method we employ the FastICA method which estimates the whole decomposition by minimizing mutual information, and estimates the individual independent components as projection pursuit directions.

The experimental results will show how the recognition rates of ICA clearly outperfom the PCA ones.

3. TWO-STEP RECOGNITION PROCESS

In our approach, each object is represented by three different appearance-based feature vectors: an *ica-color-histogram* (containing color information) and two *ica-lateral-histograms* (containing spatial information or shape information).

We called *ica-color-histogram* or $\vec{H}c$ *feature* the vector obtained after compressing the common RGB color histogram from the original object image subtracting the blue background. There are two *ica-lateral-histograms*, the *horizontal* one and the *vertical* one [Dav90a], which are calculated by summing the grey levels of the pixels in each of the columns and rows of the image respectively:

$$H(x) = \sum_{y=1}^{Y} I(x,y) \quad V(y) = \sum_{x=1}^{X} I(x,y) \tag{8}$$

The implemented recognition system basically consists of a coarse detection step (color-based) and a refinement phase (shape-based), each containing the selection of ICA features described before and the use of well-known pattern recognition techniques.

3.1.1 Coarse detection

During the initial localization, the complex scene is searched looking for the desired object applying small windows to the original image. We choose the window sizes in order to be smaller than the size of the training objects. The set of scanning windows is given as:

$$P = [P_1, P_2, \ldots P_K] \tag{9}$$

where P_j is an image matrix of window size and K is the total number of windows, which value depends on the window size and the scan step. At the same time we also obtain the uncompressed color histogram (3x256 bins) from each window, which gives us the histogram set H:

$$H = \left[\vec{H}_1, \vec{H}_2, \ldots \vec{H}_K\right] \tag{10}$$

After that, the dimension of each histogram is reduced by computing the ica-color-histogram, $\vec{H}c_j$ of each $\vec{H}_j$ from the set, projecting them over the ICA subspace obtained during the training of the system:

$$\vec{H}c_j = \vec{H}_j \cdot S \tag{11}$$

where S is the matrix of the independent components $\vec{s}_i$:

$$S = \left[\vec{s}_1, \vec{s}_2, \ldots \vec{s}_N\right] \tag{12}$$

The number of independent components N equals the number of objects used in the training step.

The next step of the process deals with the comparison of the $\vec{H}c$ feature of each window with the $\vec{H}c$ feature of the model. The goal is to find the image window j most similar to the desired object. $\vec{d}$ is the vector of distance measures:

$$\vec{d} = \left[d_1, d_2, \ldots d_K\right] \tag{13}$$

where

$$d_j = \mathrm{dist}(\vec{H}c_j - \vec{H}c_{model}) \tag{14}$$

We use two types of classifiers: the k-nearest neighbor rule [Dev82a] with L2-distance measure and the local minima, in order to select the *positive windows* or *areas of interest (AOI)*, in other words, the possible regions containing the same color distribution than the object image. Let A be the AOI set:

$$A = \left[\vec{A}_1, \vec{A}_2, \ldots \vec{A}_M\right] \tag{15}$$

The size of the AOI set is much smaller than the size of the histogram set so the dimensional reduction is achieved:

$$M << K \tag{16}$$

The value of M depends on the classifier used.

3.1.2 Refinement

Once the search has been reduced to just *M* AOI's, the adjoining AOI's are grouped in bigger windows. After that, the two ica-lateral-histograms are used for matching the exact localization of the object in the image. Let H_{lat} be the histogram lateral set:

$$H_{lat} = \left[\vec{H}_{lat1}, \vec{H}_{lat2}, \ldots \vec{H}_{latK}\right] \tag{17}$$

where each component is obtained as the concatenation of the vertical and horizontal histograms:

$$\vec{H}_{latj} = \left[\vec{H}_{Vj} \,|\, \vec{H}_{Hj}\right] \tag{18}$$

In this step, the ica-lateral-histogram models have been previously computed, then one ica-subspace is generated for each object in order to achieve better recognition rates [Vic02a]. Besides, different views from the same object are needed to compute the feature vectors. This process would be computationally expensive using the original set of windows, but not using just the AOI's obtained before.

4. EXPERIMENTAL RESULTS

For the training experiments the novel ETH-80 database [Lei03a] has been used. It contains 80 objects from 8 categories. Each object is represented by 41 views spaced evenly over the upper viewing hemisphere. This allows analyzing the performance of different recognition methods not only from a 1D circle or a few canonical viewpoints, but from multiple viewing positions.

The system was trained to recognize 8 objects from different categories. For the color features just the main image was used to compute the color-histogram models for each object. To generate the lateral-histogram models 10 views were employed, so 20 shaped features for each object vectors are required at most to achieve the refined localization.

In Table 1 we show the training results of the coarse detection step for a medium size window. The values of the recognition rates after the refined localization step are shown on Table 2. The more robust approach is ICA compression using one subspace per object and k-nearest neighbor rule.

Method	Recognition rate
1-NN L2	37.5%
5-NN L2	51.1%
10-NN L2	87.2%
20-NN L2	89.8%
50-NN L2	100%
Local minima C4	88%

Table 1. Training results of the coarse detection step for a medium size window.

Method	Recognition rate	Computational cost
1-NN L2, ICA [1]	98%	10
1-NN L2, PCA	81%	10
5-NN L2, ICA	100%	15
5-NN L2, PCA	93%	15
1-NN L2, ICA*	88.25%	1
1-NN L2, PCA*	73%	1
MLP, ICA*	97%	1
MLP, PCA*	91%	1

Table 2. Final recognition rates after the refined localization step.

5. CONCLUSIONS

The experimental results have shown the robustness of our approach. The method allows locating the target object in any complex scene.

The use of a two step process reduces computation time and allows real time applications. Apart from that, it has been proven that the ICA compression outperforms PCA for feature extraction, leading to higher recognition rates.

Future work will be focused on the categorization capabilities of the present approach (i.e whether is it possible to recognize untrained objects of a certain category).

[1] ICA and PCA use one subspace per object. In ICA* and PCA* just one subspace is used for the whole database.

6. ACKNOWLEDGMENTS

This work is supported by Generalitat Valenciana grant CTDIA/2002/03, “Automatización de tareas complejas en robótica móvil mediante sistemas inteligentes.”

7. REFERENCES

[Bie95a] Biederman, I. Visual object recognition. In S. F. Kosslyn and D. N. Osherson (Eds.). An Invitation to Cognitive Science, 2nd edition, Volume 2., Visual Cognition. MIT Press. Chapter 4, pp. 121-165. 1995.

[Dav90a] Davies E.R. Machine Vision: Theory, Algorithms, Practicalities. Academic Press. 1990.

[Dev82a] Devijver P.A. , Kittler J.V., Pattern Recognition. A statistical Approach, Prentice Hall-Englewood Cliffs.

[Hyv01a] Hyvärinen A., Karhunen J., Oja E., Independent Component Analysis. John Wiley and Sons. 2001. http://www.cis.hut.fi/projects/ica/fastica/

[Lei03a] Leibe B. and Schiele B., Analyzing Appearance and Contour Based Methods for Object Categorization. International Conference on Computer Vision and Pattern Recognition (CVPR'03), Madison, Wisconsin, June 2003.

[Mur95a] Murase H. and Nayar S.K. Visual Learning and Recognition of 3-D Objects from Appearance. International Journal od Computer Vision, 14:5-24, Januray 1995.

[Nay96a] Nayar S.H., Murase H., and Nene S.A., Parametric Appearance Representation, in Early Visual Learning, edited by S. K. Nayar and T. Poggio, Oxford University Press, February 1996.

[Tur91a] Turk M. and Pentland A., Eigenfaces for Recognition, Journal of Cognitive Neuroscience, vol. 3, no. 1, pp.71-86,1991.

[Vic02a] Vicente M.A., Reinoso O., García N., Sabater J.M., Fernández C., Jiménez L., Experiments with COIL database using PCA-based Object Recognition Techniques, ISRA’2002 3rd International Symposium on Robotics and Automation, 2002.

[Vic02b] Vicente M.A., Gil P., Reinoso O., Torres F., Objects recognition by means of projective invariants considering corner-points. Journal of WSCG, Vol.10, No.1-3,UNION Agency-Science Press, ISSN 1213-6980, 2000.