Attribute and Identity Based Face Image Retrieval

Indulekha K[1], Suhara Salim[2]

[1]Calicut University, KMCT College of Engineering, NIT (PO), Calicut, India

[2] KMCT College of Engineering, Department of Computer Science And Engineering, Calicut University, NIT (PO), Calicut, India

Abstract: ***Large-scale content-based face image retrieval is an enabling technology for many emerging applications. The existing system presents a novel approach to face recognition which considers both shape and texture information and attribute information to represent face images. Existing system contain semantic cues of the face photos to improve content- based face image retrieval by constructing semantic codewords for efficient large- scale face retrieval. Here, I propose to combine identity based face image retrieval with the existing system. Proposed system contains CBIR concepts like indexing, clustering, searching (attribute searching and identity based searching) and face image retrieval. Most of the systems related to existing system are suffering from low recall problems. One solution for this problem is to construct semantic codewords by using extra textual information. So I propose the identity based face image retrieval concept.***

Keywords: Attributes, Content Based Face Image Retrieval, Identity, Texture Information

1. Introduction

The goal of face image retrieval is to find the ranking result from most to least similar face images in a face image database given a query face image. The face area is first divided into small regions from which Local Binary Pattern (LBP) histograms are extracted and concatenated into a single, spatially enhanced feature histogram efficiently representing the face image. The recognition is performed using a nearest neighbour classifier in the computed feature space with Chi square as a dissimilarity measure. [[1], [2]. Given a query face image, content-based face image retrieval tries to find similar face images from a large image database and retrieves similar images corresponds to user query image. It is an enabling technology for many applications including automatic face annotation [2], crime investigation etc. The existing system present a novel approach to face recognition which considers both shape and texture information and attribute information to represent face images. Existing system aims to utilize automatically detected human attributes that contain semantic cues of the face photos. This is used to improve content- based face image retrieval by using semantic codewords for efficient face image retrieval. Here I propose to combine identity based face image retrieval with existing system. Proposed system contains CBIR concepts like indexing, clustering, searching (attribute searching and identity based searching) and face image retrieval. Most of the systems related to existing system are suffering from low recall problem. One solution for this problem is to construct semantic codewords by using extra textual information. So I propose the identity based face image retrieval system and also analyze the effectiveness of different human attributes across datasets and finds the informative human attributes. The purpose of the system is to increase the efficiency of the face image retrieval. The contributions of this paper include:

- Propose to combine human attributes, face features and identity to construct semantic codewords. This will overcome the low recall problem in face image retrieval.
- Identifying the problem of semi-supervised face image retrieval, introducing a general coding scheme for face retrieval problem, and further improving the performance by using identity information in the coding scheme.

1.1 Related Work

This work is closely related to several different research topics, including content-based image retrieval (CBIR) [9], Face Recognition and Face retrieval The difference between face recognition and face retrieval is that face recognition task requires completely labeled data in the training set, and it uses learning based approach to find classification result while neither training set nor learning process is needed in face retrieval task, and it provides ranking result. Traditional CBIR techniques use image content like color, texture and gradient to represent images. Active Shape Model (ASM) [6] and Active Appearance Model (AAM) [4] are two most representative face alignment models. In ASM [6], a Point Distribution model captures the shape variants and gradient distributions of a set of landmark points describe the local appearance. The shape parameters are iteratively updated by locally finding the best nearby match for each landmark point. In AAM [4], the appearance is modeled globally by PCA on the mean shape coordinates (also called “shape-normalized frame”). Automatically determining the gender of a face has been an active area. Recently, some researchers have focused on bridging the semantic gap by finding semantic image representations to improve the CBIR performance. [3] and [4] propose to use extra textual information to construct semantic codeword [5] uses class labels for semantic hashing. a method for automatically training classifiers for these and many other types of attributes was proposed [4], for the purpose of searching databases of face images Using automatically detected human attributes, the system achieve excellent performance on keyword-based face image retrieval and face verification. [7] further extends the framework to deal with multi-attribute queries for keyword-based face image retrieval. [8] proposes a Bayesian network approach to utilize the human attributes for face identification. To further improve the quality of attributes, another related work is multi-attribute space. It is used to normalize the confidence scores from different attribute detectors for similar attribute search. The works demonstrate the emerging opportunities for the human attributes but are not exploited to generate more semantic (scalable) codeword. Although these works achieve salient performance on keyword-based face image retrieval and face recognition and propose to exploit effective ways to combine low-level features and automatically detected facial attributes for scalable face image retrieval. Content-based face image retrieval is closely related to face recognition problems but they focus on finding suitable feature representations for scalable indexing systems. Because face recognition usually requires substantial computation cost for dealing with high dimensional features and generating explicit classification models, it is non-trivial to directly apply it to face retrieval tasks. Meanwhile, the photo quality in consumer photos is more diverse and poses more visual variances. [4] propose a face retrieval framework using component-based local features with identity-based quantization to deal with scalability issues. To compensate the quantization loss, the CBIR adopt GIST feature with locality sensitive hashing for face image retrieval. The system [9] propose to use component-based local binary pattern (LBP), a well known feature for face recognition, combined with sparse coding and partial identity information to construct semantic codeword for content-based face image retrieval. Although images naturally have very high dimensional representations, those within the same class usually lie on a low dimensional subspace. Sparse coding can exploit the semantics of the data and achieve promising results in many different applications such as image classification and face recognition. The system [28] propose to use sparse representation for face recognition and achieve state-of-the-art performance. Taking advantages of the effectiveness and simplicity of LBP feature [10]with the superior characteristics of sparse coding on face images, Using component-based LBP combined with sparse coding to construct sparse codeword for efficient content-based face image retrieval. However, instead of using identity information that might need manual annotations, and focus on utilizing automatically detected human attributes to construct semantic-aware sparse codeword using attribute-enhanced sparse coding. In addition, we propose another orthogonal approach to further leverage attribute information by constructing attribute-embedded inverted index in the online ranking stage. Note that the proposed methods can be easily combined with the method proposed in [5] to take advantage of both identity information and automatically detected human attributes. Also, low-level feature (i.e., LBP) can be replaced by other features such as T3HS2 descriptor.

2. Existing system

Every image in the database, the system first apply face detector to find the locations of faces. Then find 73 different attribute scores. Then Active shape model is applied to locate 68 different facial landmarks on the image. Using these facial landmarks and applying barycentric coordinate based mapping process to align every face with the face mean shape [3]. In total we have 175 grids from five components including two eyes, nose tip and two mouth corners, on the aligned image using methods proposed in [4]. This optimization problem can be efficiently solved by an online optimization algorithm.

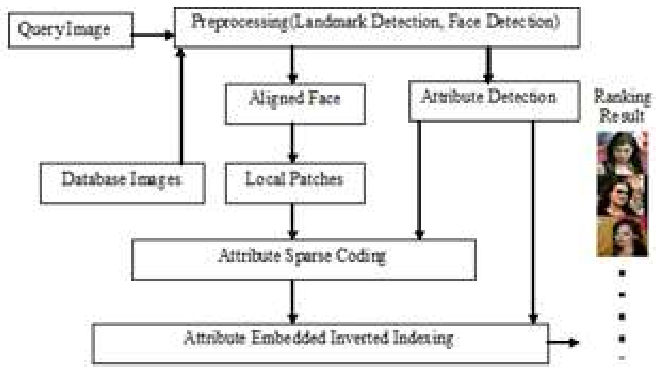

Figure 1: The Existing system framework.

From each grid, the system extract an image patch and compute a 59-dimensional uniform LBP feature descriptor as our local feature. After obtaining local feature descriptors, and quantize every descriptor into codewords using attribute-enhanced sparse coding, attribute-embedded inverted index then built for efficient retrieval. When a query image arrives, it will go through the same procedure to obtain sparse codewords and human attributes, and use these codewords with binary attribute signature to retrieve images in the index system. Figure.1 illustrates the overview of existing system.

2.1 Sparse Coding(SC)

In order to consider human attributes in the sparse representation, the system first use dictionary selection (ASC-D) to force images with different attribute values to contain different codewords. For a single human attribute, as shown in Fig. 2(b), divide dictionary centroids into two different subsets, images with positive attribute scores will use one of the subset and images with negative attribute scores will use the other. For example, if an image has a positive male attribute score, it will use the first half of the dictionary centroids. If it has a negative male attribute score, it will use the second half of the dictionary centroids. By doing these, images with different attributes will surely have different codewords. For the cases of multiple attributes, and divide the sparse representation into multiple segments based on the number of attributes, and each segment of sparse representation is generated depending on single attribute. The above goal can be achieved by solving the following optimization problem (1).

$$\min_{v} \sum_{i=1}^{n} \left\| x^{(i)} - Dv^{(i)} \right\|_2^2 + \lambda \left\| z^{(i)} o v^{(i)} \right\|_1$$

$$z_j^{(i)} = \begin{cases} \infty, & if\,(1) j \geq \left\lfloor \frac{K}{2} \right\rfloor and\, f_a(i) \geq 0 \\ & (2) j < \left\lfloor \frac{K}{2} \right\rfloor and\ f_a(i) < 0 \\ 1, & otherwise \end{cases} \quad (1)$$

where "o " denotes the pairwise multiplication between two vectors, $f_a(i)$ *is the* attribute score for i^{th} image, and z is a mask vector for deciding which codewords are allowed to be used by image *i*. By using the mask vector z, it forces the sparse representation to be zero if z is α because any other values in these dimensions will cause the objective function to become infinity. The final sparse representation $v^{(i)}$ can be found by solving a L1 regularized least square problem and only considering the dimensions where z=1

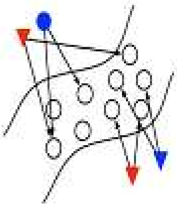

Figure 2: The attribute coding methods, Colors denote the attribute of the images and the shapes indicate the identity.

For multiple attributes, can simply define depending on multiple attributes. For example, if there are two attributes and image contains positive scores for both attributes, will become [1, …,1,∞,…∞,1....,∞]T. Although (3) can successfully encode the human attributes into the sparse representation [5], [4].

2.2. Inverted Indexing (II)

Here describe the second method that can utilize human attributes by adjusting the inverted index structure.

2.2.1 Ranking and Inverted Indexing

For each image, after computing the sparse representation we gets codeword set *c(i)* to represent it by taking non-zero entries in the sparse representation as codewords. The similarity between two images are then computed as follows,

$$S(i,j) = \|c^{(i)} \cap c^{(j)}\|. \quad (2)$$

2.2.2 Inverted Indexing

To embed attribute information into index structure, for each image, in addition to sparse codewords computed from the facial appearance, And use a dimension binary signature to represent its human attribute,

$$b_j^{(i)} = \begin{cases} 1 \; if \; f_a^{(i)}(j) > 0 \\ 0 \; otherwise \end{cases} \quad (3)$$

The similarity score is then modified into,

$$s(i,j) = \begin{cases} \|c^{(i)} \cap c^{(j)}\| \; if h(b^{(i)}, b^{(j)}) \leq T \\ 0 \; otherwise, \end{cases} \quad (4)$$

where *h(i,j)* denotes hamming distance between i and j, and T is a fixed threshold. The image ranking [7]according to (4) can still be efficiently computed using inverted index by simply doing a XOR operation to check the hamming distance before updating the similarity scores.

3. Proposed System

Here I propose to combine identity based face image retrieval with existing system. Proposed system contains CBIR concepts like indexing, clustering, searching (attribute searching and identity based searching) and face image retrieval. Here sparse-codeword is generated by combine attribute and identity of face images. Most of the systems related to existing system are suffering from low recall problem. One solution for this problem is to construct semantic codeword by using extra textual information. So I propose identity based face image retrieval system. First of all, the user chooses identity type like Aadhaar, PAN, Driving License, etc. Then user enters corresponding identity number like Aadhaar Number, PAN Card Number etc. and performs searching. Then the user retrieves images corresponding to that identity. The proposed system contains following modules.

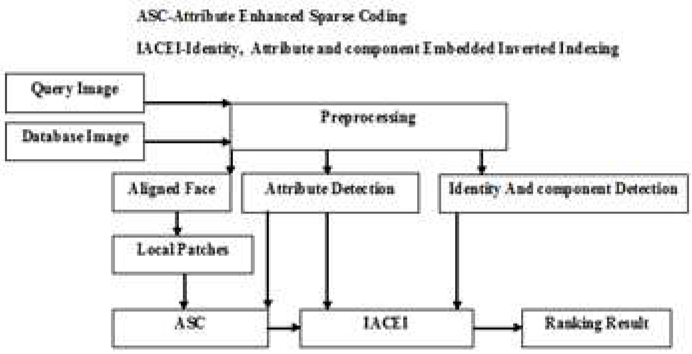

Figure 3: Proposed System Framework

- Content Based Image Retrieval(CBIR)
- Attribute Based Search
- Identity Based Search

The general image retrieval system usually consists of three modules. They are Input module, Query module, Retrieval module. In the query module, a query image is inputed, and its feature vector is extracted for the retrieval (input image is stored in the database). The feature vector of each image in the database is also stored in the database. During the retrieval process, the feature vector of the query image is compared with each vector in the database and the corresponding target (similar) images are outputed. Attribute based search is based on the attribute of the face images. If the user enters some attributes like color, age level, etc. Then retrieves the images corresponds to these attributes and gives the retrieval results. Identity based search is based on the identity information of the face images. If the user enters identity information like PAN card number, Aadhaar numbr etc. Then retrieves the images corresponds to these identity and gives the retrieval results.

3.1 Sparse coding with Identity Information

Traditional BoW model use k-means clustering algorithm to learn dictionary for quantization.

$$\min_{D,V} \sum_{i=1}^{n} \|x_i - Dv_i\|^2 \quad (5)$$

D=[d1, d2,….,dk] is a dictionary matrix with the size of 59×*K*, each column represents a centroid (totally *K*); V=[v1,….vn] is the centroid indicator matrix, each *vi* indicates that the original feature *xi* belongs to a centroid in *D*. The constraint Card (vi)= 1 means each feature can only be assigned to one centroid. This constraint is considered to be too strict because some features might be at boundary of two or more centroids, therefore many people suggest relaxing the constraint and putting L1 regularization term on *vi* instead. This then turn into another optimization problem known as sparse coding:

$$\min_{D,V} \sum_{i=1}^{n} \|x_i - Dv_i\|^2 + \lambda\|v_i\|_1 \; (6)$$

This optimization problem can be efficiently solved by an online optimization algorithm. We adopt this method to learn dictionary *D* for later use. Because we quantize feature descriptors from each grid separately, we have to learn different dictionary from them. Consequently, there are total 175 dictionaries learned. Once the dictionary is learned, we can fix D in the above formulation and minimize the objective function along with v_i separately to find the sparse representation of each feature. When D is fixed, the optimization problem turns into a least square problem with L1 regularization. Since the L1 regularization term makes the objective function non-differentiable when v_i contains zero elements, we cannot solve it by standard unconstraint optimization method. This challenge has led to many literatures using different approaches to solve this problem. Here we use LARS algorithm to solve this problem. Because of the L1 regularization term, each sparse representation v_i will only contain several nonzero elements out of K dimension. These nonzero entries are then considered as the visual words from the descriptor x_i. Since there are 175 grid locations, total dictionary size is 175×K.

4. Future Scope

In the future, we will test the scalability of our system and combine more context information into the proposed framework to further boost up the performance. In this section, we discuss the scalability of our system in two aspects, memory usage and online retrieval time.

4.1 Memory Usage

In our current implementation, each codeword only needs 16 bits to store its image ID in the index, and each image contains about 3,800 codewords on average. Total memory usage for inverted index with 13 K images is about 94.2 MB. We also need to store 40 bits attribute signature for each image (totally 0.1 MB for 13 K images). Therefore, total memory usage is about 94.3 MB for LFW dataset. For dataset with one million images, each codeword needs 20 bits to store its image ID; therefore, total memory usage for inverted index is about 9,060 MB. For attribute signatures, they take about 4.8 MB. Total memory usage is about 9,064.8 MB, which is a reasonable amount for a general computer server. Note that memory usage can be further reduced by many compression techniques in information retrieval [4] (e.g., reducing around half of memory by adopting d-gap technique).

4.2 Online Retrieval Time

Our system is implemented using C++ and operates on a 2.4 GHz Intel Xeon server. For a single query, face detection and alignment take about 0.7 seconds, LBP feature extraction takes about 0.06 seconds, computing sparse representation takes about 0.35 seconds, and retrieving index with 13 K images takes about 0.03 seconds. For attribute detection, as shown in [6], detecting a single attribute can be done within few milliseconds once the classifier is learned. For dataset with one million images, we refer to the results in [5]. In their reports, retrieving index with one million images takes 0.2 seconds. Since we have similar index structure with [3], we expect that retrieving index with one million face photos can be done in less than one second.

5. Conclusion

The existing system present a novel approach to face recognition which considers both shape and texture information and attribute information to represent face images. Existing system aim to utilize automatically detected human attributes that contain semantic cues of the face photos to improve content based face retrieval by constructing semantic codewords for efficient large scale face retrieval. I propose to combine identity based face image retrieval with existing system. Proposed system contains CBIR concepts like indexing, clustering, searching (attribute searching and identity based searching) and face image retrieval. Most of the systems related to existing system are suffering from low recall problem. One solution for this problem is to construct semantic codewords by using extra textual information. So I propose identity based face image retrieval system.

References

[1] Z. Cao, Q. Yin, J. Sun, and X. Tang, Face Recognition with Learning-Based Descriptor and Pose-Adaptive Matching, Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010.
[2] J. Chen, R. Ma, and Z. Su, Weighting Visual Features with Pseudo Relevance Feedback for cbir, Proc. ACM Intl Conf. Image and Video Retrieval, 2010.
[3] S. Gao, I. W.-H. Tsang, L.-T. Chia, P. Zhao. Local features P are not lonely laplacian sparse coding for image classification. In CVPR, 2010
[4] J. Wang, S. Kumar, and S.-F. Chang. Semi-supervised F. Semi hashing for scalable image retrieval. In CVPR, 2010.
[5] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In ICML, 2009.
[6] T. Ahonen, A. Hadid, and M. Pietikainen, Face recognition with local binary patterns. In ECCV, 2004.
[7] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In CVPR, 2001.
[8] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In CVPR, 2009.
[9] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with scale local binary patterns. 2002.
[10] D. Lowe. Distinctive image features from scale-invariant scale keypoints. In IJCV, 2003.

Author Profile

Indulekha K received the B.Tech. Degree in Computer Science and Engineering from College of Engineering, Thalassery in 2010 and pursuing M.Tech. Degree in Computer Science and Engineering from KMCT College of Engineering. During 20012-2014, her research areas include Image Processing.