Dictionary Based Video Face Recognizing Using Clustering-Based Re-Ranking and Fusion

VALENTINA | S.SATHISH KUMAR | M.VARATHARAJ
[1](Assistant professor, ECE PG department, Christ the King Engineering College, Coimbatore, INDIA, valanteenadavid@gmail.com)
[2](ECE PG department, Christ the King Engineering College, Coimbatore, INDIA, ss.sathish555@gmail.com)
[3](Head of the Department of ECE, Christ the King Engineering College, Coimbatore, INDIA, varatharaj_ms80@rediffmail.com)

Abstract— *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 efficiency 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.*

Keywords— face recognition; discriminative video signature; cumulative gain measure

1. INTRODUCTION

Nowadays With the increase in usage of camera technology in both surveillance and personal applications, enormous amount of video feed is being captured everyday. Surveillance cameras are also capturing significant amount of data across the globe. In terms of face recognition, the amount of data collected by surveillance cameras every day is probably more than the size of all the publicly available face image databases combined. One primary purpose of collecting the data from surveillance cameras is to detect any unwanted activity during the act or at least enable to analyze the events and may be determine the person(s) of interest after the act. While face recognition is a well-studied problem and several algorithms have been proposed a majority of the literature is on matching still images and face recognition from videos is relatively less explored. Recognizing the individuals appearing in videos has both advantages and disadvantages compared to still face matching the information available in a video is generally more than the information available for matching two still images. the information available in a video is generally more than the information available for matching two still images. video-to-still/still-to-video face recognition techniques can be broadly categorized into frame selection and multi-frame fusion approaches. On the other hand, in multi-frame fusion approaches, recognition results of multiple frames are fused together. In this project we propose k-mean clustering algorithm of segmentation and Euclidean distance for ranking the images and future level fusion for fusing the images. By this algorithm we can recognize the images captured in the video in a Clustering-Based Re-Ranking and Fusion method and histogram intersection for identifying the image.

2. DICTIONARY 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. 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. For a given video pair, frames from each video are extracted and pre-processed. Face region from each frame is detected and resized to 196 × 224 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 $\{F1, F2, \ldots, Fn\}$. 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.

3. LINEAR DISCRIMINANT ANALYSIS (LDA)

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 and the projection of an image onto the subspace is computed as:

$$Y = WT\,X$$

where *X* is the input image and *Y* is its subspace representation. The input frame *Fi* and all images in the dictionary are projected onto the subspace.

A. Clustering.

- Multiple frames in a 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. 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* 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.

B. Ranking.

Euclidean distance is the "ordinary" distance between two points that one would measure with a ruler, and is given by the Pythagorean formula. By using this formula as distance, Euclidean space becomes a metric space. The associated norm is called the Euclidean norm. Older literature refers to the metric as Pythagorean metric. The Euclidean distance between points **p** and **q** is the length of the line segment connecting them (**PQ).** Euclidean distance in LDA projection space, 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, we start with five different random initializations of k clusters. Finally, clusters which minimize the overall sum of square distances are selected.

C. 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. The *reliability r(Cl, j)* of a cluster *Cl, j* in ranked list *q* is computed.

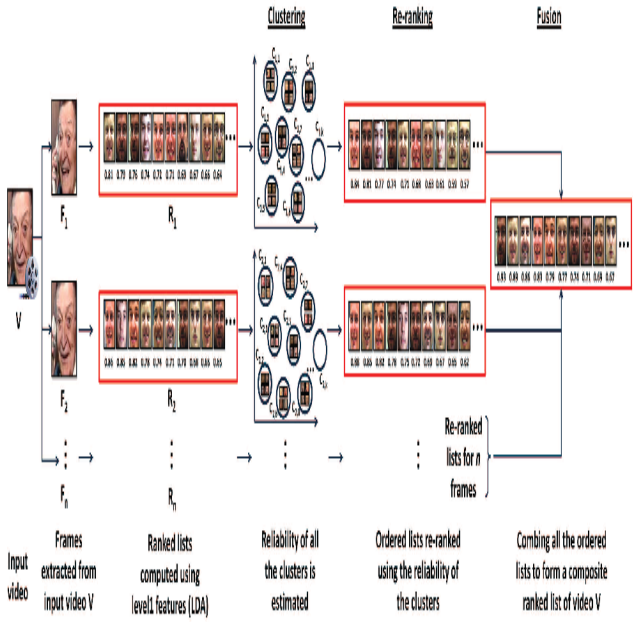
D. Fusion..

- 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
- Feature level methods are the next stage of processing where image fusion may take place. Fusion at the feature level requires extraction of features from the input images. Features can be pixel intensities or edge and texture features . The Various kinds of features are considered depending on the nature of images and the application of the fused image.
- The features involve the extraction of feature primitives like edges, regions, shape, size, length or image segments, and features with similar intensity in the images to be fused from different types of images of the same geographic area
- These features are then combined with the similar features present in the other input images through a pre-determined selection process to form the final fused image . The feature level fusion should be easy.
- However, feature level fusion is difficult to achieve when the feature sets are derived from different algorithms and data sources

E. Dictionary Based Video Face Recognition Algorithm.

- **Step-1:** For a given video pair, frames from each video are extracted and pre-processed. Face region from each frame is detected and resized to 196 × 224 pixels.
- **Step-2:** For each frame in the video, a ranked list of still face images from the dictionary is computed using level-1 features. The retrieved dictionary images are arranged in a ranked list such that the image with the maximum similarity score is positioned at the top of the list.
- **Step-3:** Ranked list across multiple frames of a video are combined to form a video signature using clustering based re-ranking and fusion.
- **Step-4:** To match two videos, their video signatures are compared using the *nDCG* measure that incorporates scores computed using both level-1 (rank) and level-2 (relevance) features

4. BLOCK DIAGRAM

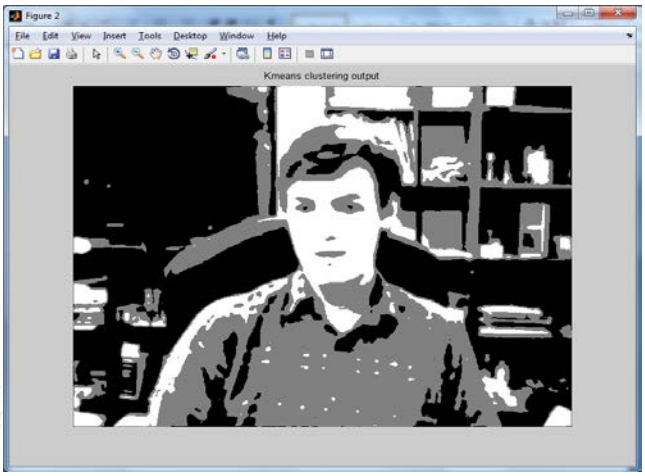


A. Experimental Results.

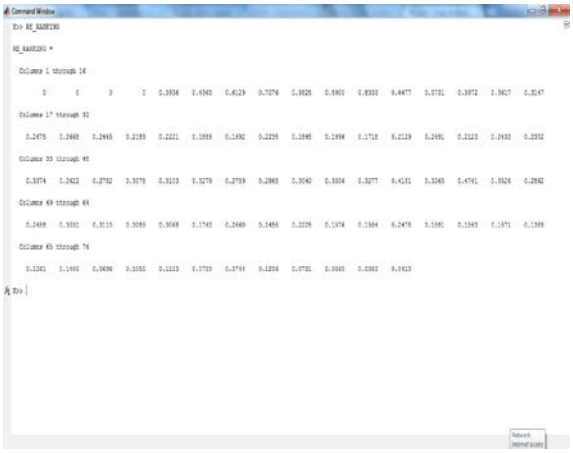
The efficacy of the proposed algorithm is evaluated on multiple databases under different scenarios such as video-to-still, still-to-video, and video-to-video. The input is given in a video mode and the proposed algorithm takes the input and process with the clustering based re-ranking and fused the image with the input given and recognize the matched person in a comparison with histogram intersection and the output is taken to identify the person correctly.

B. Input Video.

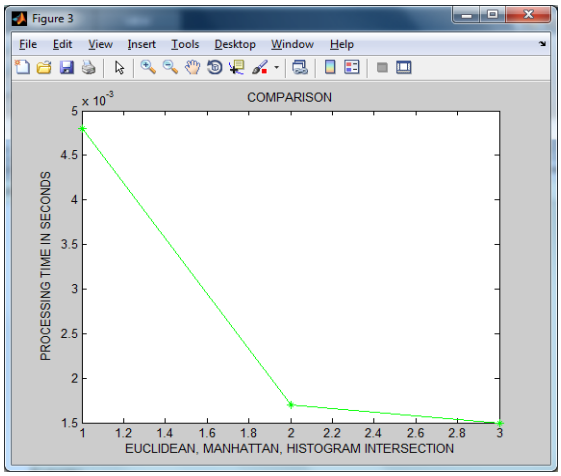
C. K-Means Clustering Output.



D. MATLAB Coding.



E. Histogram Intersection.



.

5. CONCLUSION

The proposed video based face recognition algorithm is based on the observation that a discriminative video signature can be generated by combining the abundant information available across multiple frames of a video. It assimilates this information as a ranked list of still face images from a large dictionary. The algorithm starts with generating a ranked list for every frame in the video using computationally efficient level-1 features. Multiple ranked lists across the frames are then optimized using clustering based re-ranking and finally fused together to generate the video signature. In future work the clustering based Re-ranking can be modified by fuzzy C- means clustering algorithm. To get better and efficient clustering output we may proceed with fuzzy algorithm.

REFERENCES

[1] Himanshu S. Bhatt, Richa Singh, *Member, IEEE*, and Mayank Vatsa, *Member, IEEE"*On Recognizing Faces in Videos Using Clustering-Based Re-Ranking and Fusion" *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 9, NO. 7, JULY 2014*

[2] G. Aggarwal, A. K. R. Chowdhury, and R. Chellappa, "A system identification approach for video-based face recognition," in *Proc. 17th ICPR*, 2004, pp. 175–178.

[3] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell,"Face recognition with image sets using manifold density divergence,"in *Proc. IEEE Int. Conf. CVPR*, Jun. 2005, pp. 581–58