THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Scalable Face Image Retrieval with Dynamic Allocation of Attribute Weights Using Attribute: Enhanced Sparse Codewords

S. Abirami
UG Student, Department of CSE, SNSCT, CBE, India
S. Indhu
UG Student, Department of CSE, SNSCT, CBE, India
S. Sachin Zechariah
UG Student, Department of CSE, SNSCT, CBE, India
Dr. S. Karthik
Professor & Dean, Department of CSE, SNSCT, CBE, India

Abstract:
Large-scale content-based face image retrieval is an enabling technology for many emerging applications. It utilizes automatically detected human attributes that contain semantic cues of the face photos to improve content based face retrieval by constructing semantic codeword's for efficient large-scale face retrieval. Here attributes are dynamically decided by depending upon its importance and then further exploit the contextual relationships between them. Then by leveraging human attributes in a scalable and systematic framework, two orthogonal methods named attribute-enhanced sparse coding and attribute embedded inverted indexing is proposed to improve the face retrieval in the offline and online stages. Experimental result shows better result when compare with existing one.

Key words: *Attribute -enhanced Sparse coding, Attribute embedded inverted indexing, Dynamic allocation, Face image*

1. Introduction

FACE recognition plays broad interests in pattern recognition , computer vision and machine learning areas. It is also one of the most successful applications of image analysis and understanding. A face recognition system identifies one person by comparing a query facial image with the registered images in a face database. Two major concerns in designing a face recognition system are: 1) the query images are subject to changes in illumination as well as occlusion and 2) the number of the recorded images is often tens of thousands . To address these concerns, we have to focus on two issues: 1) how to yield a robust representation for a query image, and 2) how to classify a query image as fast as possible. Simultaneously, numerous face representation and classification methods are developed.

Sparse representation for robust recognition on large-scale databases. All of the sparsity- based methods are applied to real-world face recognition problems and demonstrate to be very effective and robust to varying expression and illumination as well as occlusion and disguise. Impressive results were reported against many well-known face recognition methods. To deal with small sample problems like face recognition,where the dimension of images is larger than the training sample size, a dimensionality reduction (feature extraction) step becomes necessary before implementing SR. The choice of features is not critical, as long as the sparse representation is correctly computed and the number of features is sufficiently large. But, when the number of features is relatively small, there exist remarkable performance differences between different feature extraction methods. A small amount of representation features is preferable for the real-world face recognition problems, because it can reduce the storage requirements and improve the classification efficiency.

2. Related Work

In Face Matching and Retrieval Using Soft Biometrics, Local facial features have played an important role in forensic applications for matching face images. These features include any salient skin region that appears on the face. Scars, moles, and freckles are representative examples of the local facial features .The use of local features has become more popular due to the development of higher resolution sensors, an increase in face image database size, and improvements in image processing and computer vision algorithms. Local features provide a unique capability to investigate, annotate, and exploit face images in forensic applications by improving both the accuracy and the speed of face-recognition systems. This information is also necessary for forensic experts to give testimony in courts of law where they are expected to conclusively identify suspects .

Along with the facial marks, demographic information (i.e., gender and ethnicity) can also be considered as ancillary information that is useful in matching face images. The demographic information and facial marks are collectively referred to as soft biometric traits. Soft biometric traits are defined as characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals . The use of soft biometric traits is expected to improve the face-recognition performance when appropriately combined with a face matcher. On the other hand, when face images are occluded or partially damaged, soft biometric traits can be considered as alternative means of face matching or retrieval. Gender and ethnicity of a person typically do not change over the lifetime, so they can be used to filter the database to narrowdown the candidate face images. While some of the facial marks are not permanent, most of them appear to be temporally invariant, which can be useful for face matching and retrieval. When the face images are occluded or severely off-frontal, as is often the case in surveillance videos, utilizing the soft biometric traits is the only reliable evidence to narrow down the candidate face images Conventional face-recognition systems typically encode the face images by utilizing either local or global texture features. Local techniques first detect the individual components of the human face (i.e, eyes, nose, mouth, chin, and ears), prior to encoding the textural content of each of these components (e.g., EBGM and LFA) . Global (or holistic) techniques, on the other hand, consider the entire face as a single entity during the encoding process (e.g., PCA, LDA, Laplacianfaces, etc.)]. However, these techniques do not explicitly utilize local marks (e.g., scars and moles) and usually expect the input to be a full face image. The use and significance of soft biometric traits can be summarized

into four major categories: 1) supplement existing facial matchers to improve the identification accuracy; 2) enable fast face image retrieval; 3) enable matching or retrieval with partial or off-frontal face images; and 4) provide more descriptive evidence about the similarity or dissimilarity between face images, which can be used in the courts. Since facial marks capture the individual characteristics embedded in a face image that are not explicitly utilized in conventional face-recognition methods, a proper combination of face matcher and mark-based matcher is expected to provide improved recognition accuracy.. The presence of prominent facial marks in these two images strongly support the fact that these two images are of the same subject. The mark-based matcher helps in indexing each face image based on the facial marks (e.g., moles or scars). These indices will enable fast retrieval and also the use of textual or key-word-based query . Finally, marks can help in characterizing partial, occluded, or off-frontal face images, which will assist in matching or retrieval of tasks based on a partial face images often captured by surveillance cameras. We use demographic information (e.g., gender and ethnicity)

and facial marks as the soft biometrics traits. We labeled the gender of each face images into three categories (male, female, and unknown) and ethnicity into three categories (Caucasian, African-American, and unknown). We currently label gender and ethnicity manually as practiced in law enforcement. We propose a fully automatic facial mark extraction system using global and local texture analysis methods. We first apply the active appearance model (AAM) to detect and remove primary facial features such as eye brows, eyes, nose, and mouth. These primary facial features are subtracted from the face image. Then, the local irregularities are detected using the Laplacian- of-Gaussian (LoG) operator. The detected facial marks are used to calculate the facial similarity by their morphology and color along with the location. The mark-based matcher can be combined with a commercial face matcher in order to enhance the face matching accuracy or used by itself when the commercial face matcher fails in face matching process due to occlusion or unfavorable viewpoints. Our method differs significantly from the previous studies in the following aspects: 1) we use a number of soft biometric traits (i.e., gender, ethnicity, facial marks); 2) we extract all types of facial marks that are locally salient; 3) we focus on detecting facial marks and characterize each mark based on its morphology and color; and 4) we evaluate the performance using a state-ofthe- art face matcher on a large gallery with 10 213 subjects. The proposed soft biometric matching system will be especially useful to forensics and law enforcement agencies because it will 1) supplement existing facial matchers to improve the identification accuracy, 2) enable fast face image retrieval based on high level semantic query, 3) enable matching or retrieval from partial or off-frontal face images, and 4) help in discriminating identical twins.

3. Existing System

In existing system, Traditional CBIR techniques use image content like color, texture and gradient to represent images has been used. In order to deal with large scale data, mainly two kinds of indexing systems are used. Many studies have leveraged inverted indexing or hash-based indexing combined with bag-of-word model (BoW) and local features like SIFT , to achieve efficient similarity search. Automatically detected human attributes have been shown promising in different applications recently. The framework to deal with multi-attribute queries for keyword-based face image retrieval has been described. To do this Bayesian network approach is utilized the human attributes for face identification. Some of the issues in this method are it suffers from low recall problem due to semantic gap decrease the performance because of certain attributes (smiling, frowning, harsh lighting, etc.) and all attributes arte treated equally.

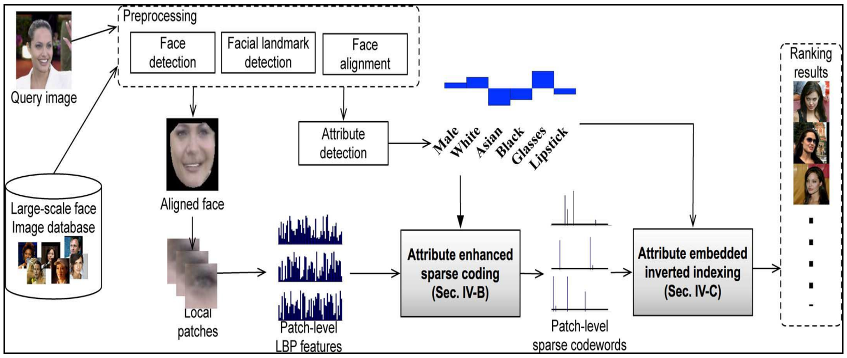

Figure 1

Fig. 1. 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 globallyin 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

4. Proposed System

Attributes are dynamically decided based on the importance of attributes and the contextual relationship can be exploited between them. To do this A score function is first used to calculate the score of each attribute value and a score matrix is constructed, and then it is transformed into a normalized score matrix. Based on the normalized score matrix, an entropy-based procedure is proposed to derive attribute weights. Furthermore, the additive weighted averaging operator is utilized to fuse all the normalized scores into the overall scores of alternatives, by which the ranking of all the given alternatives is obtained

4.1. Viola-Jones Face Detector

Rectangle features can be computed very rapidly using an intermediate representation for the image which we call the integral image.2 The integral image at location x,y
contains the sum of the pixels above and to the left of x,y inclusive:

Ii(x,y)= $\sum_{x' \leq x, \; y' \leq y} i(x' \; y')$

Where ii(x,y)is the integral image and i(x,y) is the original image. Using the following pair of recurrences:
S(x,y)=s(x,y-1)+i(x,y)
Ii(x,y)=ii(x-1,y)+s(x,y)
where s(x,y)is the cumulative row sum, s(x,-1)=0, and ii(-1,y)=0)the integral image can be computed in one pass over the original image.

4.1.1. Active Shape Model to Locate Facial Landmarks

A landmark represents a distinguishable point present in most of the images under consideration, for example, the location of the left eye pupil . We locate facial features by locating landmarks.A set of landmarks forms a shape. Shapes are represented as vectors: all the x- followed by all the y-coordinates of the points in the shape. We align one shape to another with a similarity transform (allowing translation, scaling, and rotation) that minimizes the average euclidean distance between shape points. The mean shape is the mean of the aligned training shapes (which in our case are manually landmarked faces). The ASM starts the search for landmarks from the mean shape aligned to the position and size of the face determined by a global face detector. It then repeats the following two steps until convergence (i) suggest a tentative shape by adjusting the locations of shape points by template matching of the image texture around each point (ii) conform the tentative shape to a global shape model. The individual template matches are unreliable and the shape model pools the results of the weak template matchers to form a stronger overall classifier. The entire search is repeated at each level in an image pyramid, from coarse to fine resolution.

4.2. Align Every Face with the Face Mean Shape

Using the landmarks detected by ASM, we tightly crop each face image and map it to the mean shape to simplify the mark detection and matching process. Let Si, i = 1,2….N represent the shape of each face image based on the 133 landmarks. Then, the mean shape is simply S = (1/N) i=1 Si. Each face image, Si, is mapped to the mean shape, S_I, by using Barycentric coordinate based texture mapping process. In this way, all face images are normalized in terms of scale and rotation and allows us to use the Euclidean distance based matcher in facial mark matching

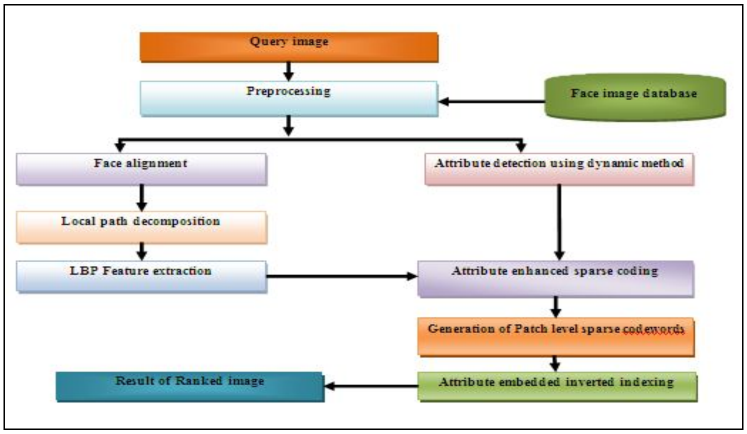

Figure 2

4.2.1. Extract an Image Patch and Computing LBP

We define a 5x7 grid at each detected component. Total we have 175 grids from five components. From each grid we extract a square image patch.Compute Local binary pattern of image patches can be computed.The local binary pattern operator works in a 3 × 3 pixel block of an image. The pixels in this block are thresholded by its center pixel value, multiplied by powers of two and then summed to obtain a label for the center pixel. As the neighborhood consists of 8 pixels, a total of $2^8 = 256$ different labels can be obtained depending on the relative gray values of the center and the pixels in the neighborhood.

$$LBP(X, Y) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p$$

4.2.2. Quantize Descriptor into Codewords using Attribute-Enhanced Sparse Coding

- Sparse Coding for Face Image Retrieval (SC)
  Using sparse coding for face image retrieval, we solve the following optimization problem:

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

Where $(x)^i$ is the original features extracted from a patch of face image i, $D \in R^{dXK}$ is a to-be-learned dictionary contains K centroids with dimensions.$V=[v^{(1)},v^{(2)},\ldots,v^{(n)}]$ 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.

5. Attribute-Embedded Inverted Index

To embed attribute information into index structure, for each image, in addition to sparse codewords $c^{(i)}$computed from the facial appearance, we used$_b$ a dimension binary signature to represent its human attribute,$b^{(i)}$

$$b_j^{(i)} = \begin{cases} 1 & \text{if } f_a^{(i)}(j) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The similarity score is then modified into,

$$S(i,j) = \begin{cases} \|c^{(i)} \cap c^{(j)}\| & \text{if } h(b^{(i)}, b^{(j)}) \le T \\ 0 & \text{otherwise.} \end{cases}$$

where h(i,j)denotes hamming distance between i and j , and T is a fixed threshold such that $0 \leq T \leq d_b$

5.1. Dynamically Weighted Attributed -Embedded Invert Index

Calculate conditional entropy in the pre-classified data set 16 for the descriptors or values of the current attribute One way to achieve this is to make Attribute Importance score as an inverse of the corresponding Conditional Entropy.Attribute Importance score is then transformed using Linear Transformation in the Interval of Variation for current attribute . The transformed score is calculated using the formula:

$$t_{h_1} = LB_i + \frac{(h_i - L_i)}{(H_i - L_i)} * (UB_i - LB_i)$$

- h_i=Attribute Importance score
- L_i=Lowest of Attribute value Importance scores for Attributei in Data Model
- H_i =Highest of Attribute value Importance scores for Attribute i in Data Model
- UB_I =Upper Bound of Interval of Variation for Attribute i
- LB_i: Lower Bound of Interval of Variation for Attribute i thiIAttribute importance score in Interval of Variation for Attribute

Normalize the Attribute Importance scores of all the attributes so that they sum up to the number of attributes. The normalized values are the dynamic weights of each of the attributes.

6. Conclusion

In this work two orthogonal methods to utilize automatically detected human attributes to significantly improve content-based face image retrieval has been described. First proposal of combining low-level features and automatically detected human attributes for content-based face image retrieval has been done. Attribute- enhanced sparse coding exploits the global structure and uses several human attributes to construct semantic-aware code words in the offline stage. In this attributes are chosen dynamically. Attribute-embedded inverted indexing further considers the local attribute signature of the query image and still ensures efficient retrieval in the online stage. Better experimental result has been achieved when compare with the existing technique.

7. 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. 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.
3. 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.
4. W. Scheirer, N. Kumar, K. Ricanek, T. E. Boult, and P. N. Belhumeur, “Fusing with context: A Bayesian approach to combining descriptive attributes,” in Proc. Int. Joint Conf. Biometrics, 2011.
5. 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.
6. 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.
7. 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, Univ. Massachusetts, Amherst, MA, USA, 2007, Tech. Rep. 07-49.
8. N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, “Attribute and simile classifiers for face verification,” in Proc. Int. Conf. Computer Vision, 2009.
9. T. Ahonen, A. Hadid, andM. Pietikainen, “Face recognition with local binary patterns,” in Proc. Eur. Conf. Computer Vision, 2004.
10. J. Zobel and A. Moffat, “Inverted files for text search engines,” ACM Comput. Surveys, 2006