A NEW BAG OF PIXELS APPROACH BASED ON EDGE DETECTION TO FACE IMAGE RETRIEVAL

Mir Hossein Dezfoulian
Image Processing Lab, Computer Engineering Department
Bu-Ali Sina University
Hamedan, Iran

Mostafa Parchami
Image Processing Lab, Computer Engineering Department
Bu-Ali Sina University
Hamedan, Iran

Saman Bashbaghi
Image Processing Lab, Computer Engineering Department
Bu-Ali Sina University
Hamedan, Iran

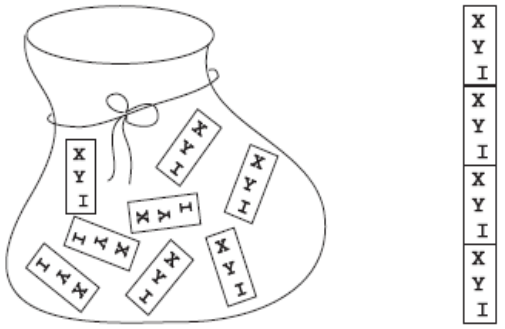
Abstract

Daily increase in number of digital images over the internet has increased the need of searching among digital image databases. As a special case, Face image retrieval is important in many applications. In this paper, a new method is proposed for human face retrieval. Also this method can be used in face recognition systems. Images in this method are represented as a bag of pixels and for this purpose images are converted to grayscale, then after calculating first order derivation of the grayscale image, Canny edge-detector is applied to highlight face organs like mouth, eyes, eyebrows and ears. Finally pixels of obtained image are clustered into seven clusters. To test the proposed method, CIE Put Image Database is used. Experimental results show that the proposed method has a good performance and high accuracy, such that its recognition rate for face images is 100% over the database.

Keywords: Face image retrieval; bag of pixels; K-means clustering; edge detection; canny filter.

1. Introduction

One of the most important parts of machine vision and image processing systems is visual data and image representation. Image representation has a direct impact on different parts of system such as segmentation, recognition and tracking. For example main changes in image that changes representation of image, has a great effect on recognition. How to represent image also is important in terms of space and time complexity of retrieval procedure. Also using less memory to represent image is desired but this should not reduce accuracy of algorithm applied. Bag of pixels is an image representation method that uses a set of N pixels with coordination (X, Y) and gray-level (I) to represent the image. In this method each image is represented by a bag of ordered triple (X, Y, I). Color images can be represented by bag of 5-tuples (X, Y, R, G, B) [1]. In contrast with other vector-based methods, bag of pixels is based on n-tuple whereas other methods store these values in a specified order. Figure 1 (b) shows vector-based method that represents image with a vector of values and Figure 1 (a) shows bag of pixels representation of images. As shown in Fig. 1. in bag of pixels, order of pixels is not important while in vector-based methods pixels have a predefined order and a change in this order changes the image. Therefore pixels in bag of pixels have no priority or order and they can be reordered so this method is robust against changing order of pixels.



(a) (b)

Figure 1. (a) Representation of image using vector for grayscale images. (b) Representation of image using bag of pixels for grayscale images.

In [2, 3] a set of pixels (X, Y, I) are used to visual representation of images. A critical problem in this method of representation is called correspondence [2, 4, 5]. For example it is possible that the relationship between a n-tuple in bag A (representation of image A) and a n-tuple in bag B (representation of image B) that should that should have correspondence, is ambiguous. Several different methods are proposed to address this problem, such as models based on physics [3, 5]. Other methods are proposed to improve models based on appearance, like converting images to vectors that is used to estimate and apply geometric adjustments [6, 7, 8, 9]. Other similar methods [8, 10, 11, 12], use calculating local correspondence to cover geometric space and disambiguate the underlying geometric model. A suitable approach is to specify an optimization criterion to show a relationship such as minimizing covariance of data [13] or minimizing length of description vector of image [10]. Approach proposed in [1] that is based on convex program, computes correspondence of all images at once in contrast with previous methods that calculate correspondence of pair of images.

Available approaches used in content-based image retrieval are not applicable to be used in this context due to similarity of shape of face of humans. All faces almost have the same shape and same basic elements and difference of human faces is more complex than other images. In most image retrieval systems, methods of human face recognition (HFR) are used. In these methods, images are represented as a large vector and PCA or LDA is applied to reduce dimensions but this process has a high time complexity and reduces efficiency of the system [14, 15, 16]. In retrieval systems, speed of retrieval is critical and retrieval time should be acceptable for user. In this paper, we propose a new method for face image retrieval that is based on bag of pixels and its response time is acceptable for retrieval task. A new method proposed to find interest points of human faces to decrease space complexity of algorithm and extracted interest points are represented as bag of pixels. Yet another step is done over these interest points to obtain indexed database. For this purpose, clustering of interest points is done to obtain a descriptor vector for each face. Experimental results show that this method has a high accuracy such that it could be used in human face recognition systems. The rest of this paper is organized as follows: problem definition, proposed method, experimental results and conclusion.

2. Problem Definition

Development of digital imaging devices made them available to everyone and caused a great growth in number of digital images over internet. This growth introduced a new problem in computer science: searching images among huge image databases over internet that can't be done non-automatically. Since 1990 searching through these databases and image retrieval has become an important research subject [17]. Face image retrieval is trickier than content-based image retrieval due to similar geometric shape and texture of face images. This fact makes face image retrieval systems become more complex than other image retrieval systems. In this paper, the goal is to design and implement fast and accurate face image retrieval that takes a query face image and retrieves similar faces stored in the database. As mentioned due to the high accuracy of the proposed method, it can be used in face recognition systems.

3. Proposed Method

In image processing systems and especially in image retrieval systems the first step is to choose the image representation method.

Here, bag of pixels is used to represent images because of its flexibility and commutative representation of pixels. The next step is to choose pixels we want to represent them in the bag of pixels. This is done during the training phase of the algorithm. In the training phase of the algorithm, an indexed database of stored images is created to make comparison faster and more accurate. This indexed database is used in the retrieval phase of the algorithm in which a query face image is given by the user and relevant images are retrieved. In the following sections these two phases are described.

3.1. Training Stage

Figure 2 shows the flowchart for the training phase of the proposed method.

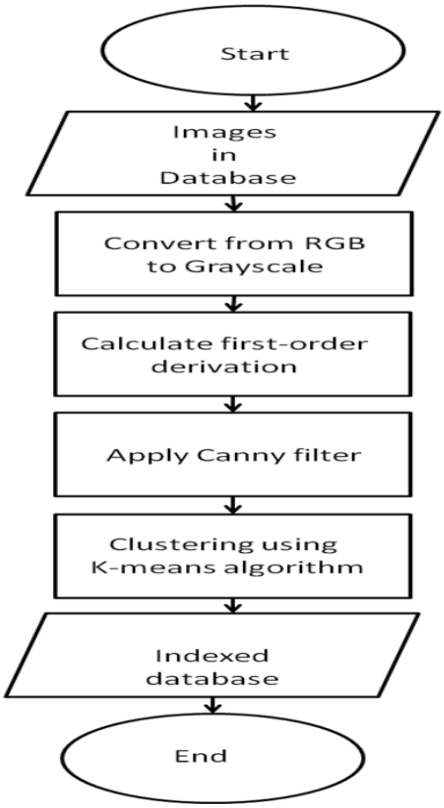


Figure 2. Flowchart of training phase.

Face images are similar in terms of geometric shapes and are homogenous. This implies that random selection of pixels as bag of pixels is not applicable. Additionally due to their homogeneity, many pixels in the face do not have any information and have low information gain. Therefore choosing these pixels will cause low accuracy of the algorithm. Pixels chosen for bag of pixels should represent face such that it is distinctive from face of other humans; these pixels should be selected wisely and should cover important points of face like ears, eyes and other parts of face that are different in face of different people. These pixels should give a complete and accurate description of a face such that it is distinct from others' faces and different aspects of a face should be included in this description. Most points used in such systems are points of edges of the image but in face images, edges do not have much information and most of the information is in organs of the face. Edge pixels are not enough because they do not have all information needed for retrieval task. For example if the person is wearing glasses or grew mustache, edges will change and retrieved images might be irrelevant. In this method, these points are extracted using known edge-detectors that are fast and accurate like canny and Sobel and other processes to extract important parts of face that helps the system to retrieve similar faces.

For each image in the database, first it is converted to grayscale to extract interest points of face image. The resulted image is shown in figure 3 (b) After converting image into grayscale, a gradient filter is applied to compute first-order derivation of the grayscale image. This step gives us pixels that their neighbor pixels have high difference.

Filter applied here is $[-1 \quad 0 \quad 1]$ and $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ and result of this step is shown in figure 3 (c) Pixels obtained from this step are not good points to use as bag of pixels because most of them are not useful and have no information to help the retrieval. After this step, canny filter is applied to the image to obtain pixels that can be represented as bag of pixels. Canny filter is used for edge detection and calculates first-order derivation of the image in its first step. Applying this filter does the magic, because it eliminates regions with linear change in their gray-levels and the resulting pixels are pixels of important parts of face such as ears, eyes, nose and eyebrows. Resulted image of this step is shown in figure 3 (d) and as shown, pixels of important organs of face are highlighted. This is result of first step of canny that calculates first-order derivation of the image. Now pixels that can describe face are obtained and can be used as bag of pixels. Thus bag of pixels is created using coordination of pixels and values of H, S and I channels in HIS color model so bag of pixels for each image is pixels obtained after applying canny filter and their values in HSI (X, Y, H, S, I).

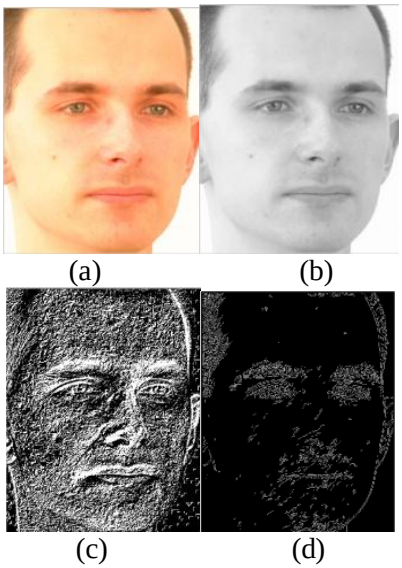


Figure 3. (a) Color image. (b) Grayscale image. (c) First-order derivation. (d) Edge-detection using canny filter.

Number of obtained pixels is too large and comparison of these pixels to retrieve relevant images takes too much time. To reduce number of comparison needed, K-means clustering is used. At this point, image is described using a bag of 5-tuples that is pixels obtained after applying canny filter. This bag can completely describe the face, but it has too much information and its dimensionality can be reduced. If image was described using vector, PCA or LDA could be applied, but here we can make use of clustering technique to cluster these points into seven clusters. Each cluster represents one part of face, for example one represents points of nose, and one represents points of right eye and so on. This way, bag of pixels can be reduced to seven cluster Centroids. Now comparison for retrieving face images can be done using center of these clusters instead of comparing all pixels in the bag of pixels. This step converts each image into seven 5-tuples that are center of clusters. These seven 5-tuples are stored to create the indexed database.

3.2. Retrieving Stage

This phase, user gives a query image and system should retrieve similar face images stored in the database. In the previous phase, the indexed database is created and will be used to find similar images. For this purpose, first the query images should be described using seven 5-tuples and then it can be compared with indexed database to retrieve similar images.

The first step is to convert the query image into grayscale image to extract interest points used to create bag of pixels. After converting face image to grayscale image, first-order derivation is applied and then canny filter applied to obtain pixels for bag of pixels. This process is the same as the process done during the training phase. After creating bag of pixels, pixels in this bag are clustered into seven clusters and center of clusters are used as descriptor of the face.

At this point, the query face image is described using seven 5-tuples and is ready to be compared with indexed database. To calculate distance between the query image and images in database, Hausdorff distance between seven 5-tuples of the query image and seven 5-tuples of the ith image in the database is calculated. Hausdorff distance formulation is given by equation 1 [19].

$$d_H(X, Y) = \{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \} \quad (1)$$

Where in Eq. (1), $d_H(X, Y)$ is Hausdorff distance between sets X and Y that X and Y are subsets of a metric space, sup and inf represent supremum and infimum respectively and d(x, y) is distance between x and y, members of sets X and Y respectively.

Finally images are sorted ascending by distances and top images that have less distance with the query image are retrieved.

4. Experimental Results

In this section, the proposed face retrieval system is evaluated to check its efficiency and accuracy. To evaluate the proposed method, CIE Put image database [20] is used. CIE Put image dataset is a standard dataset that is used in face recognition and retrieval. This database consists of images of face of 100 people with different angles, lighting conditions and different accessories. To test the proposed method over this dataset, 60 image of each person are used to created indexed database and the rest of images are used as query images.

$$CRR = \frac{\#\,\text{First relevant retrieved images}}{\#\,\text{Total queries}} \quad (2)$$

Recognition rate that is calculated using Eq. (2), for the proposed method over CIE dataset is 100%. Figure 4 shows accuracy chart for the proposed method. As shown in this graph, for each query at least 5 relevant images are retrieved in the first ten retrieved images and for 75% of queries at least 8 relevant face images are retrieved. This graph implies that the proposed method has a high accuracy and can be used for human face recognition.

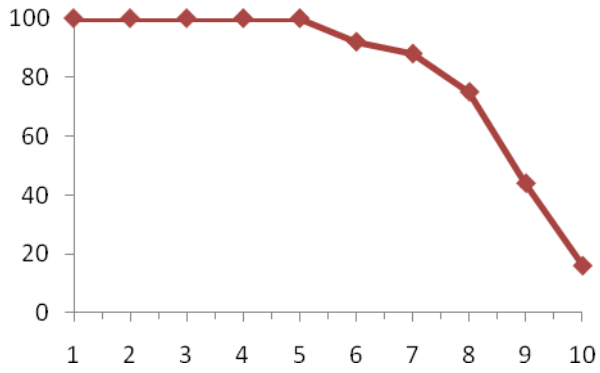


Figure 4. Recognition rate for first 1 retrieved face images.

In Figure 5 (a) query and top ten relevant retrieved images are shown. All ten retrieved images are relevant and recognition rate and accuracy for this retrieve is 100%. As this figure shows, this method can tolerate some rotation of face images.

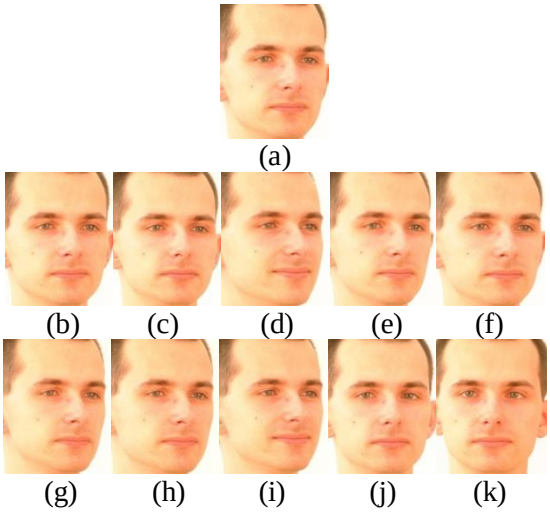


Figure 5. An example of retrieval. (a) query image. (b-k) retrieved images.

References

[1] Jebara, Tony, Images as Bag of Pixels, International Conference on Computer Vision - ICCV , pp. 265-272, 2003.
[2] Gold, S., Lu, C.P., Rangarajan, A., Pappu, S., Mjolsness, E., New algorithms for 2D and 3D point Matching, Pose estimation and correspondence, In Neural Information Processing Systems 7, 1995.
[3] Nastar, C., Moghaddam, B., Pentland, A., General-ized image matching, Statistical learning of physically-based deformations, In European Conference on Computer Vision, 1996.
[4] Beymer, D., Poggio, T., Image representation for visual learning. Science, 272, 1996.
[5] Sclaro, S., Pentland, A., Modal matching for correspondence and recognition, IEEE Transactionson Pattern Analysis and Machine Intelligence, 15(6), 1993.
[6] Murase, H., Nayar, S., Learning object models from appearance. In Proceedings of the AAAI, 1993.
[7] Moghaddam, B., Pentland, A., Probabilistic visual learning for object detection, In International Conference on Computer Vision, 1995.
[8] Frey, B., Jojic, N., Estimating mixture models of images and inferring spatial transformations using the EM algorithm. In Computer Vision and Pattern recognition, 1999.
[9] Miller, E., Matsakis, N., Viola, P., Learning from one example through shared densities on transforms. In Computer Vision and Pattern Recognition, 2000.
[10] Cootes, T., Edwards, G., Talyor, C., Active appear ance models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6) 681-685, 2001.
[11] O'Toole, A., Vetter, A., Blanz, V., Three- dimensional shape and two-dimensional surface re-ectance contributions to face recognition, An application of three-dimensional morphing. Vision Research, 39, 1999.
[12] Jones, M., Poggio, T., Hierarchical morphable mod els. In Computer Vision and Pattern Recognition, 1998.
[13] Davies, R., Twining, C., Cootes, T., Waterton, J., Taylor, C., 3d statistical shape models using direct optimisation of description length, In European Conference on Computer Vision, 2002.
[14] Tolba, A.S., El-Baz, A.H., El-Harby, A.A., Face Recognition, A Literature Review. International Journal of Signal Processing 2(2), pp. 88-103, 2006.
[15] Zhao, W., Face Recognition, A Literature Survey, ACM Computing Surveys 35(4), pp. 339 – 458, 2003.
[16] Vikram, T.N., Chidananda, G.K., Guru, D.S., Shalini, R.U., Face Indexing and Retrieval by Spatial Similarity, In International Congress on Image and Signal Processing, pp. 543 – 547, IEEE Press, Hainan, 2008.
[17] Megherbi, D.B., Miao, Y., A Distributed Technique for Recognition and Retrieval of Faces with Time-Varying Expressions, In IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, pp. 813. IEEE Press, Hong Kong, 2009.
[18] Kimmel, Ron., Bruckstein, Alfred M., On regularized Laplacian zero crossings and other optimal edge integrators, International Journal of Computer Vision, 53(3), pp. 225-243, 2003.
[19] Tyrrell Rockafellar, R., Roger, J.,Wets, B., Variational Analysis, Springer-Verlag, 2005, ISBN 3540627723, ISBN 978-3540627722, pg. 117.
[20] Kasiński, A., Florek, A., Schmidt, A., The Put Face Database, Image Processing & Communications Volume 13, Number 3-4, pp. 59-64, 2008.