MPEG-7 DESCRIPTORS BASED CLASSIFIER FOR FACE/NON-FACE DETECTION

Malek Nadil

Department of Computer Science USTHB, Algiers, Algeria

Abdenour Labed

Department of Computer Science Polytechnic School of Bordj El Bahri, Algiers, Algeria

Feryel Souami

Department of Computer Science USTHB, Algiers, Algeria

Keywords: Image retrieval, MPEG-7, Classification, Semantic.

Abstract: In this paper we present a high level Face/Non-face classifier which can be integrated to a content based image retrieving system. It will help to extract semantics from images prior to their retrieving. This two-steps retrieval allows reducing effects of semantic gaps on the performance of existing systems. To construct our classifier, we exploit a standardized MPEG-7 low level descriptor. Experiments performed on images taking from two data bases, showed that our technique outperforms others presented in the literature.

1 INTRODUCTION

Several content based images retrieving systems (CBIR) such as Virage (Bach et. al, SPIE Conf. on Vis. Commun.and Image Proc.), (Virtual Information Retrieval Image Engine) have been developed. Virage was based on color (color layout, composition), texture, and the object boundary structure information to help in the visualization management. It not only provided static images retrieval facilities but offered some video retrieval functions as well. QBIC (Flicker et. al, 1995), (Query by Image Content) is an other system developed by IBM Almaden Research Center. It was the first image database retrieval system. It provided a color similarity comparison and was very suit for a scenic photo retrieving. Photobook(Pentland et. al, 1996) proposed by MIT multimedia laboratory contained three sub-books: Appearance book, Shape book, and Texture book. It provided different retrieving algorithms and also got most closely to some domains. VisualSEEK (Smith and Chang, 1996) developed by Image and ATV Lab of

Columbia University, provided both image and video query by using examples.

However, in spite of numerous attempts to design reliable engines, all developed systems still suffer from the semantic gap between user's inquiry and results provided by these systems. The main reason is the fact that they are based on low level descriptors characterized by semantic lacking. Indeed, they can not use high level descriptors like annotation since this latter is still extracted by hand, which is practically a tedious task. That is why we introduce an idea in between. It consists of adding a classification step before starting the retrieving process. As a first step (prior to searching in a data base), we use our classifier to detect if the user's query contains human faces or not and hence minimizing irrelevant (aberrant) results suggested as answers to the query. This general problem is often referred to as "bridging the semantic gap" (Lew and Huijsmans, 1996). This classifier not only improves the pertinence of the results but shortens the time response of searching engines as well; since the original searching space has been partitioned into two smaller parts. Tests on the data bases persons and no_bike_no_person of the Institute of Electrical

Measurement and Measurement Signal Processing of Graz University of Technology (Austria) demonstrated the good performance of our classifier. This paper is organized as fellows: in section 2, we describe the classifier construction steps, beginning by features extraction and ending by validation, in section 3, we present some experimental results and we end by conclusions in section 4.

2 CLASSIFIER CONSTRUCTION

We first recall that MPEG-7 visual descriptors (ISO\IEC, 2001) standardized for the image content description are a compressed description of image features, which are represented in terms of primitive image features such as color, texture, and shape of the image. Among the MPEG-7 visual descriptors, we have chosen the edge histogram descriptor (EHD) (Won et. al, 2002) as features of images to be classified in two classes (Face/Non-Face).

To construct our classifier, an image data base is needed for its training and testing. In our work we used a data set of 470 images among which 214 contain human faces. Whilst for features (image descriptors) extraction, we used the MPEG-7 XM module (Manjunath et. al, 2001), to get the Edge Histogram Descriptor.

2.1 Edge Histogram Descriptor (EHD)

The EHD of the MPEG-7 visual descriptors represents the distribution of five edge types, namely vertical, horizontal, 45- degree diagonal, 135-degree diagonal, and non-edge types (ISO\IEC, 2001), (Won et .al, 2002). The distribution of five edge types is represented by 16 local edge histograms. Each local histogram is generated from each sub-image. A sub-image is a non-overlapping 4x4 partition of the given image space. That is, an image is divided into non-overlapping 4x4 sub-images. Then, each sub-image is used as a basic region to generate an edge histogram, which consists of five bins with vertical, horizontal, 45-degree diagonal, 135-degree diagonal and non-directional edge types. Note that the image block may or may not have an edge in it. If there is an edge in the block, the counter of the corresponding edge type is increased by one. Otherwise, the image block has monotonous gray levels and no histogram bin is increased. After examining all image blocks in the sub-image, the 5-bin values are normalized by the *total number of blocks* in the sub-image. Thus the sum of the normalized five bins is not necessarily 1. Finally, the

normalized bin values are quantized for the binary representation. Since there are 16 (4 x 4) sub-images, each image yields an edge histogram with a total of 80 (16 sub-images x 5 bins/sub-image) bins. These normalized and quantized 80 bins constitute the EHD of the MPEG-7. That is, arranging edge histograms for the sub-images in the raster scan order (bloc order is done according to that of lines), 16 local histograms are concatenated to have an integrated histogram with 80 (16x5) bins.

Once EHD descriptors have been recovered, Independent Component Analysis (ICA) is then applied to obtain independent parameters and keep just pertinent information. We have fixed a percentage of retained information in the prewhitening step by Principle Component Analysis (PCA).

For the classification step, we tried many classifiers (Bayes classifiers, support vectors machine, K-nearest neighbors,…) but we limit our selves here to a brief description of classifiers that gave satisfactory results: the Nearest Mean Classifier (NMC) and Linear Fisher Discrimenant Classifier (LFDC) . For the tests of classification scores, the cross-validation (mainly the leave one out) strategy has been adopted. All our tests were performed with the Matlab Toolbox PRTools4 (http://prtools.org/).

2.2 Nearest Mean Classifier

Nearest mean classifier calculates the centers of in-class and out-class training samples and then assigns the upcoming samples to the closest center. This classifier gives two distance values as output and should be modified to produce a posterior probability value. A common method used for K-NN classifiers can be utilized (Arlandis et. al, 2002). According to this method, distance values are mapped to posterior probabilities by the formula:

$$P(W_i \ / \ x) = \frac{1}{d_{mi}} / \sum_{j=1}^{2} \frac{1}{d_{mj}} \qquad (1)$$

where W_i refers to the ith class (i=1,2.), d_{mi} and d_{mj} are distances from the ith and j^{th} class means, respectively. In addition, a second measure recomputes the probability values below a given certainty threshold by using the formula (Arlandis et. al, 2002):

$$P(W_i \ / \ x) = \frac{N_i}{N} \qquad (2)$$

where N_i is the number of in-class training samples whose distance to the mean is greater than x, and N is the total number of in-class samples. In this way, a more effective nearest mean classifier can be obtained.

2.3 Linear Fisher Discriminant Classifier

Linear Fisher discriminant (LFD) is a well-known two-class discriminative technique. It aims to find the optimal projection direction such that the distance between the two mean values of the projected classes is maximized while each class variance is minimized.

The optimal discrimination mask can be computed explicitly in a closed form by the following formula (Duda et. al, 2004):

$$w^* = \arg\max_w J \qquad (3)$$
$$= S_W^{-1}(m_1 - m_2)$$

Where:

$$m_j = \frac{1}{l_j}\sum_{i=1}^{l_j} x_i^j \qquad (4)$$

is the mean (center) of the jth class, l the total number of training samples of all classes, and l_j the number of samples in the jth class. and:

$$S_W = \sum_{j=1}^{2}\sum_{i=1}^{l_j}(x_i - m_j)(x_i - m_j)^T \qquad (5)$$
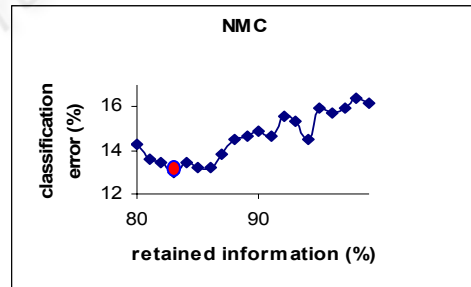
is the covariance within classes.

3 EXPERIMENTS

Table 1 summarizes the classification results for nearest mean classifier (NMC) and linear Fisher discriminant classifier (LFDC). The first column gives the percentage of retained information, the second contains the dimensionality of the space corresponding to each percentage. While the third and the fourth columns give the classification error (in %) obtained by the leave one out procedure for the NMC and the LFDC respectively. We can notice that the best scores are obtained for 83% of retained information (36 components) for both classifiers. This error is 12.979% for NMC and 13.404% for LFDC.
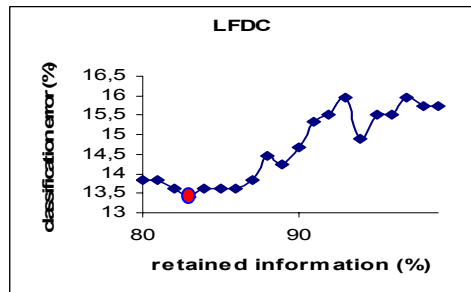
These errors are illustrated on the figure 1 below. The retained information (%) is represented on X-axis while the Y-axis represents classification error.

Table 1: Leave one out classification error for NMC and LFD based on ICA preceded by a PCA whitening.

Retained Information (%)	Number of components	Error (%)	
		NMC	LFDC
80	32	14.255	13.83
81	33	13.617	13.83
82	35	13.404	13.617
83	36	**12.979**	**13.404**
84	38	13.404	13.617
85	39	13.191	13.617
86	41	13.191	13.617
87	43	13.83	13.83
88	45	14.468	14.468
89	46	14.681	14.255
90	48	14.894	14.681
91	50	14.681	15.319
92	53	15.532	15.532
93	55	15.319	15.957
94	58	14.468	14.894
95	60	15.957	15.532
96	63	15.745	15.532
97	67	15.957	15.957
98	70	16.383	15.745
99	75	16.17	15.745

(a)

(b)

Figure 1: Classification error evolution according to variation of retained information.

Hereafter, examples of well classified and misclassified images are given for both classes (Face/Non-Face). For instance, the images below, have been well classified:

Figure 2: Examples of well classified images.

We can notice that the first pair refers to scenes in which human faces are present, while in the next pair there are no human faces.

The case of misclassified images is also illustrated below:

Figure 3: Examples of misclassified images.

4 CONCLUSIONS

We have constructed a classifier based on EHD as discriminating features and the nearest mean rule for supervised classification, to classify images from two classes: Face/Non-Face.

About 87% of good classification score as been obtained, which is slightly better than the scores obtained by the team of SHEMA Reference system(Mezaris et. al, the Schema

Reference System). Hence, we consider that our approach is able to produce satisfactory results.

In a future work, we intend to expand our approach to the case of more than two classes (sky, trees, water, animals, etc). For the classification rule we will try the classifiers combining.

REFERENCES

J.R. Bach. et.al. The Virage Image Search Engine: an OPRN Framework for Image Management. in PROC. SPIE Conf. on Vis. Commun.and Image Proc.

M. Flicker. et.al. Query by Image and Video Content: The QBIC System. IEEE Computer. Vol.28. Nº9. pp.23-32. 1995.

A. Pentland. et.al. Photobook: Content-based Manipulation of image databases. International Jounal of Computer Vision 1996.

J. R. Smith, S.F. Chang. Visual SEEK: a fully automated content-base image query system.

Proceeding of ACM International Conference Multimedia. boston. MA. pp.87-98. November. 1996.

M.S. Lew, N. Huijsmans. Information Theory and Face Detection. In *Proceedings of the International Conference on Pattern Recogntion*, Vienna, Austria, 601-605. 1996.

ISO/IEC 15938-3: Multimedia Content Description Interface - Part 3: Visual. 2001.

C. S. Won, D. K. Park, S.-J. Park. Efficient use of MPEG-7 edge histogram descriptor. *ETRI Journal*. 24. pp.23-30. 2002.

B. S. Manjunath, P. Salembier and T. Sikora. Introduction to MPEG-7: Multimedia Content Description Standard. Wiley. New York. 2001.

J. Arlandis, J.C Perez-Cortes and J. Cano. Rejection Strategies and Confidence Measures for a k-NN Classifier in an OCR Task. IEEE 2002.

R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification. New York: Wiley-Interscience, 2001.

Vasileios Mezaris et al. An Extensible Modular Common Reference System For Content-Based Information Retrieval: The Schema Reference System.