Facial Image Retrieval Through Compound Queries Using Constrained Independent Component Analysis

†Bilal Ahmed, †Tahir Rasheed, †Young-Koo Lee, †Sungyoung Lee, ‡Tae-Seong Kim*
†Ubiquitous Computing Lab, Department of Computer Engineering
‡Bio-Imaging Lab, Department of Biomedical Engineering
Kyung Hee University
Suwon, Republic of Korea
†{bilal,tahir,yklee,sylee}@oslab.khu.ac.kr, ‡tskim@khu.ac.kr

Abstract

In this work we present a new technique of facial-image retrieval using constrained independent component analysis (cICA). We have employed cICA for the online extraction of those independent components from the entire database which bear some similarity to the query-images. Instead of using any offline learning mechanism or feature extraction technique, our system works completely online. It can cater for queries formulated from both single and multiple examples, for achieving higher accuracy. For compound queries, instead of treating each query-image independently, the system is capable of finding images similar not only to the individual query-images, but also to their different combinations [1].

1. Introduction

Visual information plays a crucial role in various domains, from medical diagnosis, journalism, crime-prevention to surveillance. Whereas domain specific images carry specific semantics, the problem of interpreting visual information becomes more complex when we talk of natural images. The maxim, *'A picture is worth a thousand words'* explains this inherent problem very concisely. Indexing of large databases of images for efficient retrieval is crucial for various domains such as journalism, biomedicine, forensics etc. Manual indexing of images in such large databases can be highly subjective and time consuming. In contrast, content-based image retrieval (CBIR) focuses on the development of efficient retrieval mechanisms based on image features or meta-data used for image annotation.

Conventional approaches to CBIR represent images in the form of image-based features. These features vary from global image descriptors such as color or intensity histogram to local ones such as shape and texture. These features along with their combinations have been used previously for CBIR. For example in [8], a region-based color-descriptor modeling the color values along with their percentages in the region is proposed. Similarly in [9], multi-resolution histograms have been employed for the retrieval of textured images. In [12], the extraction of color histograms through Gaussian mixture vector quantization has been proposed. In [3] and [16] respectively, shape descriptors and shape matching algorithms have been proposed for image retrieval.

The use of low-level image features such as color histograms, shape and texture attributes introduces a semantic gap [5]. This semantic gap arises due to the inability of such low-level features to describe the objects and their inter-relations within the image. The use of such low-level features places the responsibility of achieving semantically coherent results on the user-interface. Various techniques of relevance feedback [17] have been introduced in this context. Whereas user feedback might be able to lower this gap, the overall procedure becomes subjective and requires a higher degree of user interaction.

Segmentation based techniques for image retrieval have also been used for obtaining better shape, texture and color descriptions of the image contents [7]. The motivation behind this particular approach is that: objects within an image can be segmented, and their description used for querying the database to retrieve more semantically similar images. Various segmentation techniques, such as the Normalized Cuts [19], Mean Shift Procedures, Expectation Maximization [4] algorithms have been used in image re-

*The corresponding author for this publication is Dr. Tae-Seong Kim.
[1]This research was supported by the MIC(Ministry of Information and Communicaton), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute of Information Technology Advancement)(IITA-2006-(C1090-0602-0002)).

1082-3409/07 $25.00 © 2007 IEEE
DOI 10.1109/ICTAI.2007.60

trieval. Machine-learning approaches augmented with segmentation techniques have also been used. In [21], segmentation results augmented with fuzzy logic are used to obtain *soft* similarity measures. The problem of obtaining a semantically coherent segmentation of an image still remains an open research problem and higher dependency on segmentation-results is not desirable for achieving a semantically accurate retrieval performance.

From an image-retrieval point of view, facial images have attracted a lot of attention. Various machine learning and feature extraction techniques have been employed for the efficient retrieval of facial images. Earlier retrieval systems such as the Photobook [15] uses principal component analysis (PCA) for the retrieval of facial images. In [13], feature extraction through ICA in a reduced PCA space is used for characterizing query images. The overall system comprises of classifying the input query image based on the nearest-neighbor rule using various similarity measures. Recent work on facial image retrieval [2] has focused on representing facial images as a collection of local independent components. For this purpose, the query images are decomposed into a number of overlapping and non-overlapping windows to compute the independent components. The decomposition of the image into constituent windows may result in feature distortion and loss of information.

The use of multiple images as a compound query has not been explored in much detail. Conventional CBIR systems do not provide a mechanism through which a user can specify his search criterion through multiple examples. This is analogous to multi-word queries in search-engines: the specification of a compound query helps the system in retrieving the desired results with more accuracy. Similarly, when a user cannot find a single image which can specify his search criterion he should be able to use multiple images to formulate his query. Multiple queries have been used in [20]: the approach taken is to find a combined result of the query by using the retrieved images corresponding to each query image, independently. Similarly, [2] uses multiple facial images to retrieve images similar to the independent query-images as well as to their combinations.

We have devised a system which can cater for both single and multiple exemplar image retrieval. It does not decompose the query images or the database images to windows as in [2] or uses PCA [13] for dimension reduction; so the chances of any information loss are minimal. There is also no need to store additional feature information or the need for any offline learning as in [2]. The approach is centered around the idea of cICA [14] which has the ability to extract specific independent components conforming to certain prior information (known as reference signals). Query images are provided to the cICA algorithm as references, and the output of the cICA algorithm specifies the contribution of each database image to the extracted component. Based on the magnitude of this contribution factor the database images are ranked.

2. Constrained ICA

Conventional ICA techniques perform blind-source separation (BSS) assuming a linear mixing model of the independent sources. If the observed image is represented as $X = (x_1, x_2, ..., x_n)$ and the original sources as $S = (s_1, s_2, ..., s_m)$. ICA assumes that $\mathbf{x}$ is a linear mixture of the original independent sources. Therefore,

$$X = AS \tag{1}$$

where A is the mixing matrix of size $n \times m$. Conventional ICA algorithms aim at finding an $m \times n$ demixing matrix W to recover all the ICs of the observed image.

Existing ICA algorithms find as many ICs as the number of observations. The user must manually identify which ICs represent which sources. The primary reason being the inability of the ICA algorithms in calculating the energies or signs of the ICs. This may also lead to problems where the number of sources are less than the number of observations. Deflation based ICA techniques [6][11] have also been developed, but they suffer from the arbitrary ordering of the extracted independent components.

cICA [14] has been developed to find only those independent components which are of interest to the current task at hand. This is achieved by providing some prior knowledge about these ICs to the cICA algorithm. This prior information may not be exact information, it could be the specification of statistical properties of the desired component or just a crude approximation (e.g., template). Therefore, if we have some *a priori* information about the desired sources, we can incorporate this information into cICA. The cICA algorithm uses this *a priori* information about the desired IC, encoded into reference image(s) r_i to obtain an output which is statistically independent from other sources. The closeness constraint can be written as,

$$g(W) = \epsilon(y, r) - \xi \leq 0 \tag{2}$$

where ϵ is some closeness measure and ξ some closeness threshold parameter. The measure of closeness can take any form, such as mean squared-error (MSE), correlation, or any other suitable closeness or similarity measure. The number of reference signals determine the number of independent components to be extracted from the complete set of observations.

The final mathematical model for cICA can be represented as,

$$maximize : \sum_{i=1}^{l} J(y_i)$$

$$subject \text{ to} : g(W) \leq 0, \; h(W) = 0. \tag{3}$$

where

$$J(y) \approx \rho[E\{G(y)\} - E\{G(v)\}]^2 \tag{4}$$

is the one-point contrast function for ICA introduced in [10]. W is the weight matrix to be learned; $g(W)$ is the closeness constraint and $h(W)$ constrains the output component to have unit variance. Equation (3) is a constrained optimization problem and can be solved through the use of augmented Lagrangian functions.

3. cICA-based Facial Image Retrieval

Viewing it from another perspective, the cICA framework can be used for specifying the type of information we would like to extract from huge amounts of data. The reference image(s) can be formulated as the query image(s) specified by the user. The accuracy of the extracted information depends upon the accuracy of the provided references. In our case, the image(s) provided by the user would serve this purpose, and point the cICA algorithm in the appropriate direction. The overall system architecture is depicted in Figure 1.

Since cICA extracts components y_i from the given set of observations corresponding to the provided reference images, we can ascertain the contribution of each observation by reconstructing it from the extracted component. The reconstruction procedure involves the estimation of the mixing matrix A and the reconstruction of the entire set of observations. Consider that we have n observations $X = (x_1, x_2, ...x_n)$ and i extracted components $Y = (y_1, y_2, ...y_i)$, where $i \ll n$. The mixing matrix pertaining to the extracted sources with respect to the entire set of observations can be estimated using,

$$A = XY^+ \tag{5}$$

where Y^+ is the pseudoinverse of the extracted components. Furthermore the reconstruction of X can be done using,

$$X_R = AY \tag{6}$$

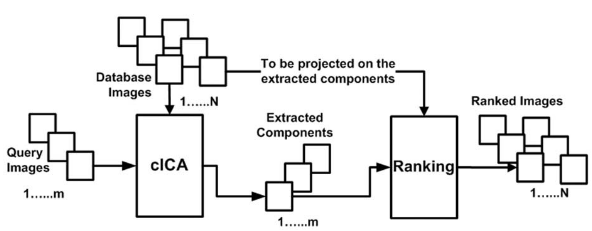
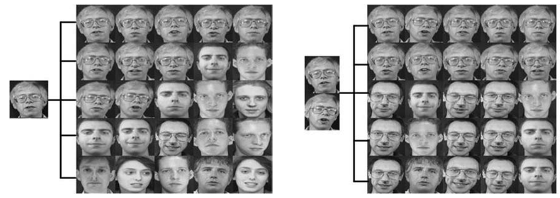


Figure 1. Overview of the cICA based CBIR system.



(a) (b)

Figure 2. cICA based CBIR applied to the ORL database.(a) A single query, here the system acts as a face recognition system. (b) Using two images of the same individual, with different pose.

where X_R is the set of reconstructed observations.

Once the data has been reconstructed, we need to estimate how well each observation has been reconstructed from the extracted sources. This involves the comparison of each reconstructed observation with its original form. Simple measures of similarity such as correlation or mutual-information can be used. In our case we have used correlation to determine the similarity between the original, and the reconstructed images.

4. Experimental Results

We have conducted extensive experiments on a publicly available facial-image database, the ORL face database [18]. The ORL database contains ten facial image of 40 individuals with varying pose, expressions and spectacles. The images were scaled to 64×64, with no pre-processing or feature extraction phase.

4.1. Homogeneous Queries

In order to retrieve the facial images of a single person from the database under varying pose and occlusion conditions a single example might not be enough. The same is also true if the database has different expressions and scale. Figure 2(a) shows the results of using a single query [2]. In the database there are ten images of the individual, with a single query image our system has been able to retrieve eight images out of the top ten retrieved images. The images which have been left out of the top ten (image(3,1), image(3,2) in Figure 2(a)) have the same individual but with his head tilted to the right side. The query image given in Figure 2(a) was unable to describe the features present in these two images.

[2]The retrieved images are shown so that the image at the top-left corner has highest rank, whereas the image at the bottom-right corner has the least.

In Figure 2(b), we have used two query images: one depicts the individual with a left tilt whereas the other depicts the same pose but in the opposite direction. All the ten relevant images have been retrieved from the database and have the highest ranking, as can be seen from the results.

4.2. Heterogeneous Queries

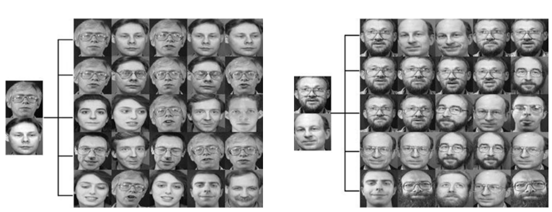
Figure 3 shows the results obtained for two heterogenous queries. In the first query of Figure 3(a), two images of two different individuals have been used. One of the subject is wearing spectacles whereas the other has none. In the retrieved images we see that the initial nine images correspond to the two individuals, with images of the second subject wearing spectacles. Similarly, after the two top rows, the system has retrieved images of individuals with and without spectacles and bearing some facial similarity to the individuals depicted in the query images.

In the second case Figure 3(b), again two different individuals are used. This time the top query image has an individual who has a beard and spectacles. Whereas, the other individual has none. The images retrieved by the system not only contain the individuals present in the query but also their various combinations. Persons having both beards and spectacles (the same as the individual in the top query image), persons having only beards, persons wearing only spectacles with no beards and persons having none of these (corresponding to the individual depicted in the lower query image).

4.3. Performance Analysis

We conducted one hundred simulations of the system with random query formulations for the homogeneous query case. A hard similarity evaluation was used, and only the retrieved images pertaining to the same individual as depicted in the query were considered relevant as opposed to [2], where it was assumed that user feedback is available. Simple measures of precision [1] and recall [1] have been used to evaluate the efficiency of the system:

$$Precision = \frac{N_{RL}}{N_R} \tag{7}$$



(a) (b)

Figure 3. Results for heterogenous queries.

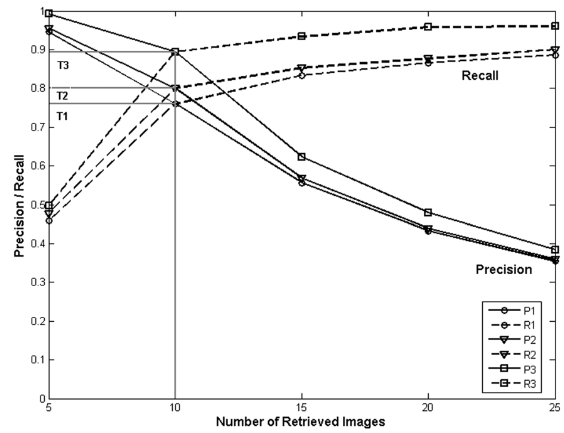


Figure 4. Precision and Recall curves for homogeneous queries. The graph shows the Precision(P) and Recall(R) values for queries formulated from 1, 2 and 3 images.

$$Recall = \frac{N_{RL}}{N_{RD}} \tag{8}$$

where N_{RL} is the number of relevant images in the retrieved images, N_R is the total number of retrieved images, and N_{RD} is the total number of relevant images in the database. In case of Figure 2(a) Precision = $\frac{10}{25}$ and Recall = $\frac{10}{10}$. Note that, when N_R equals N_{RD} the two measures become equal. This is the break-even point of the system and indicates its accuracy.

The evaluation measures for queries consisting of one, two and three images are shown in Figure 4. In the figure: T1, T2, and T3 represent the break-even points of the system for queries formulated from one, two and three images, respectively. In case of single-image queries the system has achieved an accuracy of 76%. Whereas, in the case of compound queries composed of two and three images this accuracy increases to 80% and 90% respectively. In contrast to the conventional systems, the cICA based retrieval system achieves this higher level of performance without any feature-extraction and offline-learning.

5. Conclusion

In this paper we have proposed a new technique of facial image retrieval based on cICA. Our technique requires no offline learning, pre-processing and feature extraction. The system has been designed so that none of the user-provided information is lost, and in turn more semantically accurate images can be retrieved. As our future work we would like to test the system in other domains such as the retrieval of chest x-rays and CT-scans.

References

[1] R. A. Baeza-Yates and B. A. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.

[2] J. Basak, K. Bhattacharya, and S. Chaudhury. Multiple Exemplar-Based Facial Image Retrieval Using Independent Component Analysis. *IEEE Transactions on Image Processing*, 15(12):3773–3783, December 2006.

[3] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[4] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its applications to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, 2002.

[5] Y. Chen, J. Li, and J. Z. Wang. *Machine Learning and Statistical Modeling Approaches to Image Retrieval*. Kluwer Academic Publishers, 2004.

[6] A. Cichoki, R. Thawonmas, and S. Amari. Sequential blind signal extraction in order specified by stochastic properties. *Electronics Letters*, 33(1):64–65, 1997.

[7] R. Datta, J. Li, and J. Z. Wang. Content-Based image retrieval- approaches and trends of the new age. *Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval Singapore (MIR'05).*, November 2005.

[8] Y. Deng, B. S. Manjunath, C. Kenney, M. S. Moore, and H. Shin. An efficient color representation for image retrieval. *IEEE Transactions on Image Processing*, 10(1):140–147, 2001.

[9] E. Hadjidemetriou, M. D. Grossberg, and S. K. Nayar. Multiresolution histograms and their use for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7):831–847, 2004.

[10] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[11] A. Hyvärinen and E. Oja. Simple neuron models for independent component analysis. *International Journal of Neural Systems*, 7(6):671–687, 1996.

[12] S. Jeong, C. S. Won, and R. M. Gray. Image retrieval using color histograms generated by gauss mixture vector quantization. *Computer Vision and Image Understanding*, 9(1-3):44–66, 2004.

[13] C. Liu. Enhanced independent component analysis and its application to content based face image retrieval. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 34(2):1117–1127, April 2004.

[14] W. Lu and J. C. Rajapakse. Approach and applications of constrained ICA. *IEEE Transactions on Neural Networks*, 16(1):203–212, January 2005.

[15] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: tools for content-based manipulation of image databases. *In Proc. SPIE Storage and Retrieval for Image and Video Databases II*, 2185, 1994.

[16] E. G. M. Petrakis, A. Diplaros, and E. Milios. Matching and retrieval of distorted and occluded shapes using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(11):1501 – 1516, 2002.

[17] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.

[18] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota FL*, December 1994.

[19] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[20] S. M. M. Tahaghoghi, J. A. Thom, and H. E. Williams. Are two pictures better than one?. *Proc. of the 12th Australasian Database Conference (ADC 2001)*, pages 138–144, 2001.

[21] J. Z. Wang, J. Li, and G. Wiederhold. SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947–963, 2001.