VB-BIDE: Video Based-Burglary Identification and Detection through E-mail

[1]S. Kayalvili, [2]R. Pavithra, [3]S. Thilluja, [4]N. Vaishali

[1]AP (Senior Grade), CSE, VCET, Tamilnadu.

[2, 3, 4] BE, CSE, VCET, Tamilnadu.

Received date: 22nd May, 2018, Revised Date: 5th June, 2018, Accepted Date: 9th June, 2018.

Abstract - This paper mainly addresses the building of face recognition and detection system by using Principal Component Analysis (PCA). PCA is a statistical approach used for reducing the number of variables in face recognition. In PCA, every image in the training set is represented as a linear combination of weighted eigenvectors called Eigen faces. These eigenvectors are obtained from covariance matrix of a training image set. The weights are found out after selecting a set of most relevant Eigen faces. Recognition is performed by projecting a test image onto the subspace spanned by the Eigen faces and then classification is done by measuring minimum Euclidean distance. A number of experiments were done to evaluate the performance of the face recognition system. Detection is performed by comparing the Eigen faces of captured image with the train set image in the data set.

Keywords - Eigen value, Eigen face, PCA, Covariance Matrix,

I. INTRODUCTION

Video-based face identification has broad applications, such as automatic indexing of a video, shot retrieval of a character in a TV-series, and suspect identification in surveillance videos. Unlike still images, a subject in a video generates diverse exemplars that contribute to creating a robust representation. Videos of these applications usually consist of multiple shots that involve scene and view changes. Nevertheless, most of the current video identification techniques focus on the identification task where the videos are of a single shot and the frame-by-frame face bounding boxes of the target (i.e., person of interest) are either readily provided or automatically associated using a tracking algorithm. For instance, the YouTube Faces dataset provides frame by-frame annotations, which have been used as bench marks for evaluating video-based face identification algorithms .Although bounding boxes can be automatically extracted with face detection and tracking, human supervision is often needed to ensure that the annotations are not corrupted by the failure of face detection and tracking steps. Besides, most video-based identification techniques [34]have evaluated their performances on video face datasets consisting of single-shot videos, including YouTube Faces dataset ,Point and Shoot Face Recognition Challenge(Pa SC) dataset, and Celebrity-1000 .Although tracking techniques can be used to associate the face images of a target in single-shot videos by utilizing spatial, temporal, and appearance affinity, they are not effective for associating the target's face images present in multiple shots of a video. Hence, a video-based face identification technique that utilizes face tracking in a single shot cannot fully exploit useful information contained in multiple- shot videos, such as news videos, sport broadcasts, and movie trailers.

The target in a news video of multiple shots taken in several venues. An intra-shot face association technique should establish the linkage between face images within a video shot, and an inter-shot face association method should retrieve relevant face images from a single target annotation across multiple shots. Hence, the problem of performing video-based face identification task, where the target is only annotated once in the multiple-shot video, needs further investigation. We propose a target face association (TFA) method to retrieve a set of representative face images in a video that have the same identity as the target. This set of associated face images is then utilized to generate a robust representation for face identification The TFA method leverages a linear Support Vector Machine (SVM) to obtain the associated face images in the video. This linear SVM is trained iteratively with positive and negative instances guided by the cannot-link constraints. Note that several prior works have utilized the cannot-link constraints to learn effective metrics and models. Initially, only face images corresponding to the target annotation are treated as the positive instances. The negative training instances are the target's cannot-link instances, i.e., face

images that appear with positive instances in the same frame. Face images that are classified as positive will undergo a pruning process to iteratively remove the least likely positive instance that violates the cannot-link constraint. Hence, the updated positive instances as well as their cannot-link negative instances can be used to update the linear SVM. If there is no negative training instance inferred by the target's cannot-link instances, we utilize an external face dataset as background negative instances. The idea of using a single instance against a large set of negative instances to learn the similarity function with respect to background subjects is an essential component in computing one-shot similarity ,Hence, we can learn a target-specific classifier by leveraging the background statistics in scenarios that do not have any within-video negative training instances.

II. EXISTING METHOD

In the video-based face association and identification task, we are given a single annotation of the target face in a probe video. The objective is to retrieve a set of representative face images of the target in the video, and then this set of face images is utilized to create a robust face representation of the target face for searching its corresponding subject in the gallery. In the probe video, the target face is indicated by the human-annotated face bounding box b0 in frame f0. There is m bounding boxes discovered by a face detector in a video. These face bounding boxes are denoted as b1, b2, b m, which are present in frames f1, f2. . . fm, respectively .The feature corresponding to the face image in bounding box bi is denoted as xi. We aim to learn a target-specific SVM that can be used to classify a set of face images for creating a robust face representation of the target. This approach utilizes the facial feature learned by DCNN, and thus the robust face representation of the target can be easily represented by the average of the facial features of the set of associated face images.

The following section provides the details of each component of the proposed method in detail.

A. Face Pre Association with Tracking

Since learning the target-specific SVM requires the target annotation as the initial positive training instance, a low quality target annotation, such as noisy, badly illuminated, and extreme pose face images, prevents the TFA from learning an effective SVM. A tracking technique is able to model the appearance and motion of a human head, and thus it allows us to capture subsequent face images of high quality for good initial representation. Hence, we employ an off-the self tracking technique to track the target face, and the face detection bounding boxes pre associated by tracking are utilized as the initial positive training set. In each frame, the face detection bounding box that has the highest intersection-over-union (IoU) ratio with the tracking bounding box is utilized as the pre associated face images of the target. As the tracking technique becomes vulnerable to severe occlusion and abrupt motion, we only incorporate those pre associated face detection bounding boxes in the first k frames.

Since tracking across the shot boundary can lead to unexpected pre association of face images from different subjects, we utilize a simple shot detection method by checking the absolute difference of pixel values between two consecutive frames. When the absolute difference is larger than a certain threshold, the pre association with tracking is terminated. several cases where face pre association with tracking improves the initial representation where the target annotation is corrupted due to extreme pose, noise, and occlusion. We use the set of pre associated face images as initial positive training instances to learn the linear SVM. The index set of positive training images is denoted as Sp = {0} ∪ T, where T consists of indices of those face images pre associated by tracking .Fig. 3: Pre associated face images using tracking. The first row shows the target annotation in videos, and the second row shows the pre associated face images extracted using tracking.

B. Target Face Association

We train a target-specific linear SVM from face images in the video to establish the intra/inter-shot face association of the target face. With the annotation of the target face and the pre associated face images, the index set of positive instances is initially represented as Sp= {0 } ∪T. The negative training instances can be automatically discovered by utilizing the fact that the presence of a subject is unique. We define the cannot-link relation between the Ith and j th face image as gi, j = -1, if ri, j ≤ γ, fi = fj , and I , j = {0, 1, . . . , m},0, otherwise where ri, j is the IoU ratio between bounding box bi and bj . Since the non-maximal suppression of the face detection response is not perfect, a face can be discovered by more than one bounding box. We set a tolerance threshold γ for the IoU ratio to avoid face images of similar bounding boxes in a frame being mistakenly enforced by the cannot-link constraints. Thus, gi, j = 1 indicates that the I th and j th face image are far apart and appear in the same frame, and both images should not be identified as the same subject. Hence, the index set of within-video negative instances is represented as Sn = ∪j∈ Sp{ i |gi, j = 1}.We introduce a background negative set {xi}m + li =m+1 of instances collected from an external face dataset to model the background subjects, and its corresponding index set is represented as Sb = {m + 1, m + 2, . . . , m + l}. The background negative set becomes essential when there is no within-video training

instance. We train a linear SVM with training data {(xi, yi) |i (Sp ∪ Sn ∪ Sb)}, where the data label yi is expressed as yi =-1, if i is less than Sp, −1, otherwise. We propose two models to learn the weight vector w of the linear

C. Model 1

The linear SVM is solved using the max margin framework
Min w12wTw+Cpi, Sp max 0, 1 – yi w T ¯xi] 2+ Cn _ i Sn
max [0, 1 – yi w T ¯xi]2+ Cb _i Sb max[0, 1 – yi w T ¯xi]2,
Where, Cp = C|Sp| + |Sn| + |Sb|2|Sp|,
Cn = C |Sp| + |Sn| + |Sb|+2(|Sn| + |Sb|)+|Sn| + |Sb|+2|Sn| |Sp| + |Sn| + |Sb|4|Sn| ,
and Cb = C |Sp| + |Sn| + |Sb|+2(|Sn| + |Sb|)+|Sn| + |Sb|2|Sb| =|Sp| + |Sn| + |Sb|4|Sb|

Account for the weights to compensate for the class imbalance, and C is the cost parameter in the linear SVM. The weights are inversely proportional to the number of instances in positive and negative training sets. Among the negative samples, the weights are designed to be inversely proportional to the number of instances in Sn and Sb such that the contributions of within-video negative instances and background negative instances are balanced. We normalize xi to unit norm, and then concatenate it with one to account for the bias. The normalized and augmented feature vector is represented as ¯xi = [x Ti/_xi_ 1] T.

D. Model 2

Unlike Model 1 where the background negative instances are always utilized for training, we propose Model 2 that only utilizes the background negative instances when there is no within-video negative instance. The weight vector w of the linear SVM is solved using the max-margin framework,

Note that is the indicator function. In this model, the negative training set is composed of within-video negative instances that appear with positive instances in a frame. If there is no within-video negative instance, we employ the background negative instances as negative training instances. Face images in the video that are classified as positive will be regarded as the associated face images of the target. In certain cases, the human-annotated instance x0 can be misclassified as negative due to noise and extreme pose of a face image. Hence, we enforce the index 0 to be included in A, and the index set of the associated face images is represented as A = {0} ∪ {i |w T -xi > 0, i = 1..., m}. The associated face images in a are assumed to have the same identity as the target.

E. Robust Representation

Since the associated face images are assumed to have the same identity as the target does, we can use the associated face images as positive training instances (Sp ← A). With the cannot-link constraints, we can update the index set of the negative training instances by Sn ← ∪j, Sp {i |gi, j = 1}. We can alternately update the associated face images in A and the weight vector w until the index set of the associated face images converges or the maximum number of iterations t max is attained. The detailed procedure of the TFA is described in Algorithm 1. With associated face images, we can express the robust face representation as x fa =1|A|_i Axi. (12)Note that the proposed method can handle the intra/inter shot association of face images and face tracks. Although xi represents a descriptor of a face image in this work, the proposed method can be easily extended to operate on track-level face descriptors and the cannot-link constraints among face tracks.

III. PROPOSED METHOD

In this paper, we present new video-based face identification and detection algorithm, where the clandestine user (i.e., unauthorized person) in the probe video is only annotated once with a face bounding box in a frame and the video may consist of multiple shots. Most video face identification techniques assume that the video is of single shot, and thus the bounding boxes of the target face can be extracted by tracking a face across the video frames. Nevertheless, such automatic annotation is vulnerable to the drifting of the face tracker, and the face tracking algorithm is inadequate to associate the face images of the target across multiple shots. In this paper, we propose a target face association (TFA) technique that retrieves a set of representative face images in a given video that are likely to have the same identity as the target face. These face images are then utilized to construct a robust face representation of the target face for searching the corresponding subject in the gallery. For face recognition we propose a Principal Component Analysis (PCA) algorithm is used to provide efficient result. Since two faces that appear in the same video frame cannot belong to the same person, such cannot-link constraints are utilized for learning a target-specific linear classifier for establishing the intra/inter-shot face association of the target. When the entry of clandestine user into confidential area will be captured and alerts through E-mail. JANUS CS3 dataset show that our method generates robust representations from target-annotated videos and demonstrates good performance for the task of video-based face identification and recognition problem.

A. Face Recognition Process

One of the simplest and most effective PCA approaches used in face recognition systems is the so-called eigenface approach. This approach transforms faces into a small set of essential characteristics, eigenfaces, which are the main components of the initial set of learning images (training set). Recognition is done by projecting a new image in the eigenface subspace, after which the person is classified by comparing its position in eigenface space with the position of known individuals . The advantage of this approach over other face recognition systems is in its simplicity, speed and insensitivity to small or gradual changes on the face. The problem is limited to files that can be used to recognize the face. Namely, the images must be vertical frontal views of human faces. The whole recognition process involves two steps: A. Initialization process B. Recognition process The Initialization process involves the following operations:

- Acquire the initial set of face images called as training set.
- Calculate the Eigen faces from the training set, keeping only the highest eigenvalues. These M images define the face space. As new faces are experienced, the eigenfaces can be updated or recalculated.
- Calculate distribution in this M-dimensional space for each known person by projecting his or her face images onto this face-space.

B. Eigen face Algorithm

Let a face image $\Gamma(x, y)$ be a two dimensional M by N array of intensity values. In this thesis, I used a set of image by 200 × 149 pixels. An image may also be considered as a vector of dimension M × N, so that a typical image of size 200 × 149 becomes a vector of dimension 29,800 or equivalently a point in a 29,800 dimensional space. Conversion of M × N image into MN ×1 vector

- Step1: prepare the training faces Obtain face images I1, I2, I3, I4 , ... IM (training faces). The face images must be centered and of the same size.
- Step 2: Prepare the data set Each face image Ii in the database is transformed into a vector and placed into a training set S. In My example M = 34. Each image is transformed into a vector of size MN × 1 and placed into the set. For simplicity, the face images are assumed to be of size N × N resulting in a point in dimensional space. An ensemble of images, then, maps to a collection of points in this huge space.
- Step 3: compute the average face vector The average face vector (Ψ) has to be calculated by using the following formula:
- Step 4: Subtract the average face vector The average face vector is subtracted from the original faces and the result stored in the variable
- Step 5: Calculate the covariance matrix
- Step 6: Calculate the eigenvectors and eigenvalues of the covariance matrix The covariance matrix C in step 5 has a dimensionality, so one would have eigenface and eigenvalues. For a 256 × 256 image that means that on must compute a 65, 536 × 65, 536 matrix and calculate 65,536 eigenfaces. Computationally, this is not very efficient as most of those eigenfaces are not useful for our task. In general, PCA is used to describe a large dimensional space with a relative small set of vectors.
- Step 6: consider the matrix (M × M matrix)

C. Schematic Diagram and Flowchart

Figures 1 and 2 show the schematic diagram and flowchart of the proposed method.

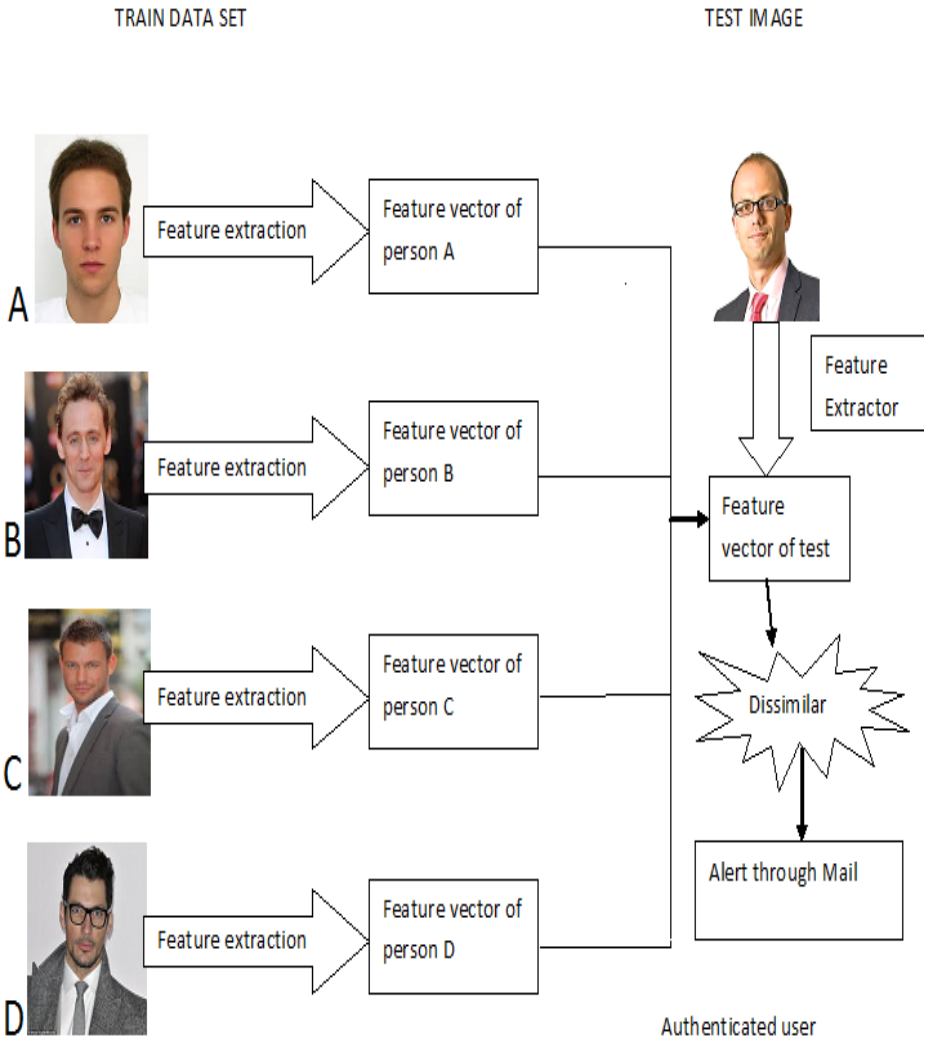

Fig. 1 Schematic diagram of the Proposed method

IV. CONCLUSION

In this paper, we present the TFA and PCA approach to assist the video-based face identification detection, task. With a single annotation of the target in the video, TFA can retrieve a set of representative face images in the video to create a robust representation of the target. Unlike tracking techniques that handle the association of face images in a video shot, the proposed method is capable of associating the face images across multiple shots in a video. The association is established by a target-specific linear classifier trained with face images of the target and background subjects in the video. The linear classifier is trained iteratively with the target's associated face images and the target's cannot-link face images. This target-specific linear classifier retrieves a set of face images to construct the robust representation of the target. Experimental results show that the target representation constructed by the associated face images is able to improve the performance of video-based face identification.

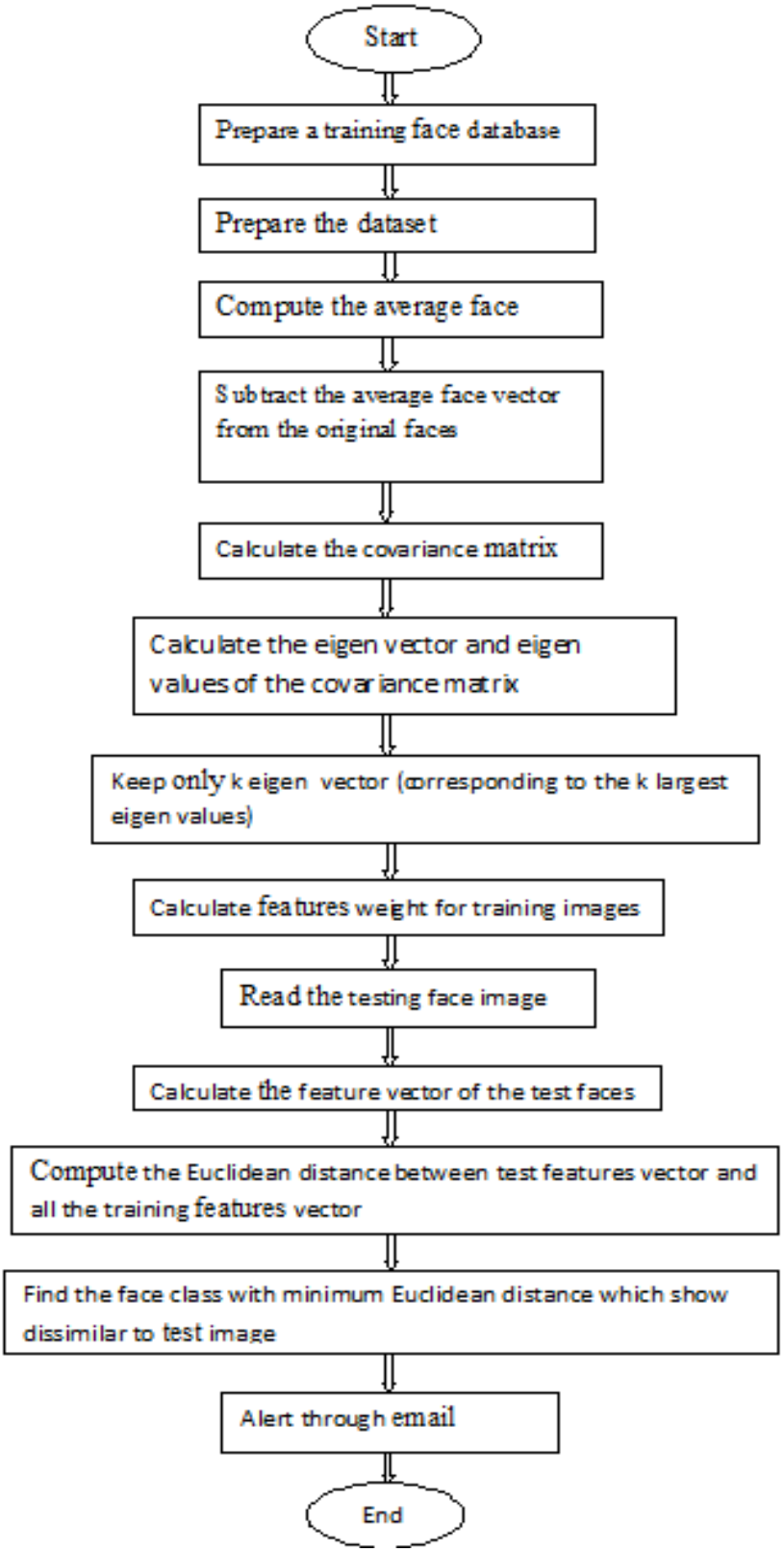

Fig. 2 Flow chart of the proposed method.

REFERENCES

[1] Ching-Hui Chen, Jun-Cheng Chen, Carlos D. Castillo, and Rama Chellappa, "Video-Based Face Association and Identification", 12th IEEE International Conference on Automatic Face & Gesture Recognition, DOI: 10.1109/FG.2017.27, 2017.

[2] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Joint face representation adaptation and clustering in videos", European Conference on Computer Vision (ECCV), 2016. Online: http://mmlab.ie.cuhk.edu.hk/projects/DeepFaceClustering/index.html

[3] J.-C. Chen, R. Ranjan, S. Sankaranarayanan, A. Kumar, C.-H. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa, "An end-to-end system for unconstrained face verification with deep convolution neural networks IEEE International Conference on Computer Vision Workshop (ICCVW), DOI: 10.1109/ICCVW.2015.55, 2015.

[4] M. B¨auml, M. Tapaswi, and R. Stiefelhagen, "Semi-supervised learning with constraints for person identification in multimedia data", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[5] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 12, pp. 2037-2041, 2011.

[6] L. Wolf, T. Hassner, and Y. Taigman, "The one-shot similarity kernel", IEEE International Conference on Computer Vision (ICCV), 2011, online: https://www.openu.ac.il/home/hassner /projects/Ossk/