IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Issue-8 E-ISSN: 2347-2693

Face Matching for Similar Faces Evaluation from Videos Using Low Level Facial Geometries

Devendra Sakharkar[1*] and SonaliBodkhe[2]

[1*,2]*Department of Computer Science and Engineering, R.T.M. Nagpur University, India*

Received: Jul /19/2015 Revised: Jul/26/2015 Accepted: Aug/16/2015 Published: Aug/30/ 2015

Abstract –The enhancement of digital devices and the popularity of social networking sites like Facebook, twitter, Instagram etc. The large numbers of peoples are shearing their images and videos by different social networking sites. The users are very much interested in uploading the images or videos on the internet in which most of the photos and videos contain faces. Thus with the rapidly growing photos and videos on the internet the large scale content base face image retrieval is a facilitating technology for many prominent applications. In this project, our aim is to detect a human face image which is present in the video frame and retrieving the similar human face images from the large scale database. By using human attributes in a systematic and scalable framework. The attribute-enhanced sparse coding is used to improve the performance of face retrieval in the offline stage. With this method the performance improvement to greater extent. Experimenting on public photo and video datasets, the result shows that the implementation of above method by using video.

Keywords—Face image, human attributes, content-based image retrieval, Face image retrieval, Face occurrences in videos

I. Introduction

Day to day the increases in the use of social networking sites like Facebook, twitter, Instagram, youtube etc. so most of the peoples are shearing the images and videos by different social networking sites. The users are very much interested in uploading the images or videos on the internet in which most of the photos and videos contain the face images. Thus with the rapidly growing photos and videos on the internet the large scale content base face image retrieval is a facilitating technology for many prominent applications. There are largely growing consumer photos in our life. Among all of these photos and videos, a large number of them are photos with human faces and some videos with the human faces (more than 70%). The large amount of face photos and videos makes manipulation (i.e. search and mining) of large-scale human face images. So it is important research problem and enables many real world applications. It is an enabling technology for many applications including automatic face annotation [1], crime investigation [2], etc.

The aim of the project is to represent the important and challenging problem i.e. large scale content based face image retrieval. When the query will be a video the content based face image retrieval tries to find the similar face images present in the video frames from a large scale database. Some face image retrieval methods use low-level features to represent faces [3],[4],[5],but low-level features having different semantic meanings and face images usually have high intra-class variations (e.g. expression, posing),so the retrieval results are unsatisfactory.

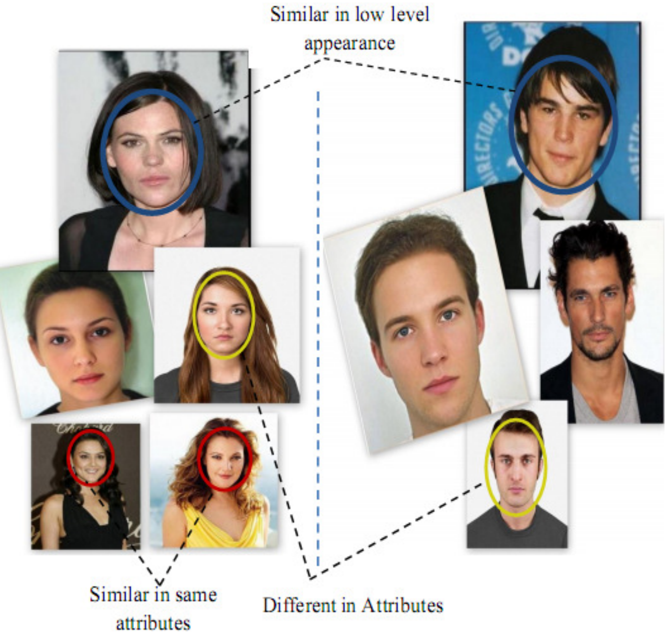


Figure.1 The face images of two different peoples are similar in low-level appearance having different attributes. By combining high-level human attributes (e.g. hair color, gender etc.) into feature representations.

In this paper, a new prospective of content base face image retrieval by combining high-level human attribute into face image representation and index structure will be implemented. The face images of different people are very close in low level features space. To achieve better retrieval result, the low-level features will be combined with high-level human attributes. The similar concept is proposed in [6] using fisher vectors with attributes for large scale image retrieval, but they use early mixture to combine the attribute score.

The objectives of our work are:

- To implement the concept of face image retrieval by using Attribute-Enhanced Sparse Codewords.
- Combining the global structure of feature space and low-level features along with several important human attributes to construct semantic codewords.
- Design a framework for face occurrences in video will be developed by extracting the frames from the videos.
- Retrieving images from the dataset by using multidimensional object features and displaying the output as an image which is present in the video frames.

Human attributes are high-level semantic descriptions about a person like gender, hair style etc. The recent works show that automatic attribute detection has sufficient quality on many different human attributes. Using these human attributes, many researchers have achieved better results in different applications.

II. Related Work

The several different researchers have working on these topics like human attribute detection, and content-based face image retrieval, content-based image retrieval (CBIR).

To deal with large-scale data some CBIR techniques use image content like color, texture and gradient for the representation of the image. To achieve efficient similarity search using hash-based indexing [7,9] or inverted indexing [8] combined with bag-of-word model (BoW) [10] and local features like SIFT [11]. Although these methods can achieve higher accuracy on rigid object retrieval, they suffer from low recall problem because of the semantic gap [12]. Some researchers have work on bridging the semantic gap by finding semantic image representations to increase the CBIR performance. The idea of [13] work is similar to the aforementioned methods, rather than using extra information that might require intensive human annotations, we try to exploit automatically detected human attributes to construct semantic codewords for the face image retrieval task.

A learning framework to find automatically describable visual attributes was proposed in [14]. They use classifiers trained on describable visual attributes and similes for face verification and image search. To determine whether two face images are of the same individual is the problem of face verification because of tremendous variability. An individual's face presents itself to a camera the pose, expression and hairstyle might differ. It makes the matter worse a minimum for researchers in biometry is that the illumination direction, camera type, focus, resolution, and image compression are all almost certain to vary as well. Because of these differences in the images of the same person have difficult for automatic face recognition and verification. Often limiting the reliability of automatic algorithms to the domain with a lot of controlled settings with following subjects [15], [16], [17].

Siddiquie et al. [18] proposed the framework for multi-attribute queries for keyword-based face image retrieval. They address the problem of image ranking and retrieval based on semantic attributes. Problem of image ranking/retrieval of people according to queries describing the physical characteristics of a person, including facial attributes (e.g. hair color, presence of eyeglasses, presence of beard or mustache etc.), body attributes (e.g. color of shirt and pants, long/short sleeves, striped shirt etc.), demographic attributes (e.g. race, gender) and even non-visual attributes (e.g. voice type, temperature) that might probably be obtained from alternative sensors. For example criminal investigation. Based on the description obtained and from eyewitnesses the law enforcement agencies gather the physical traits of the suspect. The entire video taking from surveillance cameras are scanned manually for persons with similar characteristics. This process is time consuming and can be drastically accelerated by an efficient image search mechanism.

A bayesian network approach to utilize the human attributes for face identification [19]. A bayesian formulation that incorporates information beyond soft biometrics, including non-biometric contextual data. They also introduce a Noisy-OR formulation for streamlined truth value assignment and more accurate weighting. Then they examine the accuracy of Bayesian weighting in the presence of unknown attributes. The experiments incorporate the best robust age estimation and describable visual attribute approaches that have been reported in the literature to date. They show that by incorporating additional information into the matching process. They can significantly enhance the accuracy of a leading face recognition algorithm on an identification problem.

For similar attribute search Scheirer et al. propose multi-attribute space to normalize the confidence scores from different attribute detectors [20]. They show the way to calibrate every attribute score to the probability that approximates however humans would label the image with the given attribute. Using a principled technique based on the statistical Extreme Value Theory (EVT) [21, 22], They fit a distribution to attribute scores close to but on the other side of the decision boundary for the attribute in question, e.g., the scores for images classified only slightly negatively for the "female" attribute are used to estimate the probability of being "male." similarly, the statistical fit from these "extreme values" is much more robust than one based on the strongly positive scores of a classifier. In fact, under mild assumptions, this distribution must be a Weibull. This allows for a normalization of raw classifier scores into a multi-attribute space, wherever comparisons and combinations of different attributes become "apples-to-apples." A significant advantage of our method is that it is done after-the-fact, requiring neither changes to the underlying attribute classifier nor ground attribute annotations.

A face retrieval framework using component-based local features to deal with scalability issues was proposed in [23]. They propose unique representation local and global features of images. First, they locate component-based local features that not only encode geometric constraints, but are also more robust to pose and expression variations. Second, they present a novel identity based quantization scheme to quantize local features into discriminative visual words, allowing us to index face images, a critical step to achieve scalability. Our identify-based quantization can better handle intra-class variation using multiple examples. Finally, in addition to the local features, we compute a 40-byte hamming signature for every face image to compactly represent a high-dimensional discriminative global (face recognition) feature.

III. PROBLEM DEFINITION

The works on [5], [4], [18] demonstrate the emerging opportunities for human attributes but are not generate the semantic codewords. These works achieve the better performance on keyword-based face image retrieval and face recognition. We propose to use effective ways to combine low-level features and automatically detected facial attributes for scalable face image retrieval. The prior work on [1], [3], [6] usually crop only the face into constant position and reduce the intra-class variance caused by pose and lighting variations. During this preprocessing step they ignore the rich semantic cues for face such as hair style, skin color, gender etc. As compare to the original image with the cropped version of face image the face verification performance will drop. The experiments suggest that the surrounded image context contain the important information for identifying a person. Therefore, to compensate the information loss we use automatically detected human attributes.

IV. PROPOSED WORK

For every video in the dataset will be extract into the frames and apply the Viola-Jones face detector to find the location of faces present in the frame. Extract more features by applying color map and edge map on the Viola-Jones face detector. Apply the active shape model to locate 68 different facial landmarks on the images. For every facial component (i.e. two eyes, nose tip, and two mouth corners) extract into the 7×5 grids, where every grid is a square patch. By combining there are 175 grids in total. Extract the image patch from each grid and compute 59-dimensional uniform LBP feature descriptor as local features. To quantize every descriptor into codewords by applying attribute enhanced sparse codewords after getting the local feature descriptor. Figure 2, Illustrate the system architecture.

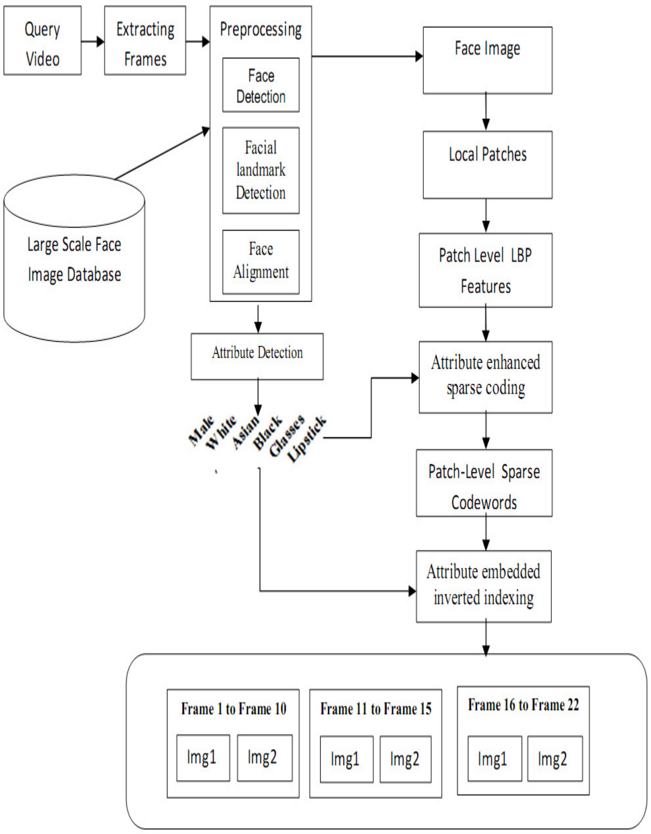
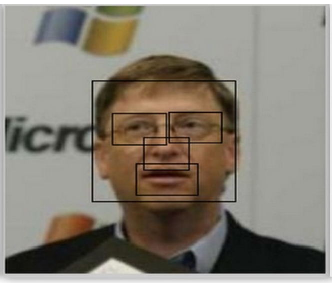


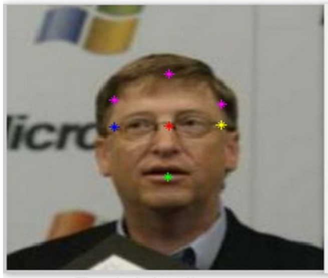
Figure 1: System architecture

The representation of human attributes in the sparse, use the dictionary selection to force images with different attribute values to contain different codewords. Then divide dictionary centroids into two different subsets, for the single human attribute. If the images with positive attribute score it will use one of the subset score and negative attribute score will use another subset. Consider an example, if an image has a positive male attribute score, they will use the first half of the dictionary centroids. If there is a negative male attribute score, it will use the second half of the dictionary centroids. By implementing this, images with different attributes will certainly have different codewords. Divide the sparse representation into multiple segments based on number of attributes, and every segment which is generated is depending on single attribute.

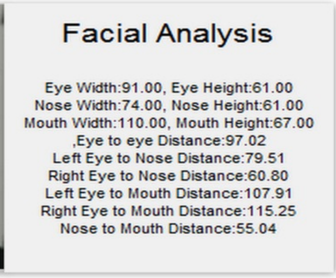
Original Face Image



Facial Parts Detection



Facial Attributes



Facial Analysis

Eye Width:91.00, Eye Height:61.00
Nose Width:74.00, Nose Height:61.00
Mouth Width:110.00, Mouth Height:67.00
,Eye to eye Distance:97.02
Left Eye to Nose Distance:79.51
Right Eye to Nose Distance:60.80
Left Eye to Mouth Distance:107.91
Right Eye to Mouth Distance:115.25
Nose to Mouth Distance:55.04

Face Features

A. *Attribute-enhanced sparse coding (ASC)*

We first introduce a way to use sparse coding for face image retrieval. We apply the same procedures to all patches in a single image and combine all these codewords together to represent the image.

We solve the following optimization problem using sparse coding for face image retrieval:

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

Wherex(i) is the original features extracted from a patch of face image i, DϵR$^{d\times K}$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. [25] Provides an efficient online algorithm for solving the above problem.

B. *Attribute Embedded Inverted Indexing (AEI)*

By using Attribute Embedded Inverted Indexing our aim to construct codewords enhanced by human attributes that may utilize the human attributes by adjusting the inverted index structure.

For every image, after computing the sparse representation we can use codeword set C(i) to represent it by taking non-zero entries in the sparse representation. Compute the similarity between two images is as follows,

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

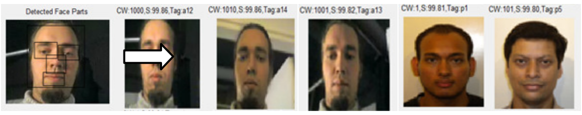
By using inverted index structure the similarity score can be efficiently found using image ranking. Attribute-embedded inverted index is built using the binary attribute signatures associated with all database images and the original codewords. The image ranking according to Equation (1) can still be efficiently computed using inverted index to check the hamming distance by simply doing a XOR operation before updating the similarity scores. As mentioned in [24], by skipping images with high hamming distance in attribute hamming space the XOR operation is faster than updating scores. The retrieval time significantly decreases.

V. Result and Discussion

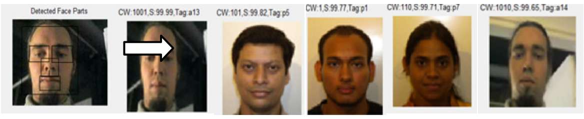
Datasets: We have used public datasets LFW for the following experiments. LFW dataset contains 13,233 face images among 5,749 people, and 12 people have more than 50 images. We take 10 images from each of these 12 people as our query set (120 images) and all other images as our database (13,113 images). Example images from the dataset can be found in Figure 5-1. The facial attribute scores of LFW are provided by [14], which use pre-trained facial attribute detectors to measure 73 attribute scores. Note that the 73 attribute scores for this datasets is also publicly available.

Experiments: by using above dataset for the images and the preprocessed video dataset which contain the face images. We are extracting the features of each video frame image by using above discussed methods and convert into the codewords and store the codewords into the dataset. The image ranking and retrieval by XORing the image codewords and the retrieval result of our system is shown below. The representation of images is from the video frames the first image is present in the video frame and the set of similar images is retrieved from the dataset.

Result from frame 1-10



Result from frame 11-20



VI. Conclusion

The video contains set of frames and each frame contains the one or more face images. To achieve a faster retrieval of face image from large scale database, by combining two different methods to use automatically detected human attributes. Combine automatically detected human attributes and low-level features for the content base image retrieval. The attribute enhanced sparse coding exploits the global structure and constructs the semantic aware codewords. By using this method we quantize the error and get better face image retrieval result. The indexing scheme can be easily integrated into inverted index and maintain scalable framework. The output will be the image which is occurring in the sequence of video frames.

REFERENCES

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

[2] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," IEEE Transactions on Information Forensics and Security,**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 con-straint," ACM Multimedia, 2011.

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

[5] 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**.

[6] Y. Freund, R.E. Schapire, "Experiments with a New Boosting Algorithm",In Proc. of the IEEE International Conference on Machine Learning (ICML), pp. 148–156, Bari, Italy, **1996**.

[7] P. Viola, M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), **2001**, pp. 511–518.

[8] J. Zobel and A. Moffat, "Inverted files for text search engines," ACMComputing Surveys, **2006**.

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

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

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

[12] 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**.

[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 PatternRecognition, **2011**.

[14] 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**.

[15] V. Blanz, S. Romdhani, and T. Vetter, "Face Identification across Different Poses and Illuminations with a 3D Morphable Model," Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition, **2002**.

[16] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 643-660, June **2001**.

[17] R. Gross, J. Shi, and J. Cohn, "Quo Vadis Face Recognition?" Proc. Workshop Empirical Evaluation Methods in Computer Vision, Dec.**2001**.

[18] 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**.

[19] 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**.

[20] 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**.

[21] W. J. Scheirer, A. Rocha, R.Michaels, and T. E. Boult. Meta-Recognition: The Theory and Practice of Recognition Score Analysis. IEEE TPAMI, 33(8):1689–1695, August **2011**.

[22] W. J. Scheirer, A. Rocha, R. Micheals, and T. E. Boult. Robust Fusion: Extreme Value Theory for Recognition Score Normalization. In ECCV, September **2010**

[23] 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**.

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

[25] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," ICML, **2009**.