Content-based facial image retrieval using constrained independent component analysis

Nguyen Duc Thang [a], Tahir Rasheed [a], Young-Koo Lee [a], Sungyoung Lee [a], Tae-Seong Kim [b,*]

[a] Department of Computer Engineering, Kyung Hee University, South Korea
[b] Department of Biomedical Engineering, Kyung Hee University, South Korea

ARTICLE INFO

ABSTRACT

Content-based image retrieval (CBIR) is a method of searching, browsing, and querying images according to their content. In this paper, we focus on a specific domain of CBIR that involves the development of a content-based facial image retrieval system based on the constrained independent component analysis (cICA). Originating from independent component analysis (ICA), cICA is a source separation technique that uses priori constraints to extract desired independent components (ICs) from data. By providing query images as the constraints to the cICA, the ICs that share similar probabilistic features with the queries from the database can be extracted. Then, these extracted ICs are used to evaluate the rank of each image according to the query. In our approach, we demonstrate that, in addition to a single image-based query, a compound query with multiple query images can be used to search for images with compounding feature content. The experimental results of our CBIR system tested with different facial databases show that our system can improve retrieval performance by using a compound query. Furthermore, our system allows for online processing without the need to learn query images.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

An efficient image retrieval system plays important roles in such fields as medical diagnosis, geography, astronomy, home security, and photograph archiving. From the inception of image retrieval in the 1970s, the text-annotated approach has been the most popular; in this approach, each image is associated with text meta-tags. However, one image may include information from multiple classes, so the task of indexing large databases is very complex. Manual indexing of images is also intractable because it is highly subjective and time consuming. Content-based image retrieval (CBIR), first developed in the 1990s, (the "early years" development of CBIR is summarized in [40]), can overcome these disadvantages by utilizing the features of query images. Comprehensive surveys of this approach can be found in [13,25]. So far, various CBIR-based search engines and academic products have been developed. For instance, TinEye was introduced by Idée Inc. [2], MUVIS by the Tampere University of Technology [20], and ALIPR by the Pennsylvania State University [22].

Content-based image retrieval begins with extracting low-level features that could be matched within a given image database. Many low-level features have been introduced, varying from simple characteristics such as color or image moments [27] to more complex ones related to shape descriptors [10,31,39] or Fourier descriptors [8,34]. However, low-level features cannot reduce the gaps between the sophisticated requirements of human perception and the visual features that can be analyzed by a computer. When the domain images contain semantic meanings, the problem of selecting the appropriate features to describe visual information becomes even more complicated. Instead of further work on the extraction of

---

* Corresponding author. Tel.: +82 31 201 3731.
   E-mail address: tskim@khu.ac.kr (T.-S. Kim).

low-level features, advanced machine learning has been investigated to improve image retrieval schemes. Using the attributed relational graph, Krishnapuram et al. [21] illustrated a fuzzy method to represent the relationship between the segmented regions and applied a fuzzy graph-matching algorithm to match pairs of images. Some other systems developed a relevance feedback solution [36,38]. During each iteration in these systems, users label results as 'relevant' or 'irrelevant,' and based on their feedback, the retrieval mechanism is updated for future queries. Alternatively, a segment-based technique was attempted to achieve better descriptors of image content [11,12]. However, obtaining a semantically coherent segmentation of an image remains a difficult problem, affecting retrieval performance.

By considering the difficulties of generalizing a method to search for images of various objects, CBIR works with specific domain data such as facial image retrieval. As an early facial image retrieval system for a single image-based query, Photobook [32] employed a principal component analysis (PCA) to extract features in low-dimensional spaces. Meanwhile, in [41], Wu et al. proposed a facial indexing scheme based on a combination of PCA coefficients, facial landmarks, and text descriptions. Others applied the discrete cosine transform (DCT) for facial representation [29]. Although these and various recent approaches [4,43,44] have been exploited to increase the high-semantic level of facial image retrieval, there have been very few efforts to address the problem of facial image retrieval with compound queries. In text-based search engines, a compound query with multiple words is always integrated as an indispensable component. Lately, studies of facial image retrieval have suggested that a set of ICs extracted using an ICA may be used to present the mixture contents of multiple images [9]. However, an ICA in either CBIR or facial recognition systems is inappropriate for compound queries with multiple images. In CBIR, the ICA algorithm only describes the image features with edges [17,18] or with the entropy of images [33], which is insufficient to present facial information. In facial recognition, the ICA algorithm has been the most popular technique for single image retrieval [7,14,24]. In [9], Basak et al. first developed a multiple facial image retrieval system based on the ICA, in which a set of query images was divided into a number of overlapping or non-overlapping windows; then, an ICA was performed in order to extract ICs from these windows. The windows belonging to the database images were reconstructed from these ICs, and finally, the total reconstruction errors for each image were ranked for retrieval. This method has two disadvantages. First, a learning step is necessary to define the suitable size, position, and number of windows. Second, some information may be lost when decomposing a facial image into constituent windows.

In this paper, we propose a new facial image retrieval system based on the constrained ICA (cICA) [26]. Because the cICA can extract specific ICs that pertain to a priori information (known as the reference), some applications have taken advantage of the cICA for identifying interesting ICs in data, such as the artifacts from electromagnetic brain signals [35] or significant components of functional magnetic resonance images [6]. The idea of utilizing a cICA for image retrieval was first reported by us, along with some preliminary results [5]. In our system, we use the query images as the references for the cICA and then evaluate the reconstruction of each database image from the extracted ICs to determine its contribution to the query images. Based on computed contributing factors, the database images are able to be ranked to answer the queries. The demonstrated advantage of our approach is that it allows the query to use the compounding contents of multiple images. This same advantage has also been explored in the ICA-based approach of Basak et al. [9]. However, in contrast to the ICA-based approach, our method uses entire images rather than a set of separated windows within the query images to improve retrieval performance and to eliminate the need to learn local features. In addition to developing a image retrieval system, to allow online retrieval, we have proposed and implemented a fast cICA algorithm to improve the computational time of the cICA algorithm.

The rest of this paper is organized as follows. In Section 2, the basics of the cICA algorithm and its implementation are introduced. Our cICA-based image retrieval system is described in Section 3. The experimental setup and results are discussed in Section 4. We present our conclusion in Section 5 and discuss it in Section 6.

2. Constrained independent component analysis

2.1. Independent component analysis

Assuming that the received signals (or images) are a linear mixture of some source signals (or images), the goal of blind source separation (BSS) is to recover the original signals (or images) from the mixture. If the observed signals are presented as $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T$ and the original signals as $\mathbf{s}(t) = (s_1(t), s_2(t), \ldots, s_m(t))^T$, an ICA solves the BSS problem by assuming that $\mathbf{x}$ is a linear mixture of the sources,

$$\mathbf{x}(t) = \mathbf{As}(t), \tag{1}$$

where $\mathbf{A}$ is the mixing matrix with size $(n \times m)$. The ICA algorithm aims to compute the $(m \times n)$ demixing matrix $\mathbf{W} = [w_1, w_2, \ldots, w_m]^T$ to recover all ICs from the observed signals,

$$\mathbf{y}(t) = \mathbf{Wx}(t), \tag{2}$$

where $\mathbf{y}(t) = (y_1(t), y_2(t), \ldots, y_m(t))^T$. Each separated output signal is $y_i = w_i^T \mathbf{x}$ or may be presented as $y = w^T \mathbf{x}$. Maximizing the non-Gaussianity of y will cause it to converge toward one of the independent components, which is close to the original signals. To measure the non-Gaussianity, the negentropy, $J(y)$, is defined as

$$J(y) = H(y_{Gauss}) - H(y), \tag{3}$$

where y_{Gauss} is a Gaussian random variable with the same variance as that of the output signal y, and $H(y)$ denotes the entropy of y. Hyvärinen [16] introduced an approximation of negentropy as

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

where ρ is a positive constant, $f(\cdot)$ is a nonquadratic function, and v is a Gaussian variable with zero mean and unit variance. Usually, the ICA identifies as many ICs as the number of observations, and there is an arbitrary ordering of the extracted ICs [26].

2.2. Constrained independent component analysis

The advantage of the cICA is that only the desired ICs are retrieved, eliminating the arbitrary ordering problem of the conventional ICA. In this section, we summarize the essentials of the cICA. More details are available in [26]. Mathematically, the basics of the cICA are formulated from the ICA algorithm with the main objective of maximizing the negentropy term in (4). Apart from negentropy, additive constraints that make uncorrelation among estimated ICs and that restrict each output to a unit variance are depicted by

$$h_{ij}(y_i, y_j) = (E\{y_i y_j\})^2 = 0, \quad \forall i, j = 1, 2, \ldots, m, \ i \neq j \text{ and}$$ (5)

$$h_{ii}(y_i) = (E\{y_i^2\} - 1)^2 = 0, \quad \forall i = 1, 2, \ldots, m,$$ (6)

where m is the number of original sources. Meanwhile, the constraints used to minimize the closeness measurement between the outputs and the references are presented by $g(y) = (g_1(y_1), g_2(y_2), \ldots, g_m(y_m))^T$,

$$g_i(y_i) = \varepsilon(y_i, r_i) - \xi_i \leqslant 0,$$ (7)

where ξ_i is the threshold, $\mathbf{r} = (r_1, r_2, \ldots, r_m)^T$ is the prior information guiding the output signals, and $\varepsilon(y_i, r_i)$ is the closeness measurement. The most common form of $\varepsilon(y_i, r_i)$ is the mean square error (MSE) $\varepsilon(y_i, r_i) = E\{(y_i - r_i)^2\}$ and the correlation $\varepsilon(y_i, r_i) = -E\{y_i r_i\}^2$.

The optimization formulation of cICA is given by

$$maximize \ \sum_{i=1}^{m} J(y_i) = \sum_{i=1}^{m} \rho(E\{f_i(y_i)\} - E\{f_i(v)\})^2$$

$$\text{subject to } h(\mathbf{y} : \mathbf{W}) = 0, g(\mathbf{y} : \mathbf{W}) \leqslant 0$$ (8)

or

$$minimize \ \mathcal{J}(\mathbf{y} : \mathbf{W}) = -\sum_{i=1}^{m} J(y_i)$$

$$\text{subject to } h(\mathbf{y} : \mathbf{W}) = 0, g(\mathbf{y} : \mathbf{W}) \leqslant 0,$$ (9)

where $h(\mathbf{y} : \mathbf{W}) = (h_{11}(y_1), h_{12}(y_1, y_2), \ldots, h_{1m}(y_1, y_m), h_{21}(y_2, y_1), \ldots, h_{mm}(y_m))^T$ and $g(\mathbf{y} : \mathbf{W}) = (g_1(y_1), g_2(y_2), \ldots, g_m(y_m))^T$. Including the Lagrange multipliers μ and λ, the optimization function is rewritten as

$$L(\mathbf{W}, \mu, \lambda) = \mathcal{J}(\mathbf{y} : \mathbf{W}) + G(\mathbf{y} : \mathbf{W}, \mu) + H(\mathbf{y} : \mathbf{W}, \lambda).$$ (10)

The explicit forms of $G(\mathbf{y} : \mathbf{W}, \mu)$, $H(\mathbf{y} : \mathbf{W}, \lambda)$ and the updated rules for $\mathbf{W}$, μ, and λ can be found in [26].

2.3. Implementation of a fast cICA

In this work, we have developed a fast version of the cICA algorithm for an online CBIR system. Our implementation of this fast cICA improved its computational time by guaranteeing a faster convergence speed than the original cICA by making some changes. The simultaneous decorrelation process for all output signals using the constraint $h(\mathbf{y} : \mathbf{W})$ in (5) and (6) with the conventional cICA was too rigid and slow to allow for convergence. Therefore, first, our cICA algorithm was reformulated by replacing a simultaneous extraction with a one-by-one extraction to extract output signals. Second, a step normalizing weight vectors was integrated into our cICA to avoid the need to use the constraint $(E\{y_i^2\} - 1)^2 = 0$ to limit the variance of the output signals to be a value of one. Lastly, we applied a preprocessing with whitening to produce the uncorrelated input signals and to keep the important information from the facial databases (the important components correspond to the largest eigenvalues computed in the whitening process [7]). Consequently, the integrated preprocessing with whitening enabled our algorithm to take less convergent steps than the conventional cICA to make the output signals uncorrelated [16] (to satisfy the constraints $(E\{y_i y_j\})^2 = 0, \forall i \neq j$). The mathematical details of our cICA algorithm for multiple references integrated with these changes are described below.

The approximate gradient of (10) for a one-unit reference is given by Huang and Mi [15]:

$$\nabla_w L = -E\{J'(y)\mathbf{x}^T\} + \mu \nabla_w g(y) + 4\lambda(E\{y^2\} - 1)E\{y\mathbf{x}^T\}.$$ (11)

We replace the last term (used to restrict outputs having a unit variance) with $\|w\|^2 = 1$ (satisfied by normalizing the weight vector w). For each output y_p, $p - 1$ uncorrelation conditions among the estimated ICs are introduced as a constraint $(E\{y_p y_j\})^2 = 0$, $\forall j = 1, 2, \ldots, p - 1$. According to the Kuhn–Tucker theorem, with the Lagrange multiplier λ_{jp}, $\forall j = 1, 2, \ldots, p - 1$, the optimization equation for $\nabla_{w_p} L = 0$ can be rewritten as

$$-E\{J'(y_p)\mathbf{x}^T\} + \mu \nabla_{w_p} g(y_p) + 2 \sum_{j=1}^{p-1} \lambda_{jp} E\{y_p y_j\} E\{y_j \mathbf{x}^T\} = 0. \tag{12}$$

We now try to solve this equation by using Newton's method. The Jacobian matrix of the left-hand side of (12) is approximated by $\nabla_{w_p}^2 L = \left(-E\{\hat{\rho} f''(y_p)\} + E\left\{ \mu g''_{y_p}(y_p) \right\} \right) \Sigma_{xx} + 2 \sum_{j=1}^{p-1} \lambda_{jp} w_j w_j^T$, where $\hat{\rho} = 2\rho(E\{f(y_p)\} - E\{f(v)\})$, and $\Sigma_{xx} = E\{\mathbf{x}\mathbf{x}^T\}$. $E\{y_p y_j\} E\{y_j \mathbf{x}^T\}$ is further simplified to $\left(w_p^T w_j \right) w_j^T$ because the pre-whitening processing transforms Σ_{xx} into an identity matrix. The update rule for each iteration is given by

$$w_p \leftarrow w_p - \nabla_{w_p}^2 L^{-1} \left(-E\{J'(y_p)\mathbf{x}^T\}^T + \mu \nabla_{w_p} g(y_p)^T + 2 \sum_{j=1}^{p-1} \lambda_{jp} \left(w_p^T w_j \right) w_j \right),$$

$$w_p \leftarrow w_p / \|w_p\|. \tag{13}$$

Our fast cICA algorithm is summarized in Table 1.

2.4. Performance analysis of the fast cICA algorithm

We verify our proposed implementation of the cICA in reducing computational time by performing the following experiments involving facial images. Four facial images of the 108th, 21th, 18th, and 72th persons from the Aleix face (AR) database [28] were considered as original sources, resized to (96×96), and converted to four (1×9216) signals, S_1, S_2, S_3, and S_4. We randomly generated a (4×4) random mixing matrix to create mixed images as shown in Fig. 1(b). We then attempted to recover the first two original images from the mixtures, using the images of the first two subjects as the references. Notice that the two reference faces were occluded by glasses. The extracted results from our algorithm can be seen in Fig. 1(d).

To compare the recovery performance of our fast cICA to that of the conventional cICA [26], we computed the individual performance index (IPI), defined as $IPI = \left(\sum_{j=1}^m \frac{|p_j|}{max_k |p_k|} - 1 \right)$, $k = 1, 2 \ldots, m$, where p_j denotes the element j of vector $P = w^T \mathbf{A}$, and the peak signal-to-noise ratio (PSNR) is defined as $PSNR(dB) = 10 log_{10} \left(\frac{\sigma^2}{MSE} \right)$, where σ^2 is the variance of the desired source signal and MSE is the mean square error between the original signal and the recovered signal. In our algorithm,

Table 1
Implementation of the fast cICA algorithm for multiple references.

1.	Pre-whitening the observed signals, set $p \leftarrow 1$, set $m \leftarrow$ number of references
2.	Initialize vector w_p with a random value from the uniform distribution, $\lambda_{jp}, \forall j = 1, 2, \ldots, p - 1$ and μ are set to 0
3.	Update $\mu \leftarrow \max\{0, \mu + \gamma g(y_p)\}$, where γ is the update rate
4.	Update $\lambda_{jp} \leftarrow \lambda_{jp} + \gamma(E\{y_p y_j\})^2 \approx \lambda_{jp} + \gamma \left(w_p^T w_j \right)^2, \quad \forall j = 1, 2, \ldots, p - 1$
5.	Update $w_p \leftarrow w_p - \nabla_{w_p}^2 L^{-1} \left(-E\{J'(y_p)\mathbf{x}^T\}^T + \mu \nabla_{w_p} g(y_p)^T + 2 \sum_{j=1}^{p-1} \lambda_{jp} \left(w_p^T w_j \right) w_j \right)$
6.	Normalize $w_p \leftarrow w_p / \|w_p\|$
7.	If w_p has not converged, go back to step 3
8.	Increment $p \leftarrow p + 1$, go back to step 2 until p equals m

Fig. 1. The results of our fast cICA algorithm on real facial image data, (a) original images, (b) mixed images, (c) reference images, and (d) recovered images.

Table 2
Comparison of PSNR and IPI performances, convergent updating steps, and running times between our algorithm and the conventional cICA with facial image data.

Algorithm	PSNR (dB)		Total IPI	Convergent updating steps	Running time (s)
	S1	S2			
Conventional cICA	26.96	23.05	0.08	41	0.61
Fast cICA with multiple references	27.09	23.85	0.08	9	0.08

because the non-Gaussianity $E\{J'(y_p)\mathbf{x}^T\}$ and the reference constraint $\nabla_{w_p} g(y_p)$ of (12) are retained in the same formulations as those in the conventional cICA, our fast cICA algorithm produces negligible effects on the recovery performance. In Table 2, the performances, indicated by the values of IPI and PSNR, are almost the same for both algorithms, validating our implementation. However, in terms of the computational time, our method requires fewer updating steps to converge than does the conventional cICA and is therefore faster than the original cICA and should be more suitable for online processing.

3. cICA-based content facial image retrieval system

In this work, we propose a new facial image retrieval system based on the cICA to extract a set of ICs from the whole database using query images as the constraints. The query images are considered as reference information specified by the users. Obviously, the more accurate the user information is, the better the results of IC extraction in terms of retrieval results will be. Then, the extracted ICs are used to reconstruct the database. Finally, the similarity between each image in the database and its reconstruction is evaluated for ranking. The overall architecture of our system is described in Fig. 2.

In the following, we formally describe the mathematics of our approach. Let each image of size $(h \times w)$ be represented by a vector $(1 \times hw)$. There are a total of m extracted ICs, $y_1, y_2, \ldots, y_m$, corresponding to m reference images. Because the cICA extracts ICs from the given set of images, we can calculate the mixing matrix $\mathbf{A}$ and determine the contribution of each image by reconstructing it from the extracted components. We have n images $\mathbf{X} = (x_1, x_2, \ldots, x_n)^T$ and m extracted components $\mathbf{Y} = (y_1, y_2, \ldots, y_m)^T$, where $m \ll n$. The mixing matrix for the extracted sources with respect to the observations can be computed using

$$\mathbf{A} = \mathbf{X}\mathbf{Y}^+, \tag{14}$$

where $\mathbf{Y}^+$ is the pseudoinverse of the extracted components. We then reconstruct the whole database of images $\mathbf{X}$ using

$$\hat{\mathbf{X}} = \mathbf{A}\mathbf{Y}, \tag{15}$$

where $\hat{\mathbf{X}} = (\hat{x_1}, \hat{x_2}, \ldots, \hat{x_n})^T$ is the set of reconstructed images. After completing the data reconstruction, we score each image according to its reconstruction from the extracted ICs. We can apply some similarity measurements such as mutual information, the Euclidean distance, or the Mahalanobis distance [24] to compute the difference between the original image and the reconstructed image. In our paper, we used the cosine distance because it has had success in facial image matching with the ICA technique [7];

$$E_{cos}(x, \hat{x}) = \frac{-x^T \hat{x}}{\|x\| \|\hat{x}\|}, \tag{16}$$

where x is the original image and $\hat{x}$ the reconstructed image.
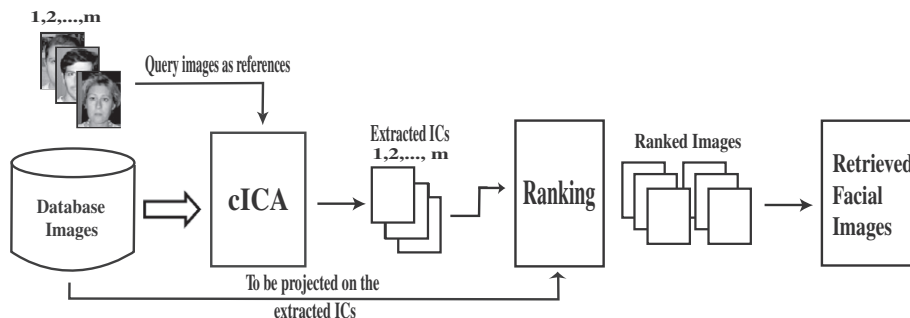

Fig. 2. Architecture of the cICA-based CBIR system.

4. Experimental setups and results

4.1. Facial databases

We performed experiments on some available public facial image databases as depicted in Fig. 3, including the ORL face database [37], the Yale face database [3], the CalTech frontal face database [1], and the AR database [28]. The ORL database consists of 400 images of 40 individual persons in various poses and expressions, normalized to (128×128). The Yale database consists of 165 images of 15 individual persons, each with a total of 11 expression images. The Yale images have been pre-processed by extracting the regions of interest and have been resized to (128×128). The CalTech database includes 342 images of 19 persons. Each image has been resized to (128×128) and has a different background. The AR database contains over 4000 color images of 126 persons. In our experiment, we choose 3094 images corresponding to 119 persons (65 men and 54 women) who participated in two photo sessions, separated by 14 days, and convert them to gray scale images of size (96×96). The front-view faces had various facial expressions, illumination conditions, and occlusions (sun glasses or scarf). The total number of pictures taken of each person was 26 (each section contained 13 pictures).

4.2. Homogeneous queries

Homogeneous query concerns involve querying facial images of the same subjects with variations in pose, expression, and illumination. Although our system has been designed to function in both single and multiple retrieval modes, using only a single face as a query image is not adequate for retrieving all of the images of the same subject under various conditions. Fig. 4(a) shows the results of a single query using a facial image of the 18th person in the ORL database. Note that, to conveniently present the results, we display the set of retrieved images in the form of a matrix and use a (row, column) pair as the index of each image. The retrieved image at the top-left corner has the highest rank, while that at the bottom-right corner has the lowest. In the database, there were a total of 10 images of each individual. In the case of a single query, our system was able to retrieve eight images out of the top 10 retrieved images. The images that were left out of the top 10 images are shown at (3,2) and (4,3) in Fig. 4(a). Those images are of the same individual but with his head tilted to the right. The query image given in Fig. 4(a) is unable to identify the features presented in these two images.

In Fig. 4(b), we used two query images of the same person. One depicts the individual with a left tilt and the other depicts the individual with a right tilt. In the obtained results, all 10 relevant images were successfully retrieved from the database with the highest ranking. This shows that the cICA-based retrieval technique can utilize the features of two different poses of the same subject.

Another aspect of the homogeneous query addresses the occlusion problem in facial image retrieval. Usually, the occlusion problem is solved by using the local feature approach: rather than considering the whole image, this method recognizes the face in parts, only considering important areas such as the lip, chin, or eyes. The use of local features has yielded some success with partial occlusions and local distortions. Some typical feature-based local representations include local feature analysis (LFA) [30], local nonnegative matrix factorization (LNMF) [23], and locally salient ICA (LS-ICA) [19]. Alternatively, in our system, the reconstructed images are the linear summations of independent components extracted through cICA. The deficient parts can be fused according to the cues of other images to overcome the occlusion. We performed occlusion testing on the AR database with the homogeneous query in Fig. 5. The two pictures are of the first person in the AR database: one in
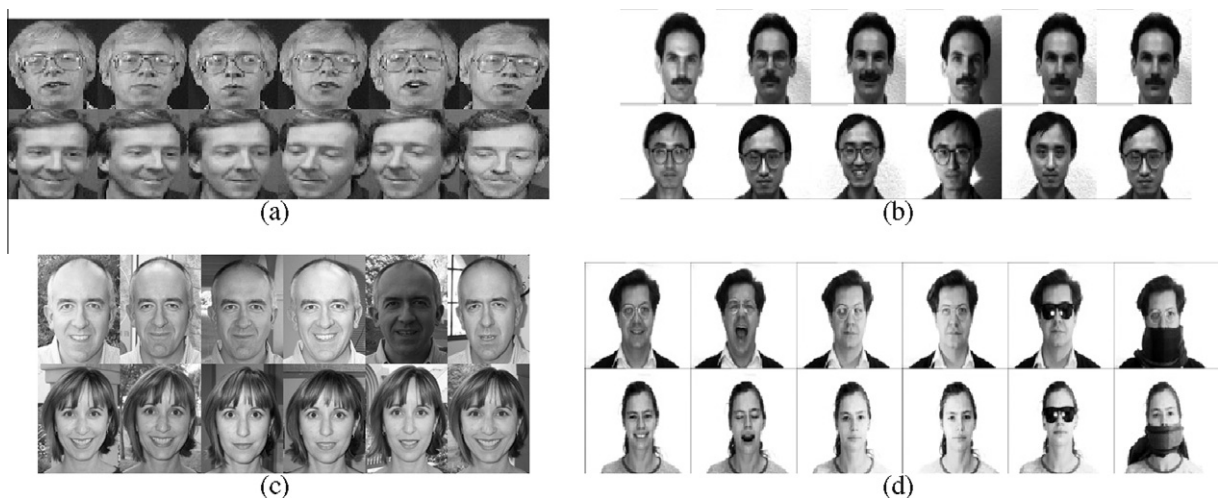

Fig. 3. Example images from the (a) ORL, (b) YALE, (c) CalTech, and (d) AR facial databases.
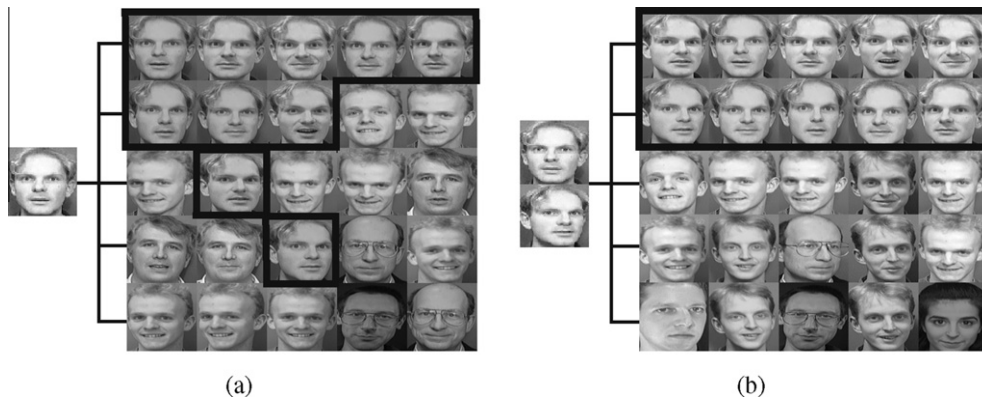
Fig. 4. Homogeneous queries with the ORL database. (a) A single query in which the system acts as a facial recognition system. Of the 10 images identified in the boxes, only eight were found in the first ten ranks. (b) Using two images of the same individual with different poses as the query images. All 10 images were retrieved in the first 10 ranks.
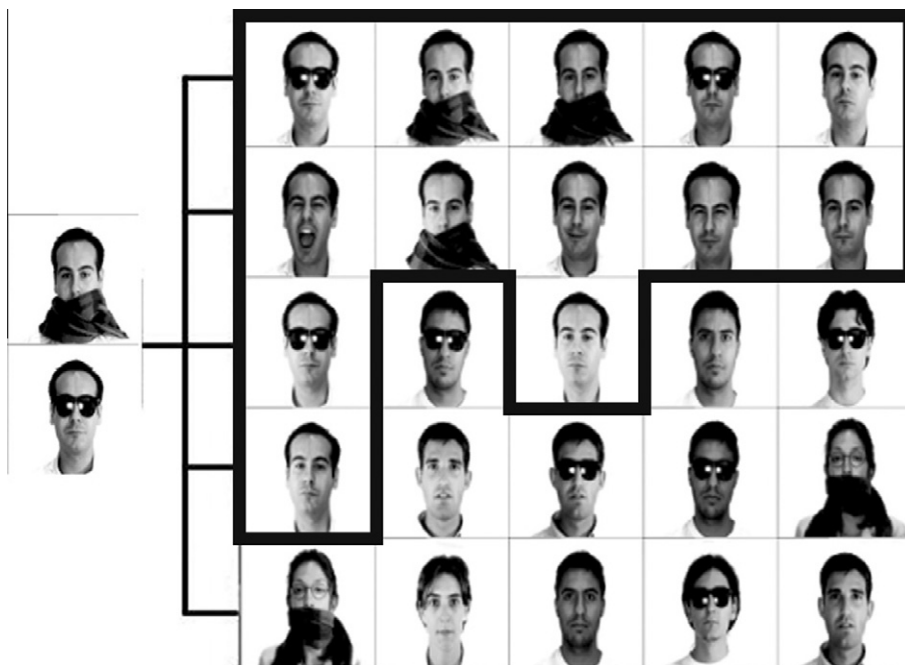

Fig. 5. Homogeneous queries in which the faces are occluded by glasses or a scarf.

which a scarf is hiding the mouth and chin and another in which glasses are worn. Looking at the outcomes, we obtained the 11 highest ranking images corresponding to the person whom we wanted to retrieve from the database. If we observe all the relevant images within the first 25 images, even with insufficient query information, we obtain the seven complete faces at $(1,5)$, $(2,1)$, $(2,3)$, $(2,4)$, $(2,5)$, $(3,3)$, and $(4,1)$, respectively.

4.3. Heterogeneous queries

The heterogeneous query deals with a retrieval using combined information of different facial images. In a search engine of facial images, providing a compound query with multiple subjects is important because it brings to users the flexibility to formulate a search for desired images. The typical application of such a query involves the identification of suspects in a criminal investigation, organization of photo archives, searching for people on the internet, etc. In this section, we show that performing a heterogeneous query is feasible in our proposed approach. In our retrieval system, a set of ICs extracted with a cICA spans the independent bases that represent the combined facial contents of the query images. By reconstructing the database images from these ICs, we can ascertain the contribution of each database image to multiple query images and

thereby rank the database images to answer the heterogeneous query. Fig. 6 depicts the experimental results tested for two heterogenous queries. In the first query (Fig. 6(a)), two images of the 4th and 27th persons of the ORL database were used. One subject is wearing spectacles and the other is not. In the retrieved images, we obtained 13 relevant images corresponding to the two individuals. Note that the retrieved images of the second subject with a feature borrowed from the first subject's face (e.g., wearing spectacles) appear at $(2,5)$ and $(5,1)$ among the first 25 retrieved images. The system also returned images of individuals with and without spectacles who carry some facial features similar to those of the individuals in the query images.

In Fig. 6(b), a similar experiment is completed with two different individuals (the 7th and 14th persons in the CalTech database). Two features are shared between the two faces rather than just one. One person is bald and the other has a beard. We obtained a total of 12 relevant images with different background conditions, facial expressions, and illuminations. The mixture characteristics of bald and having a beard commonly appear in the lower ranked group, including the six facial images at $(3,2)$, $(3,4)$, $(4,2)$, $(4,4)$, $(4,5)$, and $(5,4)$.

For comparison, in Fig. 7, we evaluated the differences between the single image-based queries performed sequentially and the heterogeneous query. For each single query, the first 15 images with the highest ranks are shown. By gathering the results of two independent queries using the facial images of the 2nd and 26th persons in the ORL database, we obtained a total of 30 facial images. Similarly, for the heterogeneous queries, the same number of retrieved images was chosen. In the latter case, a total of 18 relevant images were retrieved, approximately equal to the number of the 19 relevant images obtained in the former case. However, with the heterogeneous query, our CBIR system retrieved mixed features (e.g., wearing glasses and beard) at $(4,1)$, $(5,4)$, $(5,5)$, $(6,2)$, and $(6,5)$, revealing a significant advantage of our retrieval system in that it allows the compound query to be based on the image contents.

(a)   (b)
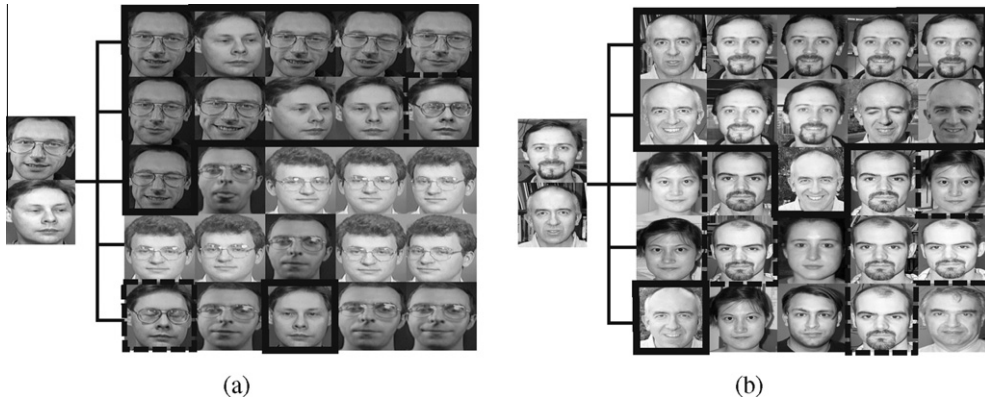
Fig. 6. Heterogenous queries with the ORL and CalTech databases. (a) Only one person wears spectacles. (b) One person is bald and the other has a beard. Here, the facial images with mixed features are indicated by a dashed line.
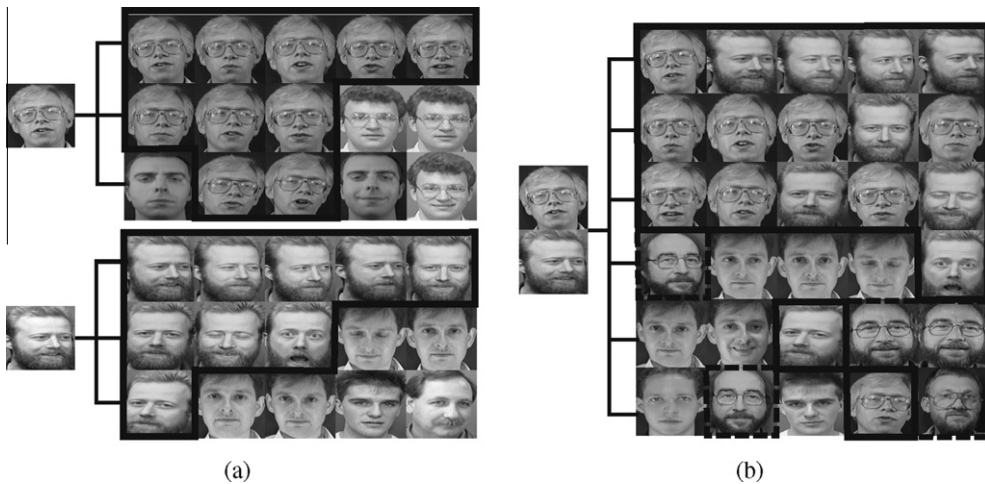

(a)   (b)

Fig. 7. Comparison results between (a) single image-based queries performed independently and (b) a heterogeneous query with two facial images. Here, the facial images with mixed features are indicated by a dashed line.

4.4. Performance analysis

For all of the experiments, we performed 100 simulations with random query formulations. System performance was evaluated using the measures of precision and recall [9], defined as

$$Precision = \frac{N_{RL}}{N_R}, \text{ and} \tag{17}$$

$$Recall = \frac{N_{RL}}{N_{RD}}, \tag{18}$$

Table 3
Accuracy performance of our facial image retrieval algorithm compared with that of the PCA, ICA, and LNMF algorithms for a single image-based query with the ORL+Yale database.

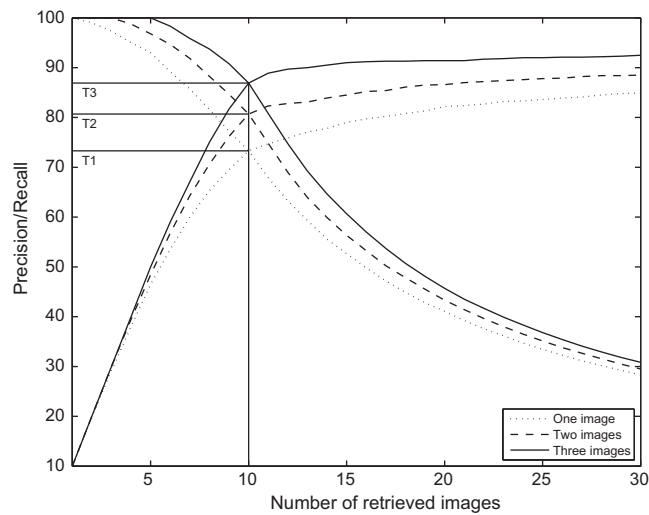|	PCA	ICA	LNMF	Our approach
Accuracy rate	75.7%	68.1%	70.1%	72.6%

Fig. 8. Performance of cICA-based CBIR for homogeneous queries with the ORL+Yale database.
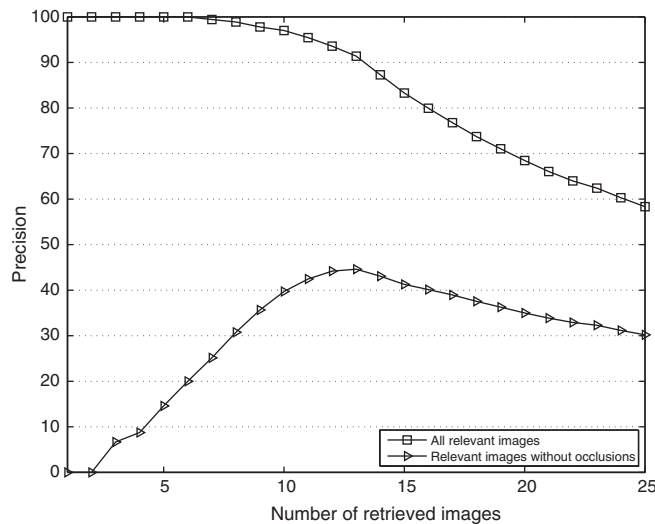

Fig. 9. Performance of the cICA-based CBIR for homogeneous queries with occlusion.

where N_{RL} is the number of relevant images among the retrieved images, N_R is the number of retrieved images, and N_{RD} is the total number of relevant images in the database. In an example in Fig. 4, we have Precision $= \frac{10}{25}$ and Recall $= \frac{10}{10}$. Note that when N_R equals N_{RD}, the two measures meet at a break-even point that indicates the accuracy of the system.

We tested our homogeneous experiments in an ORL/Yale combination database. For the Yale database, we removed one picture for each person to have the same value of N_{RD} in both databases. The facial images of the first 20 persons were randomly selected to create a query set. First, we compared our approach with some typical face recognition approaches including using PCA [32], ICA (using Architecture I) [7], and LNMF [23] with a single image-based query. Note that in all face recognition approaches and in the whitening of our approach (for both homogeneous and heterogeneous queries), 200 basic components were extracted to represent all of the database images. We used the accuracy at a break-even point to evaluate the retrieval performance of all of the examined approaches. The experimental results depicted in Table 3 indicate that, in terms of a single image-based query, our approach provides good retrieval accuracy that is competitive with those of other face recognition approaches.

However, our facial image retrieval system can perform a homogeneous query with more than an image. The evaluation performances of homogeneous queries consisting of one, two, and three images are depicted in Fig. 8. In this figure, T1, T2, and T3 present the break-even points of the system corresponding to the one-, two-, and three-image queries, respectively. As mentioned in Table 3, the system achieves an accuracy of 73% for the single image queries. With the compound queries, two- and three-image queries, the accuracy of our system improves to 81% and 87%, respectively, which is higher than those of all of the conventional face recognition systems in the previous experiment.
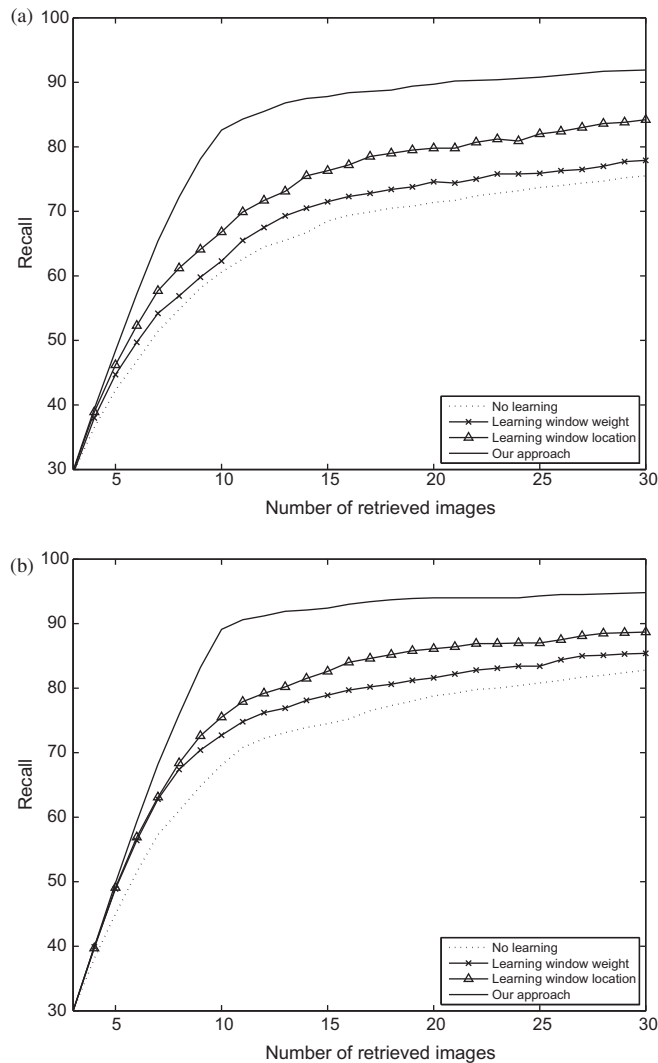

Fig. 10. Performance comparisons between our algorithm and the ICA-based approach [9] with (a) a two-image-based homogeneous query and (b) a three-image-based homogeneous query in the ORL database.

Table 4
Precision performance of our facial image retrieval algorithm compared with that of the ICA-based approach [9] with heterogeneous queries for the first 25 retrieved images.

	No learning	Learning window weight	Learning window location	Our approach
Precision rate	60.2%	68.2%	72.7%	74.9%

Previously, we discussed the advantages of a homogeneous query with our system to recognize complete facial images in occlusion conditions using a compound query. We performed this performance evaluation on the AR database by randomly choosing a set of faces wearing glasses or a scarf to create the query sets. Among all of the relevant images, we selected the facial images without occlusion and plotted them, as shown by the curve in Fig. 9. This figure shows that by using a combination of multiple images, we can retrieve complete faces using only partial information.

Continuously, we report the quantitative evaluation of our approach compared with the conventional works designed to retrieve images using the compounding contents of query images. Because there have been very few attempts to support the compound query, the algorithm presented by Basak et al. [9] has been implemented to compare with our algorithm. In their system, one needs to select a set of windows at a fixed location within the images, represent each window as a signal of length w^2 where w is the size of the window, and apply an ICA to extract the ICs from these signals. Next, each window within the database image is reconstructed from some selected ICs, and the reconstruction errors are utilized to compute the ranks of the images. There were a total of 16 windows for each image, and each window was 32×32. Two learning methods have been implemented to improve the performance of this algorithm. In the first method, the windows are arranged in a grid and are assigned different weights depending on the importance of each window. The weights were learned by using a regression fit. In the second method, the window locations are searched using a genetic-based algorithm in which each window has the same weight. In our implementation, we used 20 training images to learn the weight and location of each window according to the comments of the authors. Our approach was compared against their algorithm involving no learning (e.g., randomly choosing windows), including learning for window weight, and including learning for window location with two-image- and three-image-based homogeneous queries in the ORL databases. Fig. 10 shows the performance of our algorithm compared with that of the approach of Basak et al. The entire facial image without any information loss was used by our algorithm to improve the retrieved results.

To assess the facial retrieval performance with heterogeneous queries, we conducted experiments with the Caltech database. Each query set was randomly formulated with two or three images of different subjects. The same method used to measure the performance of the CBIR system of Basak et al. was applied to qualify the performance of our experiment. The precision scores derived from the first 25 images of our algorithm and the ICA-based algorithm of Basak et al. are given in Table 4. As the results show, our approach is superior to their algorithm, as ours does not have the need to learn the local windows.

5. Conclusion

In this paper, we introduced a new facial image retrieval system based on the cICA algorithm, which utilizes all of the information in query images to retrieve relevant images. Our system supports single and multiple query requirements. The use of multiple homogeneous queries on the same subject improves the retrieval performance and overcomes partial occlusion. When using multiple heterogeneous queries on different subjects, the compound queries can retrieve faces with the combined characteristics appearing in the query faces. The experimental results from four facial databases have shown that our algorithm achieves success for all of these queries without the requirement of a learning stage.

6. Discussion

Despite the advantages of allowing for compound queries and having no requirement for a learning stage, our algorithm has some limitations. Similar to other image retrieval algorithms, our feature extraction process is sensitive to scale, rotation, and translation (affine transformation). In future work, integrating our algorithm with the analytical Fourier–Mellin transform [42] may allow us to normalize each face to the same size, direction, and location before processing. Hence, the algorithm would be more invariant with the affine transformation of the facial images. Another interesting direction is to combine our algorithm with some conventional image features to target a hybrid system. Among the image features developed in previous studies, the textural features play an important role in representing compact information from database images [13,25]. By estimating these features and by using them with our cICA-based system, we believe that we can improve the retrieval performance of our proposed approach.

For practical applications, our system could be adopted for use in criminal investigation such as for the identification of possible suspects. For instance, to reconstruct a face from a witness description, a computer system might be used. The system first randomly offers a set of initial faces, and then, the witnesses select the faces that look similar to the suspect. From a set of selected images, our system with a heterogeneous query may be useful to identify the images similar to the selected faces to produce a new set of facial images. The process of searching a new set of images followed by further selection would

be repeated several times. Finally, the best selected faces can be combined to produce the suspect face. In similar ways, our retrieval system can be used to retrieve medical images such as X-ray, chest radiology, and CT images. However, we should note that further validation is required for these applications by using proper databases and extensive experiments. Additionally, the references for the cICA could be better designed so that more valuable information could be exploited from the query images to improve the retrieval accuracy.

Acknowledgements

References

[1] The frontal face database <http://www.vision.caltech.edu/html-files/archive.html>.
[2] The image search engine Tineye <http://www.tineye.com/>.
[3] The Yale face database <http://www.cvc.yale.edu/projects/yalefaces/yalefaces.html>.
[4] A.F. Adate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D face recognition: a survey, Pattern Recognition Letters 28 (14) (2007) 1885–1906.
[5] B. Ahmed, T. Rasheed, Y.K. Lee, S. Lee, T.-S. Kim, Facial image retrieval through compound queries using constrained independent component analysis, in: Proceedings of IEEE International Conference on Tools with Artificial Intelligence, 2007, pp. 544–548.
[6] M. Balsi, G. Filosa, G. Valente, P. Pantano, Constrained ICA for functional magnetic resonance imaging, in: Proceedings of the European Conference on Circuit Theory and Design, 2005, pp. 67–70.
[7] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face recognition by independent component analysis, IEEE Transactions on Neural Networks 13 (6) (2002) 1450–1464.
[8] I. Bartolini, P. Ciaccia, M. Patella, WARP: accurate retrieval of shapes using phase of Fourier descriptor and time warping distance, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (1) (2005) 142–147.
[9] J. Basak, K. Bhattacharya, S. Chaudhury, Multiple exemplar-based facial image retrieval using independent component analysis, IEEE Transactions on Image Processing 15 (12) (2006) 3773–3783.
[10] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (4) (2002) 509–522.
[11] C. Carson, S. Belongie, H. Greenspan, J. Malik, Blobworld: image segmentation using Expectation-Maximization and its application to image querying, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (8) (2002) 1026–1038.
[12] R. Datta, J. Li, J.Z. Wang, Content-based image retrieval: approaches and trends of the new age, in: Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval, 2005, pp. 253–262.
[13] D. Feng, W.C. Siu, H.J. Zhang, Multimedia information retrieval and management: technological fundamentals and applications, Springer, 2003.
[14] H. Hu, ICA-based neighbor preserving analysis for face recognition, Computer Vision and Image Understanding 112 (3) (2008) 286–295.
[15] D.S. Huang, J.X. Mi, A new constrained independent component analysis method, IEEE Transactions on Neural Networks 18 (5) (2007) 1532–1535.
[16] A. Hyvärinen, J. Karhunen, E. Oja, Independent component analysis, John Wiley and Sons, 2001.
[17] N. Katsumata, Y. Matsuyama, Database retrieval for similar images using ICA and PCA bases, Engineering Applications of Artificial Intelligence 18 (6) (2005) 705–717.
[18] A. Khaparde, B.L. Deekshatulu, M. Madhavilatha, Z. Farheen, S. Kumari, Content based image retrieval using independent component analysis, International Journal of Computer Science and Network Security 8 (4) (2008) 327–332. V.
[19] J. Kim, J. Choi, J. Yi, M. Turk, Effective representation using ICA for face recognition robust to local distortion and partial occlusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (12) (2005) 1977–1981.
[20] S. Kiranyaz, M. Birinci, M. Gabbouj, Perceptual color descriptor based on spatial distribution: a top–down approach, Image and Vision Computing 28 (8) (2010) 1309–1326.
[21] R. Krishnapuram, S. Medasani, S. Jung, Y. Choi, R. Balasubramaniam, Content-based image retrieval based on a fuzzy approach, IEEE Transactions on Knowledge and Data Engineering 16 (10) (2004) 1185–1199.
[22] J. Li, J.Z. Wang, Real-time computerized annotation of pictures, IEEE Transactions on Pattern Analysis and Machine Intelligenc 30 (6) (2008) 985–1002.
[23] S.Z. Li, X.W. Hou, H.J. Zhang, Q.S. Cheng, Learning spatially localized, parts-based representation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, pp. 207–212.
[24] 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) (2004) 1117–1127.
[25] Y. Liu, D. Zhang, G. Lu, W. Ma, A survey of content-based image retrieval with high-level semantics, Pattern Recognition Journal 40 (1) (2007) 262–282.
[26] W. Lu, J.C. Rajapakse, Approach and applications of constrained ICA, IEEE Transactions on Neural Networks 16 (1) (2005) 203–212.
[27] J.A. Martín, H.M. Santos, J. Lope, Orthogonal variant moments features in image analysis, Information Sciences 180 (6) (2010) 846–860.
[28] A.M. Martinez, R. Benavente, The AR face database, Tech. rep., CVC Technical Report 24, 1998.
[29] A. Mohamed, Y. Weng, J. Jiang, S. Ipson, An efficient face image retrieval through DCT features, in: Proceedings of Signal and Image Processing, 2008, pp. 189–191.
[30] P.S. Penev, J.J. Atick, Local features analysis: A general statiscal theory for object representation, Network: Computation in Neural Systems 7 (3) (1996) 477–500.
[31] S.-H. Peng, D.-H. Kim, S.-L. Lee, C.-W. Chung, A visual shape descriptor using sectors and shape context of contour lines, Information Sciences 180 (16) (2010) 2925–2939.
[32] A. Pentland, R.W. Picard, S. Sclaroff, Photobook: content-based manipulation of image databases, International Journal of Computer Vision 18 (3) (1996) 233–254.
[33] J. Perkiö, A. Hyvärinen, Modelling image complexity by independent component analysis, with application to content-based image retrieval, in: Artificial Neural Networks ICANN 2009, 2009, pp. 704–714.
[34] E.G.M. Petrakis, A. Diplaros, E. Milios, Matching and retrieval of distorted and occluded shapes using dynamic programming, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (11) (2002) 1501–1516.
[35] T. Rasheed, Y.K. Lee, S.Y. Lee, T.-S. Kim, Attenuation of artifacts in EEG signals measured inside an MRI scanner using constrained independent component analysis, Physiological Measurement 30 (4) (2009) 387–404.

[36] Y. Rui, T.S. Huang, M. Ortega, S. Mehrotra, Relevance feedback: a power tool for interactive content-based image retrieval, IEEE Transactions on Circuits and Systems for Video Technology 8 (5) (1998) 644–655.
[37] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, 1994, pp. 138–142.
[38] J.A. Santos, C. Ferreira, R.S. Torres, M.A. Gontalves, R.A.C. Lamparelli, A relevance feedback method based on genetic programming for classification of remote sensing images, Information Sciences (2010). DOI:10.1016/j.ins.2010.02.003.
[39] T.B. Sebastian, P.N. Klein, B.B. Kimia, Recognition of shapes by editing their shock graphs, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (5) (2004) 550–571.
[40] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1349–1380.
[41] J.K. Wu, Y.H. Ang, P.C. Lam, S.K. Moorthy, A.D. Narasimhalu, Facial image retrieval, identification, and inference system, in: Proceedings of the first ACM international conference on Multimedia, 1993, pp. 47–55.
[42] H. Yu, M. Bennamoun, Complete invariants for robust face recognition, Pattern Recognition Journal 40 (5) (2007) 1579–1591.
[43] X. Zhang, Y. Gao, Face recognition across pose: a review, Pattern Recognition Journal 42 (11) (2009) 2876–2896.
[44] J. Zou, Q. Ji, G. Nagy, A comparative study of local matching approach for face recognition, IEEE Transactions on Image Processing 16 (10) (2007) 2617–2628.