ISSN 2348 – 9928
IJAICT Volume -1, Issue-8, December 2014 Doi:01.0401/ijaict.2014.01.32 Published Online 05 (05) 2014

ATTRIBUTE-ENHANCED SPARSE CODEWORDS AND INVERTED INDEXING FOR SCALABLE FACE IMAGE RETRIEVAL

Ms. D. Suganya. *Ms. S. Saranya* *Dr. N. Kannan*

Jayaram College of Engineering and Technology

Pagalavadi, Trichy, Tamilnadu, India.

Abstract— Social networks have become popular due to its photo sharing facilities. People are interested to explore contents that contain images. Since internet has become a part of life people are interested in uploading images in it. Hence with the exponentially growing photos, large-scale content-based face image retrieval is a facilitating technology for many emerging applications. In this paper, our aim is to utilize automatically detected human attributes which contain semantic cues of the face photos to improve content based face retrieval by constructing semantic cues for efficient image retrieval. This technique is coupled with relevance ranking technique to enhance the efficiency further. Two orthogonal methods named attribute-enhanced sparse coding and attribute embedded inverted indexing are proposed to improve the image retrieval in both offline and online stages. Relevance ranking when added with these methods show performance improvement to greater extent.

Keywords— Face Image, Human Attributes, Content-Based Image Retrieval, Relevance Ranking.

I. INTRODUCTION

There are enormous numbers of images in internet. Among them more than 70% of images contain face images. This is because of the photo sharing facilities provided by the social networks. Our aim is to retrieve [3] face images by giving a query image to the large image database. This process of image retrieval [13] is improved by applying ranking techniques to retrieve most appropriate images as results. In crime investigation [3], Medical diagnosis, Military and automatic face annotation [2] these techniques are used extensively.

Traditional methods for face image retrieval [13] generally use low-level features for representing faces [2], [4], [5]. Low-level features does not contain semantic meanings and will also have high intra-class variance (e.g., expression, posing), hence the retrieval results are unsatisfactory.

To handle this problem, Wu et al. [4] proposed a method to use identity based quantization and Chen et al. [5] proposed a method to use identity constrained sparse coding. These methods would require cleaning training data and massive human annotations, which is practically difficult.

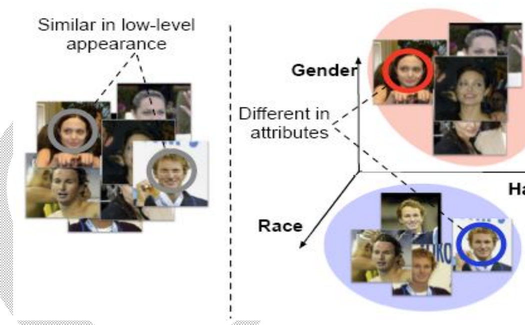

Fig. 1. (a) Because low-level features are lack of semantic meanings, face images of two different people might be close in the traditional low-level feature space.
(b) By incorporating high-level human attributes (e.g., gender) into feature representations, we can provide better discriminability for face image retrieval. (Best seen in color)

Low level features of an image describe elementary characteristics such as the shape, the color, the texture or the motion. High level features represent the human attributes like gender, race, hairstyle, etc. These attributes contain semantic descriptions about a person. While combining both low level and high level features to retrieve an image the accuracy of the retrieval process will be greater. Similar kind of method is proposed in [6] using fisher vectors with attributes for large-scale image retrieval [13], but they use early fusion to combine the attribute scores. But they do not take advantages of human attributes because their target is general image retrieval. With the use of human attributes, many researchers have achieved promising results in different applications such as face verification [7], face identification [8], keyword-based face image retrieval [9], and similar attribute search [10].

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval [13] utilize some method of adding metadata such as captioning, keywords, or descriptions to the images so that retrieval can be performed over the annotation words. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools.

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases. Content-based image retrieval is opposed to concept-based approaches.

"Content-based" means that the search analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because most web-based image search engines rely purely on metadata and this produces a lot of garbage in the results. Also having humans manually enter keywords for images in a large database can be inefficient, expensive and may not capture every keyword that describes the image. Thus a system that can filter images based on their content would provide better indexing and return more accurate results.

To improve content based face image retrieval two methods namely attribute enhanced sparse coding and attribute embedded inverted indexing are proposed [1][11] for offline and online stages respectively. Relevance ranking is the process of sorting the document results so that those documents which are most likely to be relevant to your query are shown at the top.

II. RELATED WORK

Our work is closely related to CBIR human attribute detection, and content-based face image retrieval. CBIR techniques use image content like color, texture and gradient to represent images. To deal with large scale data, mainly two kinds of indexing systems are used. Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases. Content-based image retrieval is opposed to concept-based approaches. In this paper, we propose a method that combines techniques from face detection and recognition with techniques from content-based image retrieval to allow the retrieval of images showing certain persons.

Given a set of positive example images Q+ and a (possibly empty) set of negative example images Q− a score S(Q+,Q−,X) is calculated for each image X from the database B:

$$S(Q+,Q-,X) = \sum_{q \in Q+} S(q,X) + \sum_{q \in Q-} (1 - S(q,X)). \quad (1)$$

where S(q,X) is the score of database image X with respect to query q and is calculated as $S(q,X) = e^{-\mu D(q,X)}$ with $\mu = 1.0$. D(q,X) is a weighted sum of distances calculated as

$$D(q,X) := \sum_{m=1} w_m \cdot d_m(q_m,X_m). \quad (2)$$

Here, q_m and X_m are the mth feature of the query image q and the database image X, respectively.

d_m is the corresponding distance measure and w_m is a weighting coefficient. For each

d_m, $\sum_{x \in B} d_m(Q_m,X_m) = 1$ is enforced by re-normalization.

Given a query (Q+, Q−), the images are ranked according to descending score and the K images X with highest scores S (Q+,Q−,X) are returned by the retrieval system.

Due to the lack of suitable training data, weights w_m were chosen heuristically based on experiences from earlier experiments with other data.

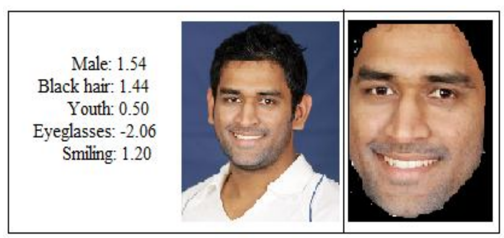

Fig. 2. (a) A face image contains rich context information (hair color, skin color, race, gender, etc.). Using automatic human attribute detection, we can describe them in high-level semantics; for example, Male: 1.54 suggests the person is likely a male and Black Hair: 1.44 implies the hair color tends to be black.
(b) The same image after preprocessing steps in prior works for face image retrieval or recognition. They normalize the position and illumination differences between the faces and exclude background contents. Such common approaches sacrifice the important context information. Using automatically detected human attributes can compensate the information loss.

Viola and Jones [15] present a new and radically faster approach to face detection based on the AdaBoost algorithm from machine learning. Boosting is a method of combining several weak classifiers to generate a strong classifier. AdaBoost is a well known algorithm to generate strong

ISSN 2348 – 9928
IJAICT Volume -1, Issue-8, December 2014 Doi:01.0401/ijaict.2014.01.32 Published Online 05 (05) 2014

classifiers from weak classifiers, while providing statistical bounds on the training and generalization error of the algorithm. The weak classifiers in the Viola & Jones algorithm [15] are based on features of three kinds. A two-rectangle feature is the difference between the sums of the values of two adjacent rectangular windows. A three-rectangle feature considers three adjacent rectangles and computes the difference between the sum of the pixels in the extreme rectangles and the sum of the pixels in the center rectangle. A four-rectangle feature considers a 2 × 2 set of rectangles and computes the difference between the sum of the pixels in the rectangles that constitute the main and off diagonals. For a 24 × 14 sub-window there could be more than 180,000 such features.

III. PROPOSED METHOD

In this paper, our aim is to utilize automatically detected human attributes which contain semantic cues of the face photos to improve content based face retrieval by constructing semantic cues for efficient image retrieval. This technique is coupled with relevance ranking technique to enhance the efficiency further. Two orthogonal methods named attribute-enhanced sparse coding and attribute embedded inverted indexing [11] are proposed to improve the image retrieval in both offline and online stages. Relevance ranking when added with these methods show performance improvement to greater extent.

3.1. Attribute enhanced sparse coding (ASC)

It describes the automatic detection of human attribute from the image and also creates the different sparse coding. These collections of sparse coding represent the original image. 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 [5], 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.

Content based image retrieval, a technique which uses visual contents of image to search images from large scale image databases according to users' interests. This paper provides a comprehensive survey on recent technology used in the area of content based face image retrieval. Nowadays digital devices and photo sharing sites are getting more popularity, large human face photos are available in database. Multiple types of facial features are used to represent discriminability on large scale human facial image database. Searching and mining of facial images are challenging problems and important research issues. Sparse representation on features provides significant improvement in indexing related images to query image.

Sparse coding with identity constraint achieves 70% relative improvement in Mean Average Precision over the baselines. In future we can boost up performance by combining high level semantics features into sparse code generation framework.

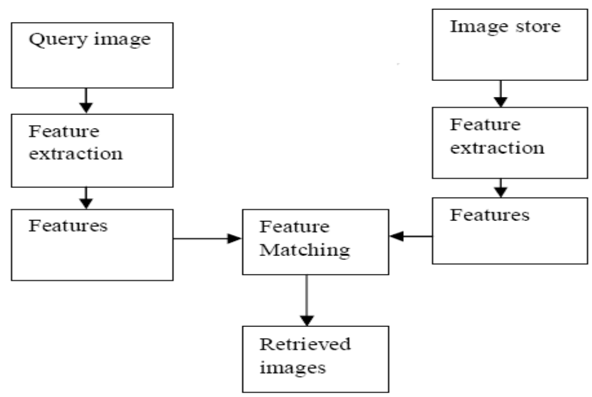

Fig. 3. Content based image retrieval system

3.2. 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. Inverted indexes [11] in image retrieval not only allow fast access to database images but also summarize all knowledge about the database, so that their discriminative capacity largely determines the retrieval performance. Editing the inverted index of a single local feature with multi-class classification scores effectively enhances its discriminative ability. This is because the co-indexing jointly considers strong cues to low-level image contents and their semantic meanings, respectively. The online query remains as efficient as before since only local features are extracted.

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. For each image, after computing the sparse representation using the method, we can use codeword set c(i) to represent it by taking non-zero entries in the sparse representation as

Corresponding Author: Ms. D.Suganya, Jayaram College of Engineering and Technology, Trichy, Tamilnadu, India.

ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.01.32 Published Online 05 (05) 2014

codewords. The similarity between two images is then computed as follows,

$$S(i; j) = c(i) \cap c(j) \qquad (3)$$

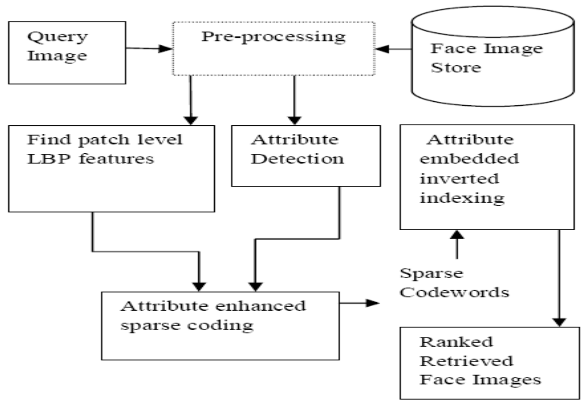

Fig. 4. Face image retrieval with sparse coding

The image ranking according to this similarity score can be efficiently found using inverted index structure [12].

Preprocessing consists of the following steps Face Detection, Compute Attribute Scores and Face Land Mark Detection.

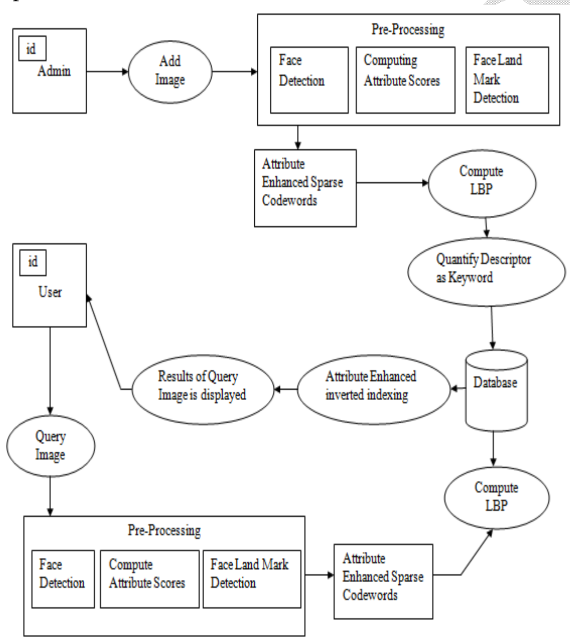

Fig. 5. The proposed system framework.

Both query and database images will go through the same procedures including face detection, facial landmark detection, face alignment, attribute detection, and LBP feature extraction. Attribute-enhanced sparse coding is used to find sparse codewords of database images globally in the offline stage. Codewords of the query image are combined locally with binary attribute signature to traverse the attribute-embedded inverted index in the online stage and derive real-time ranking results over database images.

3.3. Relevance Ranking

The user is no longer able to review several pages of results in order to manually pick the relevant objects. The bias of the search engines toward recall only exacerbates this problem. The final result is that even if a very relevant object is present in the result list, the user still could not find it, again, reducing the perceived usefulness of Learning Object Repositories (LORs)[23].

While doing a stricter filtering of results (increasing precision at the expense of recall) could solve the oversupply problem [24], it could also lead again to the initial problem of scarcity. A proven solution for this problem is ranking or ordering the result list based on its relevance. In this way, it does not matter how long the list is, because the most relevant results will be at the top and the user could manually review them. As almost all search engines use this method, searchers are not only used to work with these sorted lists of results but expect them. To help the user find relevant learning objects, Duval proposed the creation of Learn Rank, a ranking function used to define the relevance of learning objects similarly to how Page Rank defines the relevance of web pages. Also, in a previous paper [24], the authors explore how Contextualized Attention Metadata (CAM) could be mined to obtain meaningful information about the relevance of a specific learning object for a specific user and context.

This paper provides important progress in this direction, proposing and testing a set of multidimensional relevance ranking metrics [24]. These metrics use external sources of information in addition to what is explicitly stated in the user query to provide a more meaningful relevance ranking than current query-matching implementations.

IV. CONCLUSION AND FUTURE WORK

The method combine two orthogonal methods to utilize automatically detected human attributes to significantly improve content-based face image retrieval. To the best of our 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 code words 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 code words generated by the proposed coding scheme, the method can

ISSN 2348 – 9928
IJAICT Volume -1, Issue-8, December 2014 Doi:01.0401/ijaict.2014.01.32 Published Online 05 (05) 2014

reduce the quantization error and achieve salient gains in face retrieval on two public datasets; the proposed indexing scheme can be easily integrated into inverted index, thus maintaining a scalable framework. During the experiments, the method also discovers certain informative attributes for face retrieval across different datasets and these attributes are also promising for other applications. Current methods treat all attributes as equal. The method will investigate methods to dynamically decide the importance of the attributes and further exploit the contextual relationships between them.

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, “Semisupervised 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] J. Zobel and A. Moffat, “Inverted files for text search engines,” ACM Computing Surveys, 2006.

[12] Bor-Chun Chen, Yan-Ying Chen, Yin-Hsi Kuo, Winston H. Hsu, “Scalable Face Image Retrieval using Attribute-Enhanced Sparse Codewords,” IEEE transactions on multimedia vol:pp no:99 year 2013.

[13] 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,” IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[14] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, “Context based vision system for place and object recognition,” International Conference on Computer Vision, 2003.

[15] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” IEEE Conference on Computer Vision and Pattern Recognition, 2001.

[16] S. Milborrow and F. Nicolls, “Locating facial features with an extended active shape model,” European Conference on Computer Vision, 2008.

[17] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online dictionary learning for sparse coding,” ICML, 2009.

[18] A. Coates and A. Y. Ng, “The importance of encoding versus training with sparse coding and vector quantization,” ICML, 2011.

[19] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” Annals of statistics, 2004.

[20] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, “Locality constrained linear coding for image classification,” IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[21] H. Jegou, M. Douze, and C. Schmid, “Hamming embedding and weak geometric consistency for large scale image search,” European Conference on Computer Vision, 2008.

[22] Shumeet Baluja and H. Rowley, “boosting sex identification performance,” International Journal of Computer Vision, 2007.

[23] F. Neven and E. Duval, “Reusable Learning Objects: A Survey of LOM-Based Repositories,” Proc. 10th ACM Int’l Conf. Multimedia (MULTIMEDIA ’02), pp. 291-294, 2002.

[24] L. Sokvitne, “An Evaluation of the Effectiveness of Current Dublin Core Metadata for Retrieval,” Proc. VALA (Libraries, Technology, and the Future) Biennial Conf., 2000.

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

Corresponding Author: Ms. D.Suganya, Jayaram College of Engineering and Technology, Trichy, Tamilnadu, India.