iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** VII **Month of publication:** July 2014

DOI:

Vol. 2 Issue VII, July 2014

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Face Image Retrieval Techniques : A Survey

Abdul Afeef N

Computer Science Department, Viswajyothi College of Engineering and Technology

MG University, Kerala

***Abstract*— Due to popularity of digital devices, there are largely growing photos available in our life. Among them, a big percentage of are photos with human faces. Face image retrieval is an enabling technique to retrieve face images corresponding to a query face image. The survey investigates existing face image retrieval techniques developed in past years by various computer vision research communities across the world. Each technique has its own strengths and weaknesses. Different face image retrieval techniques uses different face feature extraction technique and feature indexing techniques for effective retrieval.**

***Keywords*—Content based face image retrieval, face image**

I. Introduction

Due to the popularity of digital devices and the rise of social network and photo sharing services (e.g., Facebook, Flickr), there are largely growing consumer photos available in our life. Among all those photos, a big percentage of them are photos with human faces. The importance and the sheer amount of human face photos make manipulations (e.g., search and mining) of large-scale human face images a really important research problem and enable many real world applications.

Given a query face image, content based face image retrieval techniques aims to retrieve similar face images to query from large database. It is an enabling technology for many applications including automatic face annotation, crime investigation, etc. Different methods uses different face feature representation like soft biometric, low level features and high level human attributes. Each technique has its own advantages and disadvantages. Face matching and retrieval using soft biometric has the problem of low descriminability in large databases but has the advantage of embedding soft biometric information in face feature extraction. The technique explained in [3] has advantage of improved retrieval results compared to [2] but has the limitation of requiring clean training data and massive human annotation. Sparse coding technique is used in [4] which embed identity constraints in sparse representation of ace feature representation has the advantage of better retrieval results and the limitation of requiring clean training data and massive human annotation. The method described in [5] has the advantage of combining attributes and fisher vectors for face retrieval technique. Scalable face image retrieval using attribute enhanced sparse codewords [6] uses both low level feature and high level human attributes for effective face image retrieval.

II. State of the Art

Several techniques are proposed in recent years for face image retrieval which uses different face feature representation and face retrieval strategy. Face matching and retrieval using soft biometrics [2] uses soft biometrics embedded in face such as gender and facial marks and this information combined with face matching score. Scalable face image retrieval with identity based quantization and multi-reference re-ranking [3] uses new scalable face representation using both local and global features. It uses a new multi-reference distance to re-rank candidate images using the hamming signature. Semi supervised face image retrieval using sparse coding with identity constraints [4] integrate partial identity information in sparse feature representation for better face image retrieval results. Combining attributes and

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

fisher vectors [5] improves the performance of retrieval of particular objects as well as categories. Scalable face image retrieval using attribute enhanced sparse codewords uses both low level features such as uniform LBP features and high level human attributes for better retrieval results.

A detailed study of few face image retrieval techniques are given below:

A. Face Matching and Retrieval using Soft Biometrics

Soft biometric traits embedded in a face (e. g., gender and facial marks) are ancillary information and are not fully distinctive by themselves in face-recognition tasks. However, this information can be explicitly combined with face matching score to improve the overall face-recognition accuracy. Moreover, in certain application domains, e.g., visual surveillance, where a face image is occluded or is captured in off-frontal pose, soft biometric traits can provide even more valuable information for face matching or retrieval. Facial marks can also be useful to differentiate identical twins whose global facial appearances are very similar. The similarities found from soft biometrics can also b e useful as a source of evidence in courts of law because they are more descriptive than the numerical matching scores generated by a traditional face matcher. Propose to utilize demographic information (e.g., gender and ethnicity) and facial marks (e.g., scars, moles, and freckles) for improving face image matching and retrieval performance. An automatic facial mark detection method has been developed that uses 1) the active appearance model for locating primary facial features (e.g., eyes, nose, and mouth), 2) the Laplacian-of-Gaussian blob detection, and 3) morphological operators.

Each facial mark is represented with an enclosing rectangular bounding box. The ground truth labeling process is performed by using the following ten categories provided by a forensics expert like freckle, mole, scar, whitening etc. Freckle is a single or a set of dark spots. When there is a dense set of spots in a small region, label each of the prominent dark spots rather than labeling the entire set with a single bounding box. Mole is referred to as an extruded region with typically dark skin color. In a 2-D facial image, it is difficult to distinguish between a spot and a mole. A mole typically appears larger in size and darker in colour compared with spots. Scar is the discoloured region of skin induced from a cut or injury. Pockmark is a crater-shaped scar. Acne is a red region caused by pimples or zits and stays for a few days to several months. Whitening represents a skin region that appears brighter compared with the surrounding region; it is observed more often with dark skinned people. When a larger region of skin is observed as dark, it is labelled as dark skin. While abrasion is not temporally invariant, it can later be related to the scars that are possibly caused by abrasions. Consider only large wrinkles and ignore small wrinkles especially around the eyes and mouth. Ignore beards and facial hair in constructing the ground truth. All other facial marks that do not belong to the nine groups mentioned above are labeled as other.

Face marks appear as salient localized regions on the face image. Therefore, a blob detector based on Difference of Gaussian (DoG) or LoG operator can be used to detect the marks. However, a direct application of a blob detector on a face image will generate a large number of false positives due to the presence of primary facial features (e .g., eyes, eye brows, nose, and mouth). Therefore, first localizing the primary facial features and then extracting facial marks in the rest of the face region is necessary for successful mark detection.

1) Advantages: Soft biometric features are used for feature extraction.

2) Disadvantages: Low descriminability for large database.

B. Scalable Face Image Retrieval with Identity Based Quantization and Multi Reference Re rankings

State-of-the-art image retrieval systems achieve scalability by using bag-of-words representation and textual retrieval methods, but their performance degrades quickly in the face image domain, mainly because they 1) produce visual words with low discriminative power for face images, and 2) ignore the special properties of the faces. The leading features for face recognition can achieve good retrieval performance, but these features are not suitable for inverted indexing as they are high-dimensional and global, thus not scalable in either computational or storage cost.

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

This paper aim to build a scalable face image retrieval system. For this purpose, develop a new scalable face representation using both local and global features. In the indexing stage, exploit special properties of faces to design new component-based local features, which are subsequently quantized into visual words using a novel identity-based quantization scheme. Use a very small hamming signature (40 bytes) to encode the discriminative global feature for each face. In the retrieval stage, candidate images are firstly retrieved from the inverted index of visual words. Use a new multi-reference distance to re-rank the candidate images using the hamming signature. On a one-million face database, show that local features and global hamming signatures are complementary, the inverted index based on local features provides candidate images with good recall, while the multi-reference re-ranking with global hamming signature leads to good precision.

Five facial components (two eyes, nose tip, and two mouth corners) are located on a detected face by a neural network based component detector. The face is then geometrically normalized by a similarity transform that maps the positions of two eyes to canonical positions. Define a 5x7 grid at each detected component. In total it have 175 grids fromfive components. From each grid extract a square image patch. A T3hS2 descriptor (responses of steerable filters) is then computed for each patch. All descriptors are quantized into visual words that are subsequently inverted indexed. Notice that the existing interest point based local feature detectors are not suitable for the face image. Such detectors tend to detect features in regions with rich textures or high contrast. They do not perform as well on face images since they contain mostly smooth textures. To enforce geometric constraints among features, assign each grid a unique ID, which is called position id. The position id will be concatenated with the feature quantization id to form the a visual word. By doing so, each visual word carries strong geometric information - two features can be matched only if they come from the same component and are extracted from the same grid in that component. This work proposed an identity-based quantization scheme using supervised learning.

1) Advantages: Results are improved compared to results in soft biometric based face image retrieval in [2].

2) Disadvantages: It requires clean training data and massive human annotations.

C. Semi Supervised Face Image Retrieval using Sparse Coding with Identity Constraint

This technique aim to develop a scalable face image retrieval system which can integrate with partial identity information to improve the retrieval result. To achieve this goal, first apply sparse coding on local features extracted from face images combining with inverted indexing to construct an efficient and scalable face retrieval system. Propose a novel coding scheme that refines the representation of the original sparse coding by using identity information. Using the proposed coding scheme, face images with large intra-class variances will still be quantized into similar visual words if they share the same identity. Experimental results show that system can achieve salient retrieval results on LFW dataset (13K faces) and outperform linear search methods using well known face recognition feature descriptors.

For every face image in database, first employ a frontal face cascade detector to find the location of the face, and then active shape model is applied to locating five facial components, including two eyes, two mouth corners and nose tip. The face image is then aligned using the locations of the eyes. After alignment, define 5x7 grids around each facial component to get a total of 175 grids. From each grid, extract a 59 dimension uniform LBP feature descriptor. All 175 descriptors then are first quantized into visual words separately by using sparse coding (Nonzero entries in the sparse representation are considered as visual words). Images with identity information are further refined to embed the information. Inverted index is built using the final sparse representation. Note that each descriptor from different grid locations will b e quantized separately, hence two visual words will match only if it is extracted from the same grid location. Through this way, it can encode important geometric information of face into visual words. Given a query image, apply the same detection, alignment, and feature extraction methods described above. The extracted features are then quantized into visual words by sparse coding and used for

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

retrieving inverted index. In the retrieval stage, simply use histogram intersection to compute similarity score.

1) Advantages: Identity constraints included in traditional sparse coding for face image retrieval. Results are improved compared to sparse coding for face image retrieval without identity constraints.

2) Disadvantages: It requires clean training data and massive human face annotations.

D. Combining Attributes And Fisher Vectors For Efficient Image Retrieval

Attributes were recently shown to give excellent results for category recognition. In this paper, demonstrate their performance in the context of image retrieval. First, show that retrieving images of particular objects based on attribute vectors gives results comparable to the state of the art. Second, demonstrate that combining attribute and Fisher vectors improves performance for retrieval of particular objects as well as categories. Third, implement an efficient coding technique for compressing the combined descriptor to very small codes. Experimental results on the Holidays dataset show that approach significantly outperforms the state of the art, even for a very compact representation of 16 bytes per image. Retrieving category images is evaluated on the web-queries dataset. It shows that attribute features combined with Fisher vectors improve the performance and that combined image features can supplement text features.

Fisher vectors are a means of aggregating local descriptors into a global descriptor. Local descriptors are computed by extracting orientation and scale invariant Hessian affine interest points and by describing their neighbourhoods using the SIFT descriptor (reduced to 64 dimensions by PCA). The position information of the points is not included in the descriptor. During a preliminary learning stage, a 64 centroid Gaussian mixture model (GMM) was computed to fit the distribution of local descriptors in a dataset of unrelated images. The distribution of local descriptors of an image has a likelihood with respect to this GMM. The Fisher descriptor is the derivative of this likelihood with respect to the GMM parameters. Then restrict the parameters for which it compute derivatives to the means of the Gaussians, so descriptor has 64x64 = 4096 dimensions. Fisher descriptors were shown to outperform BOF as a global descriptor for image classification and retrieval.

Each attribute corresponds to a term from a vocabulary. For an image, the attribute descriptor encodes how relevant each term is to describe that image. Attribute descriptors are computed from image classifiers built for each of the terms. Images are often associated with text. For example, on photo sharing sites there are tags and user comments, in photo banks indexing terms are associated with the images, and on random web pages, the text surrounding an image is likely to be relevant. If a dataset has text associated with the images, build a basic text descriptor from these annotations. Remove punctuation and convert all text to lowercase, tokenize it into words and build a dictionary from all the words found in the corpus. Remove stopwords and words that are too rare in the corpus. Describe each image with a (sparse) histogram of the words appearing in its annotations. The histogram is L2 normalized and apply TF-IDF weighting to favour infrequent words. Histograms are compared with scalar products. Images are represented by global descriptors, i.e., Fisher vectors and attribute features. Retrieval consists in finding the nearest neighbours in a high-dimensional descriptor space. To combine Fisher vectors and attribute features, each of them should be normalized. Attribute vectors contain SVM classification scores. These scores are approximately Gaussian distributed. It have empirically observed that normalizing the vectors with L2 or L1 norm decreases the retrieval performance. Fisher vectors and attribute vectors can now be compared with the L2 distance, and have the same magnitude on average. To combine them, simply add up the squared L2 distances. The performance can be improved by using a weighting factor to increase the contribution of the Fisher vector. To accelerate retrieval, project the vectors to a lower dimension. For non-sparse vectors up to a few thousand dimensions compared with the L2 distance, a good choice is to apply a PCA transform. Dimensionality of VLAD descriptors (similar to Fisher) can be reduced using PCA with only a small loss of discriminative power.

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

1) Advantages: Attributes and fisher vectors are combined for better general image retrieval.

2) Disadvantages: Do not take advantages of human attributes.

E. Attribute and Simile Classifiers for Face Verification

It present two novel methods for face verification. First method - "attribute classifiers" uses binary classifiers trained to recognize the presence or absence of describable aspects of visual appearance (e.g., gender, race, and age). Second method - "simile classifiers" removes the manual labelling required for attribute classification and instead learns the similarity of faces, or regions of faces, to specific reference people. Neither method requires costly, often brittle, alignment between image pairs; yet, both methods produce compact visual descriptions, and work on real-world images. Furthermore, both the attribute and simile classifiers improve on the current state-of-the-art for the LFW data set, reducing the error rates compared to the current best by 23:92 percent and 26:34 percent, respectively, and 31:68 percent when combined. For further testing across pose, illumination, and expression, introduce a new dataset termed "PubFig" of real-world images of public figures (celebrities and politicians) acquired from the internet. This data set is both larger (60,000 images) and deeper (300 images p er individual) than existing data sets of its kind. Finally, present an evaluation of human performance.

The first step of this approach is to extract low-level features from different regions of the face, e .g., normalized pixel values, image gradient directions, or histograms of edge magnitudes. Its aim is to design a face verification method that is tolerant of image changes, second step is to use these low-level features to compute high-level visual features, or traits, which are insensitive to changes in pose , illumination, and expression. These visual traits are simply scores of trait classifiers (attribute or simile). To perform face verification on a pair of images, compare the scores in both images. Steps are formalized below.

1. Extract low-level features : For each face image I, extract the output of k low-level features and concatenate these vectors to form a large feature vector.
2. Compute Visual Traits : For each extracted feature vector F(I), compute the output of n trait classifiers.
3. Perform Verification : To decide if two face images I1 and I2 are of the same person, compare their trait vectors using a final classifier.

1) Advantages: Due to the usage of attributes, better results obtained.

2) Disadvantages: Large amount of training data required or attribute and simile classifiers.

F. Scalable Face Image Retrieval Using Attribute-Enhanced Sparse Codewords

The goal of paper is to address one of the important and 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. Traditional methods for face image retrieval usually use low-level features to represent faces.

This paper provide a new perspective on content-based face image retrieval by incorporating high-level human attributes into face image representation and index structure. 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, this method can able to find better feature representations and achieve better retrieval results. Human attributes (e.g., gender, race, hair style) are high-level semantic descriptions about a person. The recent work shows automatic attribute detection has adequate quality (more than 80 percentage accuracy) on many different human attributes.

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

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

automatically detected by attribute detectors for improving content-based face image retrieval, this paper 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 online 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, this paper build a large-scale content-based face image retrieval system by taking advantages of both low-level (appearance) features and high-level (facial) semantics.

1) Advantages: Better results in large image database and it is very scalable

2) Disadvantage: Attribute detection needs large amount of training data.

III. Conclusions

This survey analysed various face image retrieval techniques. Different face image retrieval techniques used corresponding feature extraction methods from the face images to give the description of face. Different face indexing techniques also used to index the face descriptors to retrieve face images corresponding to query image. Various advantages and disadvantages of different face image retrieval techniques are also analysed. It is also helpful to explore new face image retrieval technique.

Acknowledgment

The author would like to express her sincere thanks to HOD, group tutor and staff in Computer Science department, Viswa Jyothi College of Engineering and Technology for many fruitful discussions and constructive suggestions during the preparation of this manuscript.

References

[1] Bor-Chun Chen, Yan-Ying Chen, Yi n-Hsi Kuo, and Wi nston H. Hsu, "Scalable Face Image Retrieval Using Attribute-Enhanced Sparse Codewords", in IEEE Trans on Multimedia, vol. 15, no. 5, Aug 2013.

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

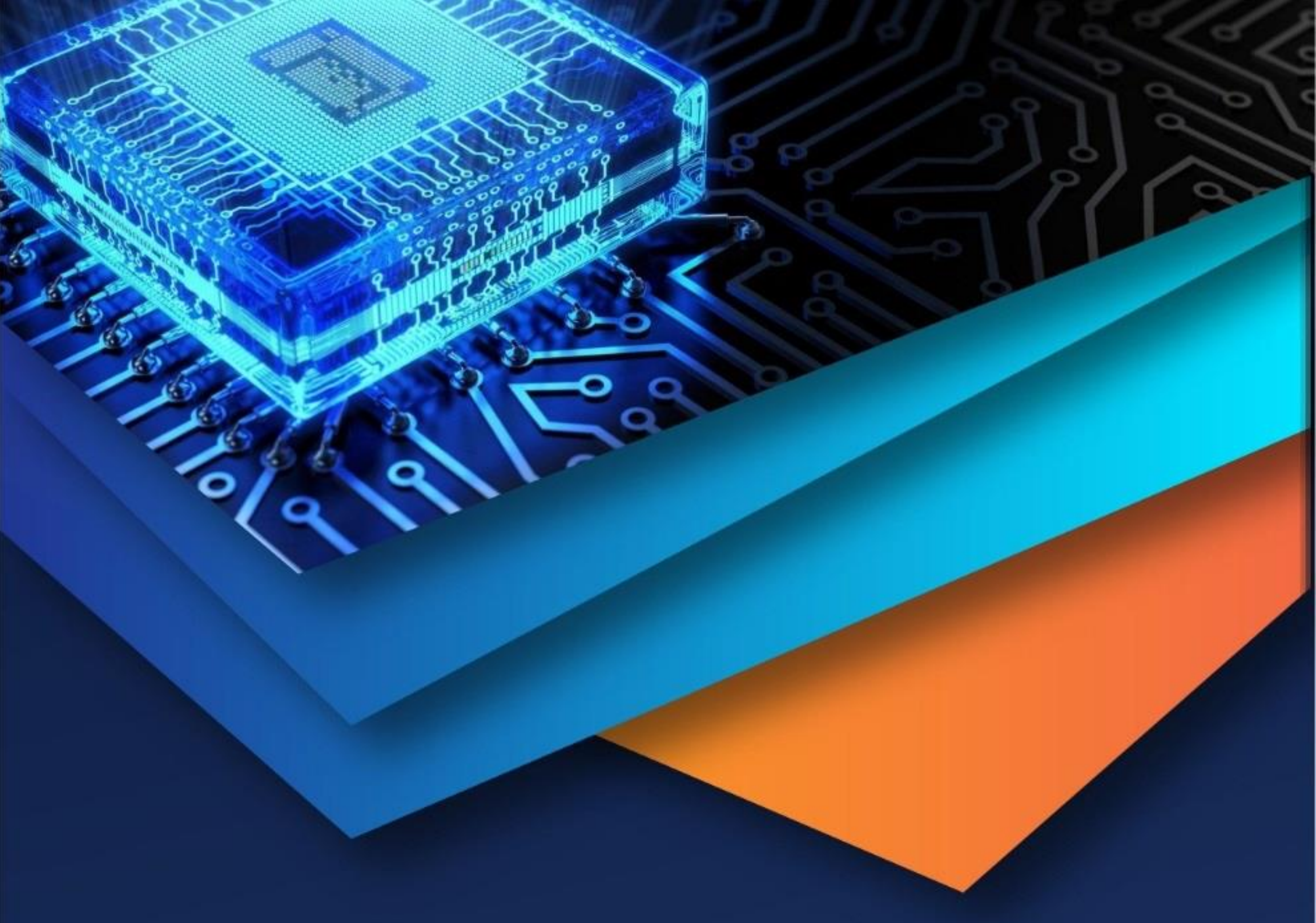
[3] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, "Scalable face image retrieval with identity-based quantization and multi-reference re-ranking", inProc. IEEE Conf. Computer Vision and Pattern Recognit, 2010.

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

[5] 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.

[6] 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.

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



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY