APPEARANCE-BASED RECOGNITION WITH VARYING PATTERNS

M.A. Vicente, A. Gil, C. Fernández, L. Jiménez and R. Ñeco
Industrial System Engineering Dpt., Miguel Hernández University
Elche, Alicante, (Spain)
suni@umh.es

Abstract

An appearance-based visual recognition system for discrimination between free-form objects is presented. The system has been developed in order to be run in a scene where the number of patterns is considerably large and they are not stable, as new patterns are introduced and others are eliminated along the time. In appearance-based systems, the learning step requires the acquisition of large image sets and an intensive computational cost due to the feature extractors. Generally, learning is performed off-line so computing time is not a problem. However, with varying patterns, learning has to be performed several times and under such circumstances the training time is relevant. The proposed method uses random projection for dimensionality reduction and a k-NN classifier to identify the object correctly. We show that despite the computational simplicity of random projection, it allows to keep enough information of the original appearance-based vectors in order to distinguish between the patterns. Experimental results obtained with several image databases show the validity of the approach.

Keywords
object recognition, feature extraction, dimensionality reduction, random projection, computer vision

1. INTRODUCTION

Appearance-based recognition approaches are a powerful alternative to model-based techniques when it is difficult to obtain geometrical models of the objects and when the images have non-controlled backgrounds. Such systems are able to manage changes in illumination conditions, shape, pose and reflectance [14] and even to handle translation and partial occlusions [17].

Appearance-based object recognition methods work on high dimensional spaces called *image spaces*, where each dimension corresponds to a pixel of the object image. The number of dimensions is thus extremely large, even when low resolution images are used. Besides, it is common to use multiple channels of information in order to improve the recognition rates [2]. Examples of frequently used channels include the grayscale values of the original images; color information (using histograms different color spaces can be used like RGB, YUV,...); binary or grayscale border images (again, different border extractors may be used); texture information, etc. The main idea is to use as many sources of information as possible in order to easily distinguish the different objects: some of them may be distinguishable according to the grayscale pixel values, others may require color information to be distinguished, etc. Under such circumstances, the number of dimensions of the image space is even larger (roughly the number of channels times the number of pixels) and the direct use of a classifier becomes computationally unfeasible, as the number of attributes for the classifier equals the number of dimensions of the image space.

Different techniques have been used to reduce the dimensionality of the data, allowing the classifier to work on a smaller number of attributes. The goal is to discard the irrelevant or redundant data and to keep as much information as possible. The most popular technique for dimensionality reduction purposes is Principal Component Analysis (PCA) [11]. This technique has been used both in general object recognition applied to robotic tasks [14][15] and in face recognition applications [10][21]. Another widely used technique is Linear Discriminant Analysis (LDA) which takes into account class memberships in computing the subspace. Some applications in face recognition can be found in [3] and [20]. Dimensionality reduction is in fact a feature extraction process, as new features are extracted from the original attributes.

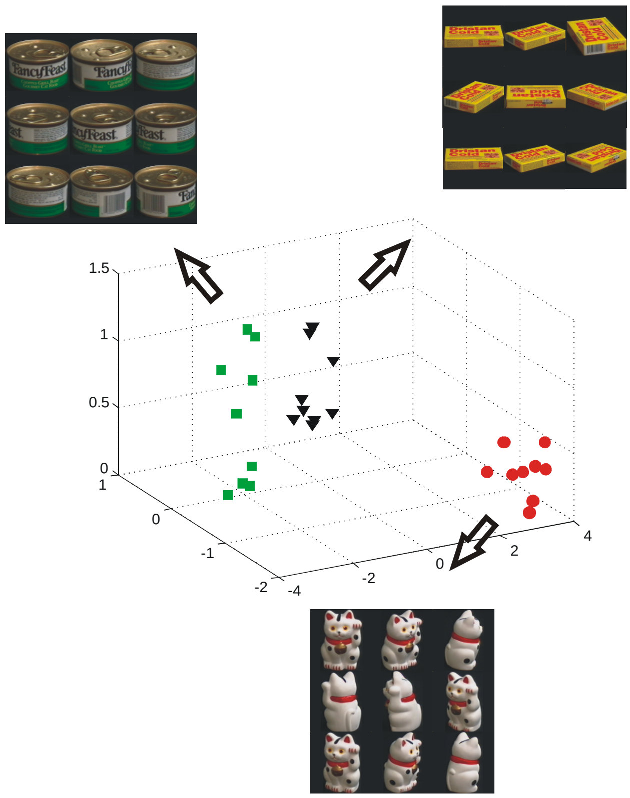


Fig. 1. Concept of the image space: a two-dimensional image may be viewed as a vector in such a very high dimensional space.

Such feature extracction processes require an intensive computational cost. Generally, this processes are performed off-line, so computing time is not a problem. However, with varying patterns, feature extraction has to be performed several times and under such circumstances the computing time is relevant. The presented technique offers a simple and computationally efficient solution to this problem.

The proposed method uses random projection for dimensionality reduction and a k-NN classifier to identify the object correctly. We show that despite the computational simplicity of random projection, it allows to keep enough information of the original appearance-based vectors in order to distinguish between the patterns. Besides, this feature extraction method has a great advantage over the classical techniques used, mainly PCA and LDA, due to its independence from the set of patterns available in each moment. Using random projection it is not necessary to re-train completely the system when the set of patterns is modified.

2. RANDOM PROJECTION .

Random projection (RP) is a method for dimensionality reduction used in the preprocessing of textual data, prior to applying latent semantic indexing (LSI), which may be computed either by PCA or SVD, [19]. RP has been also used as a preprocessing step in SOM systems [9][13]. Other applications of random projection include [4][7][12].

Basically, in RP, the original data is projected to a low dimensional subspace using a random matrix in order to reduce the original dimension of the data.

Let **X** be the original data matrix, where each column contains the *m* pixel values of one image vector, **x**. The RP transform of the data matrix **X** of size *m x n* is defined as

$$\mathbf{Y}^{RP}_{k\,x\,n} = \mathbf{R}_{k\,x\,m}\mathbf{X}_{m\,x\,n} \tag{1}$$

where n is the number of observations (images) and $k << m$. This transform is allowed due to the Johnson-Lindentrauss lemma [8][5]: if points in a vector space are projected onto a randomly selected subspace of suitable high dimension, then the distances between the points are approximately preserved. The equation (1) is not a real projection of the data **X** because the matrix **R** is generally not orthogonal, so the RP transformation can cause significant distortions in the original data. However, it has been shown that **R** is almost orthogonal [6] so it is valid to consider the transformation as a projection of the original data.

As we can see, RP is computationally very simple: computing the random matrix **R** and projecting the data matrix **X**. One of the key points of the approach is the choice of the random matrix. The elements of **R** are often gaussian distributed, although other distributions can be used as it has been show by [1] recently.

3. EXPERIMENTAL COMPARISON

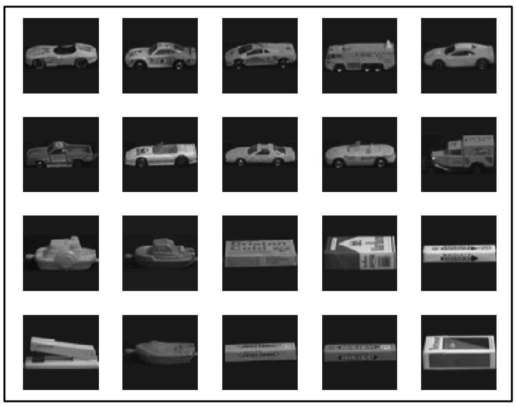
In this section, the validy of RP as a dimensionality reduction technique in an appearance-based object recognition is tested. Two different datasets are used for the comparison: first, a subset of 20 objects of the well-known COIL-100 dataset [16] (the most similar objects have been chosen in order to avoid extremely high recognition rates that would made the comparison useless); and second, a set of 30 face images from the ORL database [18]. Fig. 2 shows some examples of both databases.

To check whether RP works properly or not, a nearest neighbour classifier is used with different input data:

- raw data or data in the original image space: the grayscale pixel values of the images (size 64x64 for the COIL images and 112 x92 for the ORL faces)
- features extracted using PCA from the the grayscale original images : different tests are performed with a number of components ranging from 6 to the total number of eigenvectors (equal or lower than the number of images used in the training data matrix)
- features extracted using RP from the the grayscale original images: different tests are performed with a number of dimensions ranging from 6 to 1000.

In all cases, ten repetitions of a cross-validation test are performed in order to obtain reliable results. These results are shown on Fig. 3.a. and Fig. 3.b. It is possible to see that both PCA and RP techniques allow to obtain results similar to those obtained when the classifier is fed with raw data. Such result shows that both approaches are valid for feature extraction. However, RP requires a larger number of features than PCA to obtain the same classification results and particularly, with the COIL dataset, the results obtained with PCA slightly outperformed the best results obtained with RP. The number of features required when RP is used, even being greater than that of PCA, is much smaller than the number of pixels of the original image, so dimensional reduction is achieved and the classifier is not slowed down.

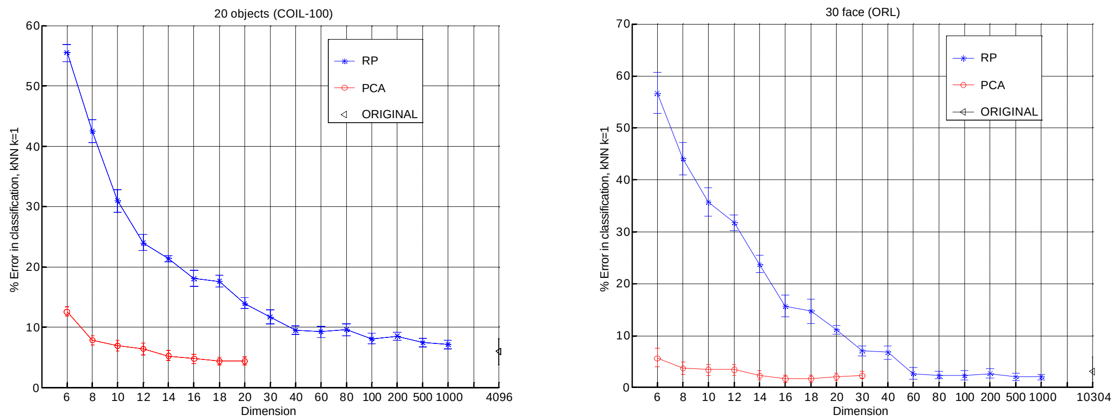
(a)



(b)

Fig. 2. Image examples of the databases used in the experiments: a) ORL face image database, b) 20 most similar objects from the COIL-100

As a conclusion, PCA should be used when off-line computation time (training step) is not relevant, as the number of features required is smaller; and RP should only be used when off-line computation time is important: i.e. when the training examples are not prior fixed and training has to be performed continuosly. In such cases, it is enough to increase the number of features in order to obtain similar results to those obtained by PCA avoiding the computational time required to the principal components.



(a) (b)

Fig. 3. Experimental results of the comparison between raw data, PCA and RP features: a) results for the most similar objects from COIL-100 dataset, b) results for the ORL face image database.

4. APPLICATIONS OF RP IN APPEARANCE-BASED SYSTEMS

The main advantage of RP is its independence from the data: the transformation matrix **R** is not relative to the data in **X**; that's not the case of PCA, where the matrix transformation is composed by the eigenvectors computed from the covariance matrix of **X**. So using RP it is not necessary to re-train completely the system when the set of patterns is modified.

The method presented may be successfully used in applications with varying patterns, where learning has to be performed several times, and under such circumstances the training time is relevant. These applications include those related to access control, e.g. where it is desired to maintain a register of the people which are inside a building, or the cars in a garage.

5. CONCLUSIONS

- Appearance-based object recognition requires a high amount of image information (multiple channels) in order to obtain the best results.
- Traditional dimensionality reduction techniques like PCA are not applicable when training has to be performed continuosly due to their high computational load. RP is an alternative in such environments as similar results to those of PCA can be obtained and the training computational load is highly reduced.
- The use of RP allows to increase the number of channels of information used and thus the global results may outperform those of PCA when a single channel is used.

6. AKNOWLEDGEMENT

This research is sponsored by the spanish Ministerio de Educación y Ciencia (Project reference: DPI2004-07433-C02-01. Title: HERRAMIENTAS DE TELEOPERACION COLABORATIVA. APLICACION AL CONTROL COOPERATIVO DE ROBOTS).

References

[1] Achlioptas, D., “Database-friendly random projections”, *Proc. ACM Symp. on the Principles of Database Systems*, 2001, pp. 274-281.

[2] Bartlett, W.M., "SEEMORE:Combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition”, *Neural Computation*, 1997, no. 9, pp. 777-804.

[3] Belhumeur, P.N., Hespanha, J.P., and Kriegman, D.J., ”Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1997, vol. 19, no. 7, pp. 711-720.

[4] Bingham, E. and Mannila, H., “Random projection in dimensionality reduction: Applications to image and text data”**,** *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2001)*, 2001, San Francisco, CA, USA, pp. 245-250.

[5] Frankl, P. and Maehara, H., “The Johnson-Lindenstrauss lemma and the sphericity of some graphs”, *Journal of Combinatorial Theory, Ser. B*, 1988, vol. 44, pp. 355-362.

[6] Hecht-Nielsen, R., “Context vectors: general purpose approximate meaning representations self-organized from raw data”, *Computational Intelligence: Imitating Life*, J.M. Zurada, R.J. Marks II and C.J. Robinson Ed. , 1994, pp.43-56.

[7] Indyk, P. and Motwani, R., “Approximate nearest neighbors: towards removing the curse of dimensionality”, *Proc. 30th Symp. on Theory of Computing*, 1998, pp. 604-613.

[8] Johnson W.B. and Lindenstrauss, J., “Extensions of Lipshitz mapping into Hilbert space”, *Conf. in modern analysis and probability, vol. 26 of Contemporary Mathematics*, Amer. Math. Soc., 1984, pp.189-206

[9] Kaski, S., “Dimensionality reduction by random mapping”, *Proc. Int. Joint. Conf. on Neural Networks*, 1998, vol.1, pp 413-418.

[10] Kirby, M. and Sirovich, L. , ”Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1990, vol. 12, no.1, pp. 103-108.

[11] Kirby, M., Geometric Data Analysis, 2001, Wiley Interscience.

[12] Kleinberg, J.M., “Two algorithms for nearest-neighbor search in high dimensions”, *Proc. 29th ACM Symp. on Theory of Computing*, 1997, pp. 599-608.

[13] Kurimo, M., “Indexing audio documents by using latent semantic analysis and SOM”, *Kohonen Maps*, E. Oja and S.Kaski Ed., Elsevier, 1999, pp.363-374.

[14] Murase, H., and Nayar, S.K., ”Visual Learning and Recognition of 3D Objects from Appearance”, *Int’l J. Computer Vision*, 1995, vol. 14, pp. 5-24.

[15] Nayar, S.K., Nene, N.A., and Murase, H., ”Subspace Methods for Robot Vision”, *IEEE Trans. Robotics and Automation*, 1996, vol. 12, no.5, pp. 750-758.

[16] Nene, S. A., Nayar, S. K. and Murase H., “Columbia object image library (COIL-100)”, *Technical Report CUCS-006-96*, http://www1.cs.columbia.edu/CAVE/research/softlib/coil-100.html, February 1996.

[17] Ohba, K., and Ikeuchi, K., ”Detectability, uniqueness, and reliability of eigen windows for stable verification of partially occluded objects”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1997, vol. 19, pp. 1043-1048.

[18] The ORL database of faces, AT&T Laboratories Cambridge. http://www.uk.research.att.com/facedatabase.html.

[19] Papadimitriou, C.H., Raghavan, P., Tamaki, H., and Vempala, S., “Latent semantic indexing: A probabilistic analysis”, *Proc. 17th ACM Symp. on the Principles of Database Systems*, 1998, pp. 159-168.

[20] Swets, D.L. and Weng, J.J., ”Using Discriminant Eigenfeatures for Image Retrieval”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1996, vol. 18, no.8, pp. 831-836.

[21] Turk, M. and Pentland, A. , ”Eigenfaces for Recognition”, *J. Cognitive Neuroscience*, , 1991, vol. 3, no. 1, pp. 71-86.