Face Image Retrieval using CBIR with Attribute Enhanced Sparse Codewords

Mahalakshmi S ,Pramod M
Kalabhavi,Roopavani B , Tarranum Bano
T.John institute of technology
Bangalore, India

Ms.Poonamsree S, Asst Prof,Dept of CSE
T.John institute of technology
Bangalore, India

***Abstract*— These days snaps with natives are the main concern of people. Amid every one of those photos, a large percentage of them are snaps with person faces. The significance and the pure quantity of person face snaps construct manipulations of big scale person face descriptions a truly imperative research difficulty and permit various actual world applications. Therefore, by the exponentially rising pictures, great scale contented based face picture reclamation is an allowing technology for several promising relevance. Content-based image retrieval (CBIR) is a technique to automatically index images by extracting their (low-level) visual content, such as color, texture, and shape, and the retrieval of images is based upon the indexed image features. By leveraging human attributes in a scalable and systematic framework, we propose two orthogonal methods named attribute-enhanced sparse coding and attribute- embedded inverted indexing to improve the face retrieval in the offline and online stages. In this work, we aim to utilize automatically detected human attributes that contain semantic cues of the face photos to improve content-based face retrieval by constructing sparse code-words for efficient large-scale face retrieval, the results show that the proposed methods can achieve up to 43.5% relative improvement in MAP compared to the existing methods.**

Keywords- Face image, facial attributes content-based image retrieval, attributes enhanced sparse coding, and attribute embedded inverted indexing, attribute detection;

I. INTRODUCTION

Nowadays we have many multimedia devices such a camera, cellular phone, audio/video player and so on. Images play vital role in our daily communication. Impression is more proved by an image rather than a thousand words as stipulated by the statement “One picture is worth more than a thousand words”. Social networks such as Facebook twitter etc. are widely used in our day to day life. Many of them use human face images for their profile. And also people use celebrity’s faces. Now a day’s human faces are mostly used for manipulations such as searching and mining. Face image retrieval using content based method is an emerging technology in many real world applications. Due to different people having similar faces, problems can be faced while we arrive to retrieve for similar faces. To solve this technology such as Retrieval based face annotation use common outline for same categories of image. For example kid cap can be set as constrain to retrieve children’s, long hair for women’s.

Our goal in this paper is to address one of the important and the challenging problems – large-scale content-based face image retrieval. Given a query face image, content-based face image retrieval tries to find similar face images from a large image database. It is an enabling technology for many applications including automatic face annotation, crime investigation, etc. In this work, we provide a new perspective on content- based face image retrieval by incorporating high-level human attributes into face image representation and index structure. As shown in Figure 1, face images of different people might be very close in the low-level feature space. By combining low-level features with high-level human attributes, we are able to find better feature representations and achieve better retrieval results. The similar idea is proposed in using fisher vectors with attributes for large-scale image retrieval, but they use early fusion to combine the attribute scores. Also, they do not take advantages of human attributes because their target is general image retrieval.

Human attributes (e.g., gender, race, hair style) are high- level semantic descriptions about a person. Some examples of human attributes can be found in Figure 2 (a). The recent work shows automatic attribute detection has adequate quality (more than 80% accuracy) on many different human attributes. Using these human attributes, many researchers have achieved promising results in different applications such as face verification , face identification , keyword-based face image retrieval , and similar attribute search.

These results indicate the power of the human attributes on face images. In Table I, we also show that human attributes can be helpful for identifying a person by the information- theoretic measures.

Although human attributes have been shown useful on applications related to face images, it is non-trivial to apply it in content-based face image retrieval task due to several reasons. First, human attributes only contain limited dimensions. When there are too many people in the dataset, it loses discriminability because certain people might have similar attributes. Second, human attributes are represented as a vector of floating points. It does not work well with developing large- scale indexing methods, and therefore it suffers from slow response and scalability issue when the data size is huge.

To leverage promising human attributes automatically detected by attribute detectors for improving content-based face image retrieval, we propose two orthogonal methods named attribute-enhanced sparse coding and attribute-embedded inverted indexing. Attribute-enhanced sparse coding exploits the global structure of feature space and

uses several important human attributes combined with low-level features to construct semantic codewords in the offline stage. On the other hand, attribute-embedded inverted indexing locally considers human attributes of the designated query image in a binary signature and provides efficient retrieval in the online stage. By incorporating these two methods, we build a large-scale content-based face image retrieval system by taking advantages of both low- level (appearance) features and high-level (facial) semantics.

II. CONTENT BASED IMAGE RETRIEVAL

Content-based image retrieval (CBIR) is a technique to automatically index images by extracting their (low-level) visual content, such as color, texture, and shape, and the retrieval of images is based solely upon the indexed image features. Mainly two kinds of indexing systems are used, to deal with large scale data. Many studies have move with Inverted indexing or hash-based indexing combined with bag-of-word model (Bow) and local features like scale-invariant feature transform (SIFT), to achieve efficient similarity search. The bag-of-words model is a well-known and popular feature representation method for image categorization and annotation tasks. The key idea is to quantize high-dimensional local features into one of visual words, and then represent each image by a histogram of the visual words. For this purpose, a clustering algorithm (e.g., K-means), is generally used to generate a codebook (or vocabulary) by converting the visual features to codewords or visual words. However, traditional Bow-like methods not succeed to address issues related to noisily quantize visual features and also problems related to variations in viewpoints, lighting conditions, etc., commonly observed in large-scale image datasets. These methods can achieve high precision on rigid object retrieval, but they suffer from low recall rate due to the semantic gap. In recent times, some researchers have focused on bridging the semantic gap by finding semantic image representations to improve the CBIR performance.

A. *Human Attribute Detection*

Human attributes are high-level semantic descriptions about a person (e.g., gender, age, hair style, skin color). The recent work shows automatically detected human attributes have achieved promising results in different applications. The advantages of a describable human attributes are manifold: they are composable; they are generalizable, as from large image collections one can learn a set of attributes and then apply them to recognize new objects or categories without any further training; and attributes are also efficient. N. Kumar et al. [3] propose a learning framework to automatically detect describable aspects of visual appearance. In their approach, 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 label new images. First large number of images are collected from Internet using various online tools which having vast variations. Then, commercial face detector used to extract faces and fiducial points from downloaded images and stored in the Columbia Face Database. Facial images from Columbia Face Database are submitted to the Amazon Mechanical Turk (MTurk) service, for labelling images with attributes and identity. From these attribute and identity labels and face database, two publicly available face data Sets created, namely Face Tracer and Pubfig data sets, respectively which have been publicly released for non-commercial use. A set of labeled positive and negative images for each attribute are requires for training attribute classifiers. For that purpose, all types of low-level features from the whole face are extracted and automatic, iterative selection procedure designed to select best features from a rich set of low-level feature options. The selected features are used to train the attribute or simile classifier. Using automatically detected human attributes with the help of attribute classifiers, they achieve excellent performance on face verification and keyword-based image search. B. Siddiquie et al. further extend the framework for ranking and retrieval of images based on multi-attribute queries. They propose a principled approach for multi-attribute keyword-based face image retrieval which explicitly models the correlations that are present between the different attributes which leading to improved ranking/retrieval performance. This recent works demonstrate the emerging opportunities for the human attributes but are not used to generate more semantic (scalable) codewords. Although these works achieve salient performance on keyword-based face image retrieval and face recognition, Chen et al. further extend framework to exploit effective ways to combine low-level features and automatically detected facial attributes for scalable content based face image retrieval. To further improve quality of attributes, techniques based on the statistical Extreme Value Theory used by W. Scheirer et al. to propose multi-attribute space to normalize the confidence scores from different attribute detectors for similar attribute search.

III. ATTRIBUTE ENHANCED SPARSE CODEWORDS

Existing face image retrieval system use low level facial features to represent face image. But these low level features are lack of semantics meaning and affects retrieval performance because of facial images have high inter class variations. Face images of different people might be very close in the low-level feature space. In this paper, proposed system utilized high level human facial attributes into face image representation and index structure. Human facial attributes (e.g., hair, age, gender, personal, race) are provide high-level semantic descriptions about a human face. By combining low-level features with high-level human attributes can provides better feature representations. Because certain people might have similar attributes it loses discriminability among too many face images in database.

To address these problems two orthogonal methods are proposed: 1. Attribute-enhanced sparse coding 2. Attribute-embedded inverted indexing. Attribute-enhanced sparse coding uses facial attributes combined with texture features to construct semantic codewords.

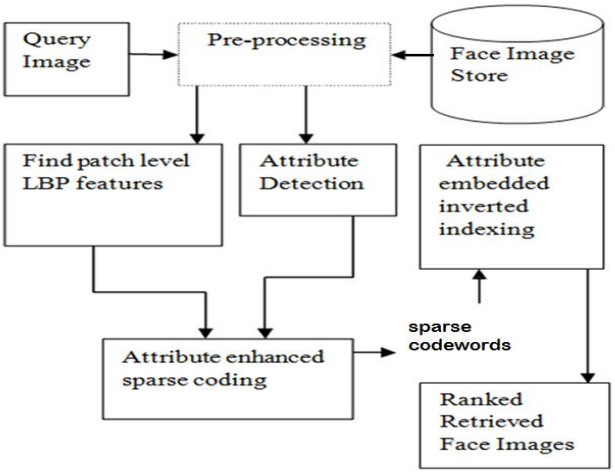

Fig.1 System overview

1) Sparse coding for face image retrieval (SC): Using sparse coding for face image retrieval, we solve the following optimization problem:

$$\min_{D,V} \sum_{i=1}^{n} ||x^{(i)} - Dv^{(i)}||_2^2 + \lambda ||v^{(i)}||_1$$
$$\text{subject to } ||D_{*j}||_2^2 = 1, \forall j \quad (1)$$

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 contains K centroids with d dimensions. $V = [v^{(1)}, v^{(2)}, v^{(n)}]$ is the sparse representation of the image patches. The constraint on each column of D (D_{*j}) is to keep D from becoming arbitrarily large. Using sparse coding, a feature is a linear combination of the column vectors of the dictionary. Provides an efficient online algorithm for solving the above problem.

Note that the Equation (1) actually contains two parts: Dictionary learning (find D) and sparse feature encoding (find V). In, Coates ET. Al. found that using randomly sampled image patches as dictionary can achieve similar performance as that by using learned dictionary (< 2.7% relative improvement in their experiments) if the sampled patches provide a set of over complete basis that can represent input data. Because learning dictionary with a large vocabulary is time-consuming (training 175 codebooks with 1600 dimension takes more than two weeks to finish), we can just use randomly sampled image patches as our dictionary and skip the time-consuming dictionary learning step by fixing D in the Equation (1) and directly solve V. When D is fixed, the problem becomes a L1 regularized least square problem, and can be efficiently solved using LARS algorithm. After finding v^{for} each image patch, we consider nonzero entries as codewords of image i and use them for inverted indexing. Note that we apply the above process to 175 different spatial grids separately, so codewords from different grids will never match. Accordingly, we can encode the important spatial information of faces into sparse coding. The choice of K is investigated in Section V-A. We use K = 1600 in the experiments, so the final vocabulary size of the index system will be $175 \times 1600 = 280,000$.

IV. ATTRIBUTE EMBEDDED INVERTED INDEXING

It collects the sparse code words from the attribute enhanced sparse coding and check the code words with the online feature database and retrieve the related images similar the query image.

For every image in the database face detector is used to detect the location of face region. 73 possible attributes can be taken. For example hair, color, race, gender etc. Active shape model is used to mark the facial landmarks and by using that land mark alignment of the face is done. For each face component 7*5 grid points are taken. Each grid will be a square patch. These grid components include eyes, nose, mouth corners etc. LBP feature descriptor is used to extract features from those grids. After extracting the features we quantize it to code words known as sparse coding. All these code words are summed and generate a single pattern for the image. These steps are obtained by using attribute enhanced sparse coding. Before storing the image in database an index number will be provided to it and by using that index number we can identify the image. All these process will be performed in offline stage. Attribute embedded inverted indexing will be performed in online stage which compares the sparse codeword of query image and the database image and finally provide all the similar faces from the database. This technology is the emerging one that is used in real time applications.

V. CONCLUSION AND FUTURE ENHANCEMENT

Conventional face matching system generate only numeric matching scores as a similarity between face images but in the proposed method combines two methods to utilize automatically detected human attributes to significantly improve content-based face image retrieval .Here I have used and adopted the rule of maximization of mutual information to obtain a compact and discriminative dictionary. Some dictionary atoms are also considered as attributes in the paper. Since I have used the sparse coding feature which speeds up the process. To the best of my knowledge, this is the first proposal of 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. Attribute-embedded inverted indexing further considers the local attribute signature of the query image and still ensures efficient retrieval in the online stage. The experimental results show that using the codewords generated by the proposed coding scheme, we can reduce the quantization error and achieve effective results. Current methods treat all attributes as equal. I will investigate methods to dynamically decide the importance of the attributes and further exploit the contextual

relationships between them. My ongoing work includes 1) studying the image resolution requirement for facial mark detection.2) since I used automatic detection of spontaneous asymmetric expressions my further work includes Analyzing few basic words in kids before they talk with the help of expressions. 3) improving the efficiency of the retrieval of images even

REFERENCES

[1] Y.-H. Lei, Y.-Y. Chen, L. Iida, B.-C. Chen, H.-H. Su, and W. H. Hsu, "Photo search by face positions and facial attributes on touch devices," ACM Multimedia, 2011.

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

[3] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," IEEE Transactions on Information Forensics and Security, 2010.

[4] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, "Scalable face image retrieval with identity-based quantization and multi-reference re-ranking," IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[5] 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," ACM Multimedia, 2011.

[6] M. Douze and A. Ramisa and C. Schmid, "Combining Attributes and Fisher Vectors for Efficient Image Retrieval," IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[7] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Describable visual attributes for face verification and image search," in IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Special Issue on Real-World Face Recognition, Oct 2011.

[8] W. Scheirer, N. Kumar, K. Ricanek, T. E. Boult, and P. N. Belhumeur, "Fusing with context: a bayesian approach to combining descriptive attributes," International Joint Conference on Biometrics, 2011.

[9] B. Siddiquie, R. S. Feris, and L. S. Davis, "Image ranking and retrieval based on multi-attribute queries," IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[10] W. Scheirer and N. Kumar and P. Belhumeur and T. Boult, "Multi-Attribute Spaces: Calibration for Attribute Fusion and Similarity Search," IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[11] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[12] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," International Conference on Computer Vision, 2009.

[13] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," European Conference on Computer Vision, 2004.

[14] J. Zobel and A. Moffat, "Inverted files for text search engines," ACM Computing Surveys, 2006.

[15] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," VLDB, 1999.

[16] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," International Conference on Computer Vision, 2003.

[17] D. Lowe, "Distinctive image features from scale-invariant key points,"International Journal of Computer Vision, 2003.

[18] O. Chum, J. Philbin, J. Sivic, M. Isard and A. Zisserman, "Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval," IEEE International Conference on Computer Vision, 2007.

[19] L. Wu, S. C. H. Hoi, and N. Yu, "Semantics-preserving bag-of-words models and applications," Journal of IEEE Transactions on image processing, 2010.