www.ioirp.com
International Journal of Innovative Research in Computer Science and Engineering (IJIRCSE)
ISSN: 2394-6364, Volume – 1, Issue – 6. May 2015

Identification of face in Videos Using Color Edge Based Re-Ranking and Fusion with PSO-SVM

Remya Chandran
ME CSE, Maharaja Prithvi Engineering College, Avinashi, India
nivedika@gmail.com

***Abstract*— Video-based face recognition has gained significant attention. A three-stage approach is proposed for optimizing ranked lists across multiple video frames and fusing them into a single composite ordered list to compute the video signature. This signature embeds diverse intra-personal variations and facilitates in matching two videos with large variations. For matching two videos, a discounted cumulative gain measure is utilized, which uses the ranking of images in the video signature as well as the usefulness of images in characterizing the individual in the video. This face recognition technique support vector machine and particle swarm optimization (SVM-PSO) for developing real-time face recognition systems. The integrated scheme aims to adopt the SVM-PSO method to improve recognition systems on dynamically visual perception. The efficacy of the proposed algorithm is demonstrated on the YouTube faces database and the MBGC v2 video challenge database that comprise different types of video-based face recognition challenges such as matching still face images with videos and matching videos with videos.**

Keywords*— *Face Recognition, Video, Surveillance, PSO-SVM

I. Introduction to Image Processing

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them.

It is among rapidly growing technologies today, with its applications in various aspects of a business. Image Processing forms core research area within engineering and computer science disciplines too.

Image processing basically includes the following three steps.

1. Importing the image with optical scanner or by digital photography.
2. Analyzing and manipulating the image which includes data compression and image enhancement and spotting patterns that are not to human eyes like satellite photographs.
3. Output is the last stage in which result can be altered image or report that is based on image analysis.

Purpose of Image processing

The purpose of image processing is divided into 5 groups. They are:

- Visualization - Observe the objects that are not visible.
- Image sharpening and restoration - To create a better image.
- Image retrieval - Seek for the image of interest.
- Measurement of pattern – Measures various objects in an image.
- Image Recognition – Distinguish the objects in an image.

The two types of methods used for Image Processing are Analog and Digital Image Processing. Analog or visual techniques of image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. The three general phases that all types of data have to undergo while using digital technique are Pre- processing, enhancement and display, information extraction.

In modern sciences and technologies, images also gain much broader scopes due to the ever growing importance of scientific visualization (of often large-scale complex scientific/experimental data). Examples include microarray data in genetic research, or real-time multi-asset portfolio trading in finance.

A. Related Research

Face recognition in unconstrained videos is challenging due to large variations in pose, illumination, expression etc. In this face recognition system , address the problem from two different aspects: To handle pose variations, learn a Structural-SVM based detector which can simultaneously localize face fiducially points and estimate face pose. By adopting a different optimization criterion from existing algorithms, possible to improve localization accuracy.Face variations model of other kinds using intra-personal/extra-personal dictionaries. The proposed framework is advantageous in terms of both accuracy and scalability. Itdemonstrate through experiments that the algorithm outperforms state-of-the-art approaches on challenging public databases.

B. Research Contributions

This system presents a video-based face recognition algorithm that computes a discriminative video signature as an ordered list of still face images from a large dictionary. A three-stage approach is proposed for optimizing ranked lists across multiple video frames and fusing them into a single composite ordered list to compute the video signature. This signature embeds diverse intra-personal variations and facilitates in matching two videos with large variations. For matching two videos, a discounted cumulative gain measure is utilized, which uses the ranking of images in the video signature as well as the usefulness of images in characterizing the individual in the video. The efficacy of the proposed algorithm is evaluated under different video-based face recognition scenarios such as matching still face images with videos and matching videos with videos. The efficacy of the proposed algorithm is demonstrated on the YouTube faces database and the MBGC v2 video challenge database that comprise different types of video-based face recognition challenges such as matching still face images with videos and matching videos with videos. Performance comparison with the benchmark results on both the databases and a commercial face recognition system shows the efficiency of the proposed algorithm for video-based face recognition.

II. DICTIONARY AND PSO-SVM BASED VIDEO FACE RECOGNITION

A. Dictionary

Dictionary is a large collection of still face images where every individual has multiple images capturing a wide range of intra-personal variations i.e. pose, illumination, and expression variations. In this approach, definition of dictionary is different from the dictionary in sparse representation based approaches. They represent a dictionary as a collection of atoms such that the number of atoms exceeds the dimension of the signal space, so that any signal can be represented by more than one combination of different atoms. In this research, the dictionary comprises 38, 488 face images pertaining to 337 individuals from the CMU Multi-PIE database. OpenCV's boosted cascade of Haar-like features provide the face boundaries and eye-coordinates. These boundaries are used to detect and crop faces from the dictionary images and eye-coordinates are used to normalize the detected image with respect to rotation. The normalized face images are resized to 196×224 pixels with inter-eye distance of 100 pixels.

B. Computing Ranked List

Let V be the video of an individual comprising n frames where each frame depicts the temporal variations of the individual. Face region from each frame is detected and preprocessed. Face regions corresponding to different frames across a video are represented as {F1, F2, . . . , Fn }. To generate ranked lists, each frame is compared with all the images in the dictionary. Since the dictionary consists of a large number of images and each video has multiple frames; it is essential to compute the ranked list in a computationally efficient manner. Linear discriminant analysis (LDA), a level-1 feature, is therefore used to generate a ranked list by congregating images from the dictionary that are similar to the input frame. A linear discriminant function is learnt from the dictionary images that captures the variations in pose, illumination, and expression. The linear discriminant function learns these variations and retrieves images from the dictionary that are similar to the input video frame i.e. images with similar pose, illumination, and expression. The ranking of retrieved images from such a dictionary is found to be more discriminative for face recognition than that of a signature based on the pixel intensities or some image features. Each column of the projection matrix W represents a projection direction in the subspace.

www.ioirp.com
International Journal of Innovative Research in Computer Science and Engineering (IJIRCSE)
ISSN: 2394-6364, Volume – 1, Issue – 6. May 2015

C. Clustering

Multiple frames in video exhibit different intra-personal variations; therefore, each ranked list positions dictionary images based on the similarity to the input frame. Images in the ranked list are further partitioned into different clusters such that if an image in a cluster has high similarity to the input frame, then all images in that cluster tend to be more similar to the input frame. The main idea behind clustering is to congregate images in a ranked list into different clusters where each cluster represents a particular viewing condition i.e. a specific pose, illumination or expression. Let **R**i be the i th ranked list of a video corresponding to frame Fi, then {Ci,1,Ci,2, . . . ,Ci,k } form k clusters of **R**i . In this research, K-means clustering which is an unsupervised, nondeterministic technique for generating a number of disjoint and flat (non-hierarchical) clusters is used to cluster similar images with an equal cardinality constraint. To guarantee that all clusters have equal number of data points, k centroids are initially selected at random. For each point, similarity to the nearest cluster is computed and a heap is build.

A data point is drawn from the heap and assigned to the nearest cluster, unless that cluster is already full. If the nearest cluster is full, distance to the next nearest cluster is computed and the data is re-inserted into the heap. The process is repeated till the heap is empty i.e. all the data points are assigned to a cluster. It guarantees that all the clusters contain equal number of data points (±1 data points per cluster). K-means clustering is used as it is computationally faster and produces tighter clusters than hierarchical clustering techniques. After clustering, each ranked list **R**i has a set of clusters Ci,1, Ci,2, . . ., Ci,k, where k is the number of clusters. K-means clustering is affected by the initialization of initial centroid points; however,start with five different random initializations of k clusters. Finally, clusters which minimize the overall sum of square distances are selected.

D. Re-Ranking

Clusters across multiple ranked lists overlap in terms of common dictionary images. Since the overlap between the clusters depends on the size of each cluster, it is required that all the clusters should be of equal size. Higher the overlap between the clusters, more likely that they contain images with similar appearances (i.e. with similar pose, illumination, and expression). Based on this hypothesis, the reliability of each cluster is computed as the weighted sum of similarities between the cluster and other clusters across multiple ranked lists.

E. Fusion

The ranked lists across multiple frames have redundant information and matching such ranked lists across two videos can be computationally inefficient. Therefore, it is imperative to compute a composite ranked list as the video signature. Once the similarity scores of images are adjusted across all the ranked lists, multiple ranked lists are fused into a final composite ranked list, R_. The final similarity score of an image d (denoted as SSd) is the average of adjusted similarity scores of image d across all the ranked lists.

Algorithm: Fusing Ranked Lists With Clustering and Re-Ranking

Input:A set of ranked lists R1,R2,….,Rn from multiple frames in a vedio V.
Iterate: i=1 to n(number of ranked list)
Clustering: Partition ranked list Ri into different clusters $C_{i,1}, C_{i,2,\ldots}, C_{i,k,}$ where k is the number of clusters.
End iterate.
Iterate: i=1 to n, j=1 to k.
Reliability: Compute reliability of cluster $r(C_{i,j})$.
Re-ranking:Adjust the similarity score of each image d based on the reliability of the cluster it belongs.
$Sim_j^*(d)=Sim_i(d)*(1+r(C_{c,j})),d € C_{i,j.}$
end iterate.
Fusion: Compute an ordered composite ranked list R where similarity score of an image d is given as:

Output:Final composite ranked list R' for video V.
PSO-SVM

$$Ss_d=\frac{\sum_{i=1ton} Sim_j^*(d)}{n}$$

The face recognition technique makes use of support vector machine and particle swarm optimization (SVM-PSO) for developing real-time face recognition systems. The integrated scheme aims to adopt the SVM-PSO method to improve recognition systems on dynamically visual perception.

III. Experimental Results

The efficacy of the proposed algorithm for video based face recognition is evaluated in verification mode.

1. YouTube Faces Database: The performance of the proposed algorithm is evaluated using the experimental protocol defined by Wolf et al.. In this experiment both gallery and probe consist of videos and training is performed as two class problem with 'same'/'not-same' labels. In this experiments, ten splits provided along with the database are used. Training is performed on nine splits and the performance is computed on the tenth split. The final performance is reported as an average of 10 folds. In this protocol, the information about the subject's label associated with the video is discarded and only the information about whether a pair is 'same' or 'notsame' is retained. The experiments are also performed to evaluate the performance enhancement due to different stages of the proposed algorithm on the YouTube faces database. Firstly, to evaluate the performance gain due to clustering based re-ranking and fusion steps, the performance is compared when ranked list across multiple frames are combined using the MNF approach. Secondly, to evaluate the effectiveness of the DCG measure, the performance is evaluated when two ranked lists are compared using the distance measure proposed by Schroff et al. Their distance measure only considers the overlap between two ranked lists and ignores other information such as relevance of images in the ranked list. It should be noted that while evaluating the gain in performance due to an individual step, all other steps in the proposed algorithm remain the same.

2. Multi Biometric Grand Challenge v2 Database: Multiple experiments are performed on this database to evaluate the efficacy of the proposed algorithm. Specifically, the algorithm is evaluated for two different scenarios: (1) matching still face images with videos and (2) matching videos with videos. a) Matching still face images with videos: In many real world applications, such as surveillance, it is required to match still face images with videos for authenticating the identity of individuals. In this experiment, still face images from the MBGC v2 still portal and videos (comprising both walking and activity videos) from the MBGC v2 video challenge database pertaining to 147 subjects are used. To evaluate the efficacy of the proposed algorithm, experiments are performed with 10 times repeated random subsampling (cross validations). In each experiment, training is performed on 47 subjects and the performance in reported on the remaining 100 subjects. This experiment further comprises two different subsets: Matching video probe with still gallery images: In this experiment, the probe is a video of an individual whose identity is to be matched against a gallery of still face images. The ranked list of an image in the gallery is computed by positioning the images retrieved from the dictionary based on their level-1 similarity scores. The composite ranked list of a probe video is then compared with the ranked list computed for each of the gallery images. The experiment is further divided as: 1) probe comprises 618 walking videos pertaining to 100 subjects and 2) probe comprises 513 activity videos pertaining to 100 subjects. In both the cases the gallery consists of 100 still face images, one image per subject.

Matching still probe with video gallery: In this experiment, the probe is a still face image and the gallery comprises videos. The ranked list of a still probe image is compared with the composite ranked list of each video in the gallery. The experiment is divided as: 1) gallery comprises 100 walking videos and 2) gallery comprises 100 activity videos. In both the cases, the probe comprises 1543 still face images pertaining to 100 subjects. b) Matching videos with videos: The proposed algorithm is evaluated for face verification when both gallery and probe comprise videos of individuals. The performance of the proposed algorithm on the MBGC v2 video challenge database is evaluated under three different scenarios, 1) walking vs walking (WW), 2) walking vs activity (WA), and 3) activity vs activity (AA).

IV. Conclusion

Multiple frames in a video provide temporal and intra-class variations that can be leveraged for efficient face recognition. The face recognition method can be corporate with support vector machine and particle swarm optimization (SVM-PSO) for developing real-time face recognition systems. The integrated scheme aims to adopt the SVM-PSO method to improve recognition systems on dynamically visual perception. Face recognition method can modified by detecting the global feature: Left Eye, Right Eye, Nose and mouth.

References

[1] "On Recognizing Faces in Videos using Clustering based Re-Ranking and Fusion" Himanshu s .Bhatt,RichaSingh,member ,IEEE, and MayankVatsa, member IEEE.

[2] "Image sets alignment for video-based face recognition" Z. Cui, S. Shan,H. Zhang, S. Lao, and X. Chen.

[3] "Memetically optimized MCWLD for matching sketches with digital face images" H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa.

[4] "Dictionary based face recognition from video" C. Chen, V. M. Patel, P. J. Phillips, and R. Chellappa.

[5] 'Fusing robust face region descriptors via multiple metric learning for face recognition in the wild" Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen

[6] "Face Recognition with Image Sets Using Manifold Density Divergence"Ognjen Arandjelov,Gregory,John Fisher.

[7] "Face Recognition Using Eigen Faces and Artificial Neural Network" MayankAgarwal, Nikunj Jain, Mr. Manish Kumar and HimanshuAgrawal.

[8] **"**The Kernel Orthogonal Mutual Subspace Method" Kazuhiro Fukui1and Osamu Yamaguchi.

[9] "Probabilistic recognition of human faces from video" ShaohuaZhou,Volker Krueger, and Rama Chellappa.