ISSN:0975-9646

Content-Based Face Image Retrieval Using Attribute-Enhanced Sparse Codewords

Bharti S. Satpute
ME Computer Student
Padmashree Dr. D.Y. Patil Institute of Engineering & Technology, Pimpri, Pune
SavitriBai Phule Pune University, India

Archana A. Chaugule
Asst. Professor in Computer Dept
Padmashree Dr. D.Y. Patil Institute of Engineering & Technology, Pimpri, Pune
SavitriBai Phule Pune University, India

***Abstract*— Due to the popularity of image capturing digital devices and the ease of social network/photo sharing services (e.g., Facebook, Twitter, Flickr), there are largely growing consumer photos available in database. Thus, with the exponentially growing photos, large-scale content-based face image retrieval is an enabling technology for many emerging applications. In this face image retrieval system, aim is to utilize automatically detected human attributes that contain semantic cues of the face photos to improve face retrieval by constructing semantic codewords for efficient large-scale face image retrieval. In this work, low level features and high level attributes are used to represent facial images and regression technique is applied to improve retrieval result. Experimental result shows that, proposed face image retrieval framework achieved more accuracy as compared to the existing methods.**

Keywords—content-based image retrieval; human attribute detection; sparse coding; content-based face image retrieval; sparse codewords.

I. INTRODUCTION

Due to the popularity of digital devices, people can easily capture a photo and share it using the internet by various online tools like facebook, flickr, quicker, twitter, etc. Among these vast digital images and photos shared on the internet, a big percentage of them are photos related to human faces. Because human faces are closely related to social activities of human beings. The exponential growth of facial images has created many research problems and opportunities for a variety of real-world applications.

Facial images usually have high intra-class variances caused by expressions, poses and illuminance (lighting variations). Due to these intra-class variances in face images, content-based face image is very challenging problems. Aim in this work is to address large-scale content-based face image retrieval problem which is very important and challenging task. In content-based face image retrieval system, when input face image is given then it tries to find most similar facial images from a large image database. Face image retrieval is enabling technology for many applications like automatic face annotation, crime investigation, etc.

Traditional face image retrieval methods usually use low-level features to represent face images [1], [2], [3], but these local features are lack of semantic meanings and face images have high intra-class variance, so the retrieval results are not satisfactory. To address this problem, Z. Wu et al. [2] propose framework for face image retrieval which use identity based quantization and B. C. Chen et al. [3] propose to use identity-constrained sparse coding combined with component based local binary pattern, but these methods might require clean training data and massive manual annotations.

In this paper, a novel approach for content-based face image retrieval is provided by incorporating high-level human attributes (for example, gender, hair color, skin color) into face image representation. Face image of two different persons might be very close in the traditional low level feature space, because low-level features are lack of semantic meanings. By combining low-level features with high-level human attributes, we can find better feature representations and achieve more relevant retrieval results. Ramisa et al. [4] proposed similar idea by combining fisher descriptors with attributes for retrieval of particular objects as well as categories. Their experiments shows that retrieving images of particular objects based on attribute vectors gives results comparable to the state of the art and also demonstrate that combining attribute and fisher vectors improves performance for image retrieval. But they use early fusion for combining attribute and fisher descriptors. Also, they do not take advantages of human attributes because their target was general image retrieval and their selection and training process for attributes features is somewhat adhoc.

High-level human attributes are semantic descriptions about a person (e.g. hair color, gender, race, smiling etc). Recent work shows that automatic attribute detection helps in different applications to achieved better result. Using these automatically detected human attributes, many researchers have achieved promising result in different applications such as face identification, face verification [5], keyword-based face image retrieval [6], and similar attribute search [7]. This recent work shows power of high-level semantic human attributes on face images. To improve content-based face image retrieval, we proposed attribute-enhanced sparse coding with identity constraint. This system used automatically detected human attributes of face images using learned attribute classifier. Attribute-enhanced sparse coding method is used to represent images in offline stage using sparse codewords. For constructing sparse codewords for image representation this method used both, low-level features as well as several important

high-level human attributes of images. On the other hand, regression method provides efficient retrieval result in the online stage.

The rest of the paper is organized as follows. Section II discusses related work in areas like content based face image retrieval and human attribute detection. Section III describes proposed methodology including attribute-enhanced sparse coding. Section IV gives brief idea about experimental setting and result from this proposed framework. Section V concludes this paper.

II. RELATED WORK

Traditional content-based image retrieval techniques use (low-level) image content like color, texture and shape to represent images. Inverted indexing and hash indexing, these two types of indexing systems are used to deal with large-scale data. To achieve efficient similarity search, many studies have go with inverted indexing or hash-based indexing combined with bag-of-word model (BoW) and local features like scale-invariant feature transform (SIFT). The bag-of-words(BoW) model is a well-known feature representation method for image categorization and annotation tasks. These methods can achieve high precision on rigid object retrieval, but they suffer from low recall rate due to the semantic gap. Recently, several researchers have focused on bridging this semantic gap by finding semantic image representations to improve the CBIR performance. L. Wu et al. [8] propose a novel Semantic-Preserving Bag of Word model to learn optimized BoW models, by considering the distance between semantically identical features as a measurement of the semantic gap, and they attempt to learn an optimized codebook by minimizing this gap, aiming to achieve the minimal loss of the semantics using effective distance metric learning. Y. H. Kuo et al. [9] propose a framework to enhance each database image with semantically related auxiliary visual words (AVWs). The idea of our proposed work is similar to the abovementioned methods, but instead of using extra information that might require intensive human annotations (and tagging), we try to make use of automatically detected human attributes to construct semantic codewords for the face image retrieval task.

Recently, many work shows automatically detected human attributes have achieved promising results in different applications. N. Kumar et al. [5] propose a learning framework to automatically detect describable aspects of visual attributes. In their framework, an extensive vocabulary of visual attributes is used to label a large data set of images, which is then used to train classifiers which measures the presence, absence, or degree to which an attribute is expressed in images and then these attribute classifiers can automatically detect human attributes of new face images and label that new images. Using these automatically detected human attributes with the help of attribute classifiers, they achieve outstanding performance on face verification and keyword-based image search. Siddiquie et al. [6] further extend the framework for ranking and retrieval of face images using multi-attribute queries. They propose a framework for multi-attribute keyword-based face image retrieval which explicitly models the correlations that are present between the different attributes which leads to improved ranking/retrieval performance. This recent works shows the emerging opportunities for the human attributes but are not used to generate more semantic (scalable) codewords for image representation. Although these works achieve salient performance on keyword-based face image retrieval and face recognition, we further extend framework to exploit effective ways to combine low-level features and automatically detected facial attributes for scalable content based face image retrieval.

Now a day the rise of photo sharing/social network services (e.g. flickr, picasa, quickr, snapdeal etc.) leads to rises the strong needs for large-scale content-based face image retrieval. Content-based face image retrieval task is intimately related to face recognition task. The difference between face image retrieval and face recognition is that face recognition requires completely labeled data in the training set, and it uses learning process to find classification result while in face image retrieval task neither training set nor learning based approach is required and it gives a ranking result. Generally, face retrieval task focus on finding suitable feature representations for scalable indexing systems because of its high dimensionality. Facial images are more diverse and pose more visual variances (e.g. in poses, expressions, lighting conditions etc).

Existing methods for face image retrieval usually use low-level features to represent face images which have lack of semantic meanings and face images generally have high intra-class variance (e.g., in expressions, posing, illuminance etc.), so the face image retrieval results are not satisfactory. To tackle this problem, Z .Wu et al. [2] propose a framework to use identity based quantization scheme and multi-reference re-ranking for scalable face image retrieval. Using bag-of-words representation and textual retrieval methods content-based image retrieval systems achieve scalability, but performance of such a system degrades quickly in the face image domain, mainly because they produce visual words with low discriminative power for face images, and also they ignore the special properties of the faces. This paper [2] develops a new scalable face representation using both local and global features. They exploit special properties of face images to design new component-based local features and then use identity-based quantization scheme to quantize local features into discriminative visual words. In addition to local features, they also use a small hamming signature (40 bytes) to encode the discriminative global feature for each face and after that re-rank the top retrieved candidate face images using hamming signature. This improves the precision and also not losing the scalability. But in this identity-based quantization scheme for face image retrieval, construction of visual word vocabulary requires manual annotation (and tagging). Automation of this process can further improve the visual word vocabulary for face.

Wang et al. [1] investigated the retrieval-based face annotation problem and propose a promising framework to deal with this challenge by mining massive weakly labeled facial images freely available on Internet. A novel Weak

Label Regularized Local Coordinate Coding (WLRLCC) algorithm was proposed to improve the annotation performance. This WLRLCC algorithm effectively exploits the principles of both local coordinate coding and graph-based weak label regularization.

B. C. Chen et al. [3] developed a scalable face image retrieval framework using component-based local binary pattern (LBP) with the help of sparse coding and which can integrate with partial identity information to improve the face retrieval result. In their approach, they first apply sparse coding on local features extracted from facial images combining with inverted indexing to construct an efficient content-based face image retrieval system. Then propose a novel sparse coding scheme that refines the representation of the original sparse coding by using partial identity information. Their experimental results shows that the system can achieve salient retrieval results on LFW dataset (13K faces) and also achieve better performance than linear search methods based on well-known face recognition feature descriptors. But disadvantage of this proposed coding scheme is that, face images with more intra-class variances will still be quantized into similar visual codewords if they share the same identity and use of identity information might need manual annotations.

Similar approach as Chen et al. used in [3] is adopted —using component-based LBP features combined with sparse coding to construct sparse codewords for efficient content-based face image retrieval. We focus on taking advantage of automatically detected human attributes to construct semantic-aware sparse codewords using attribute-enhanced sparse coding method.

III. PROPOSED SYSTEM

When input query image is given, then the goal of face image retrieval is to find the ranking result from most to least similar face images in a face image database.

Proposed scalable face image retrieval system has following objectives:

- Design sparse coding method which combines automatically detected high level features with low level features for better image representation to achieve improved face retrieval result.
- Use regression technique for efficient and fast retrieval result.
- Use randomly sampled image patches as dictionary instead of learned dictionary. Because learning dictionary with a large vocabulary is time-consuming process (training 175 codebooks with 1600 dimension takes more than two weeks to finish)[14], so aim to just use randomly sampled image patches as dictionary and skip the time-consuming dictionary learning step by fixing dictionary D.
- Design dictionary selection process to force images with different attribute values to contain different codewords.

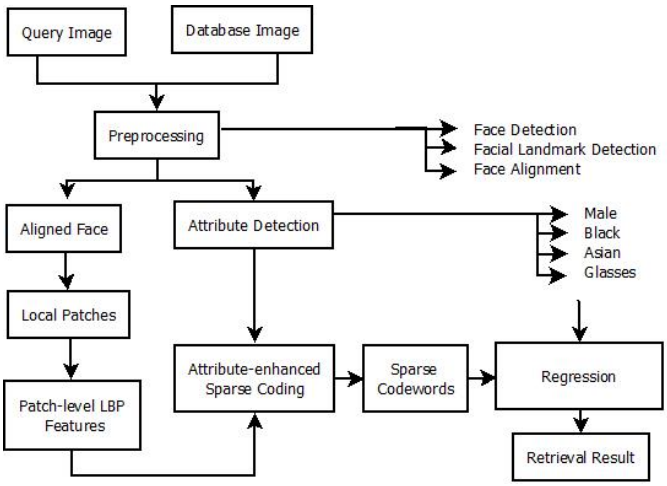

Fig. 1. Block Diagram of Proposed System

A. System Overview

Both database images and query image will go through same procedure as shown in fig 1: First step is image pre-processing. In pre-processing stage, first step is to find location of face from input image and then find components of face. Human attributes are detected from given face. After that, face alignment is done to extract low level features. From detected facial components local patches extracted and 9-dim LBP features computed. These are called local feature descriptors. After obtaining local low-level LBP features and attributes scores, our aim to combine these features to semantically represent image. To the best of our knowledge, this is the first proposal of such combination for content-based face image retrieval. For that purpose, proposed attribute-enhanced sparse coding method is applied to construct sparse codewords for that image.

Query image will go through the same procedure to obtain sparse codewords and human attributes, and use these codewords to retrieve images from dataset. For efficient retrieval from dataset images regression technique is used.

B. Methodology

Step1: Image preprocessing
Step2: Attribute detection
Step3: Patch-level LBP feature computation
Step4: Attribute-enhanced sparse coding
Step5: Regression technique
Step6: Retrieval result

1) Image preprocessing:

For every image in the database, first Viola-Jones face detector [11] is applied to find the locations of faces. To locate 68 different facial landmarks on the face image Active shape model [12] is applied. Using these facial landmarks, barycentric coordinate based mapping process is applied to align every face with the face mean shape [13].

2) Attribute detection

Before face alignment in preprocessing stage attribute detection is needed. Attribute detection framework from [5] used to find human attributes from located face. For

automatic attribute detection, attribute classifier is trained using various labeled images from internet which measures the presence, absence, or degree to which an attribute is expressed in images.

3) Patch-level LBP feature computation

In our framework three components detected from face image, nose tip, and two eyes. From each detected facial component 81 grids are extracted, where each grid is square patch [2]. Hence, in total we have 243 grids from aligned face. From each grid, an image square patch is extracted and a 9-dimensional uniform LBP features computed.

4) Attribute-enhanced sparse coding

This coding method is applied to all patches in a single image to find different codewords and finally combine all these codewords together to represent the image.

In order to enhancing sparse coding with human attributes to represent image, first dictionary selection needed to force images with different attribute values to contain different codewords. For a single human attribute, dictionary centroids divided into two different subsets, images with negative attribute scores will use one of the subset and images with positive attribute scores will use another subset. For example, if an image has a positive female attribute score, it will use the first half of the dictionary centroids. If it has a negative female attribute score, it will use the second half of the dictionary centroids. With help of this, images with different attributes will definitely have different codewords. For considering multiple attributes, sparse representation is divided into multiple segments based on the number of attributes, and each segment of sparse representation is generated depending on single attribute. This goal of finding sparse representation can be achieved by solving the following optimization problem (1) [10]:

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

Where $x^{(i)}$ is the original features extracted from a patch of face image *i*, $D \in R^{d \times K}$ is a to-be-learned dictionary having K centroids with d dimensions. V= (v^1, v^2, …, v^n) is the sparse representation of the image patches. ”o” is the pairwise multiplication between two vectors. $f_a(i)$ denotes attribute score of i^{th} image and $z^{(i)}$ is mask vector for deciding codewords allowed to be used by image *i*. We first assign a half of the dictionary centroids to have +1 attribute score and use them to represent images with the positive attribute; the other half of the dictionary centroids are assigned with -1 to represent images with the negative attribute. After the assignment, we can use the distance between attribute scores of the image and the attribute scores assigned to the dictionary centroids as the weights for selecting codewords. Because the weights are decided by attribute scores, two images with similar attribute scores will have similar weight vector, and therefore have a higher chance to be assigned with similar codewords and result in similar sparse representations. Images with similar attributes will be assigned with similar centroids, but images with erroneous attributes might still be able to retrieve correct images if their original features are similar.

We first define an attribute vector a ∈ {+1, -1}, where a_j contains the attribute scores of the j^{th} centroid. z^i can be calculated using following equation:

$$z^i_j = \exp\left(\frac{d(f_a(i), a_j)}{\sigma} \right) \qquad (2)$$

The final sparse representation $v^{(i)}$ can be found by solving a L1 regularized least square problem and only considering the dimensions where $z_j^{(i)} > 0$. After finding $v^{(i)}$ for each image patch only considered non-zero entries as codewords of image and save that codewords of dataset images in result file.

5) Regression

For each image, after computing the sparse representation using the method described in Section III-B4, we can use codeword set $c^{(i)}$ to represent it by taking non-zero entries in the sparse representation as codewords. The similarities between query image and dataset images are find out using regression technique.

6) Retrieval result:

When query image arrives, then this proposed system finds sparse codewords. Then using multiple regression technique images those are similar to input image are retrieved.

IV. EXPERIMENTS

A. Experimental setup

1) Datasets:

We are using two different public datasets, LFW and Pubfig for experiments. Various images from these two datasets are taken for database and query set those are having large intra-class variances.

2) Parameter Setting:

We need to decide parameter lambda and dictionary size K for methods SC and AESC. We set lambda= [10^{-6}, 10^{-2}] and dictionary size K= [400, 3200].

B. Result

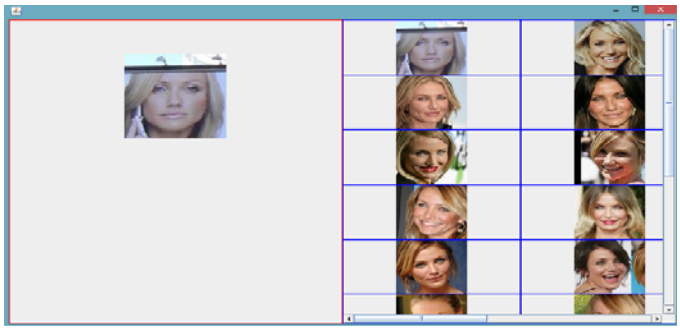

Fig. 2. Generated result of Proposed System

To evaluate performance of proposed system we checked accuracy of proposed system with existing system. We tested our result on various query images against dataset images. Fig. 2 shows result generated from proposed system. Experimental result shows, proposed system having more accuracy as compared to existing system as shown in fig 3.

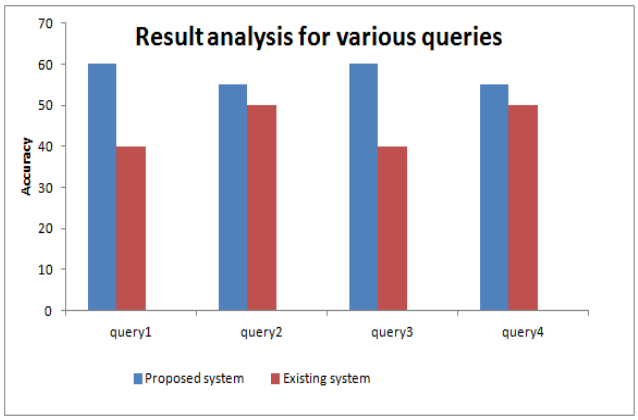

Fig. 3. Result analysis

V. CONCLUSION

Scalable face image retrieval framework proposed to utilize automatically detected human attributes to significantly improve content-based face image retrieval. Proposed system combining low-level features and automatically detected human attributes for content-based face image retrieval. Attribute-enhanced sparse coding exploits the global structure and uses several human attributes to construct semantic-aware codewords in the offline stage and also used regression to further improve result. Experimental result shows improved result than existing methods.

ACKNOWLEDGMENT

Authors would like to thanks to those peoples who are directly and indirectly involved for their productive suggestions which are helped us to improve the quality and presentation of this work.

REFERENCES

[1] D. Wang, S. C. Hoi, Y. He, and J. Zhu, "Retrieval-based face annotation by weak label regularized local coordinate coding," in Proc. ACM Multimedia, 2011.

[2] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, "Scalable face image retrieval with identity-based quantization and multi-reference re-ranking," in Proc. IEEE Conf. Computer Vision and Pattern Recognit., 2010.

[3] B.-C. Chen,Y.-H. Kuo, Y.-Y. Chen, K.-Y. Chu, and W. Hsu, "Semi-supervised face image retrieval using sparse coding with identity constraint," in Proc. ACM Multimedia, 2011.

[4] M. Douze, A. Ramisa, and C. Schmid, "Combining attributes and fisher vectors for efficient image retrieval," in Proc. IEEE Conf. Computer Vision and Pattern Recognit., 2011.

[5] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Describable visual attributes for face verification and image search," IEEE Trans. Pattern Anal. Mach. Intell., Special Issue on Real-World Face Recognition,vol. 33, no. 10, pp. 1962–1977, Oct. 2011.

[6] B. Siddiquie, R. S. Feris, and L. S. Davis, "Image ranking and retrieval based onmulti-attribute queries," in Proc. IEEE Conf. Computer Vision and Pattern Recognit., 2011.

[7] W. Scheirer, N. Kumar, P. Belhumeur, and T. Boult, "Multi-attribute spaces: Calibration for attribute fusion and similarity search," in Proc. IEEE Conf. Computer Vision and Pattern Recognit., 2012.

[8] L. Wu, S. C. H. Hoi, and N. Yu, "Semantics-preserving bag-of-words models and applications," IEEE Trans. Image Process., vol. 19, no. 7, pp. 1908–1920, Jul. 2010.

[9] Y.-H. Kuo, H.-T. Lin, W.-H. Cheng, Y.-H. Yang, and W. H. Hsu, "Unsupervised auxiliary visual words discovery for large-scale image object retrieval," in Proc. IEEE Conf. Computer Vision and Pattern Recognit., 2011.

[10] B. C. Chen, Y. –Y. Chen, Y. –H Kuo, and Winston H. Hsu, "Scalable Face Image Retrieval Using Attribute-Enhanced Sparse Codewords", IEEE Transactions On Multimedia, vol. 15, no. 5 Aug 2013.

[11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Computer Vision and Pattern Recognit.*, 2001.

[12] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proc. Eur. Conf. Computer Vision*, 2008.

[13] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 406-415, Sep 2010.

[14] A. Coates and A. Y. Ng, "The importance of encoding versus training with sparse coding and vector quantization," in *Proc. ICML*, 2011.