Volume 4, Issue 11, November 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Face Recognition of Various Poses in a Video

Ms. Madhavi R. Bichwe, Ms. Ranjana Shende (Assistant Professor**)**
Department of Computer Science and Engineering
G. H. Raisoni Institute of Engineering And Technology for Women,
Nagpur, Maharashtra, India

Abstract*— *In recent years, multi-camera networks have become increasingly common for biometric and surveillance systems. Multi view face recognition has become an active research area in recent years. In this paper, an approach for video-based face recognition in camera networks is proposed. Traditional approaches estimate the pose of the face explicitly. A robust feature for multi-view recognition that is insensitive to pose variations is proposed in this project. The proposed feature is developed using the spherical harmonic representation of the face, texture mapped onto a sphere. The texture map for the whole face constructed by back-projecting the image intensity values from each of the views onto the surface of the spherical model. A particle filter is used to track the 3D location of the head using multi-view information. Videos provide an automatic and efficient way for feature extraction. Data redundancy renders the recognition algorithm more robust. The similarity between feature sets from different videos can be measured using the reproducing Kernel Hilbert space.

Keywords*— *multi camera networks , face recognition, spherical harmonic, particle filter, Kernel Hilbert space

I. INTRODUCTION

Face detection is the first stage of a face recognition system. A lot of research has been done in this area, most of which is efficient and effective for still images only & could not be applied to video sequences directly. In the video scenes, human faces can have unlimited orientations and positions, so its detection is of a variety of challenges to researchers [1][2]. In recent years, multi-camera networks have become increasingly common for biometric and surveillance systems. Multi view face recognition has become an active research area in recent years. In this paper, an approach for video-based face recognition in camera networks is proposed. Traditional approaches estimate the pose of the face explicitly. A robust feature for multi-view recognition that is insensitive to pose variations is proposed in this paper. The proposed feature is developed using the spherical harmonic representation of the face, texture mapped onto a sphere. The texture map for the whole face is constructed by back-projecting the image intensity values from each of the views onto the surface of the spherical model. A particle filter is used to track the 3D location of the head using multi-view information. Videos provide an automatic and efficient way for feature extraction. In particular, self-occlusion of facial features, as the pose varies, raises fundamental challenges to designing robust face recognition algorithms. A promising approach to handle pose variations and its inherent challenges is the use of multi-view data.

II. RELATED WORK

The term multi-view face recognition, in a strict sense, only refers to situations where multiple cameras acquire the subject (or scene) simultaneously and an algorithm collaboratively utilizes the acquired images/videos. But the term has frequently been used to recognize faces across pose variations. This ambiguity does not cause any problem for recognition with still images. A group of images simultaneously taken with multiple cameras and those taken with a single camera but at different view angles are equivalent as far as pose variations are concerned. However, in the case of video data, the two cases diverge. While a multi-camera system guarantees the acquisition of multi-view data at any moment, the chance of obtaining the equivalent data by using a single camera is unpredictable. Such differences become vital in non cooperative recognition applications such as surveillance. With the prevalence of camera networks, multi-view surveillance videos have become more and more common. Most existing multi-view video face recognition algorithms exploit single-view videos. The different methods for face recognition are given below:

A. Still image-based recognition:

This method will also require the poses and illumination conditions to be estimated for both face images. This "generic reference set" idea has also been used to develop the holistic matching algorithm, where the ranking of look-up results forms the basis of matching measure. There are also works which handles pose variations implicitly without estimating the pose explicitly [3].

B. Video-based recognition:

Video contains more information than still images. A straightforward way to handle single view videos is to take advantage of the data redundancy and perform view selection. Then, for each of the candidates, a face detector specific to that pose is applied to determine if it is a face. Only the frontal faces are retained for recognition. The continuity of pose

variation in video has inspired the idea of modelling face pose manifolds. The typical method is to cluster the frames of similar pose and train a linear subspace to represent each pose cluster. Here, the piecewise linear subspace model is an approximation to the pose manifold. The linearity is measured as the ratio of geodesic distance to Euclidean distance, and the distances are calculated between a candidate neighbour and each existing sample in the cluster. The 3D model can be then used in a model-based algorithm to perform face recognition [4].

C. Multi-view-based recognition:

In contrast to single view/video-based face recognition, there are relatively a smaller number of approaches for recognition using multi view videos. Frames of a multi-view sequence are collected together to form a gallery or probe set. The recognition algorithm is frame- based PCA and LDA fused by the sum rule. In, a three-layer hierarchical image-set matching technique is presented. The first layer associates frames of the same individual taken by the same camera. The second layer matches the groups obtained in the first layer among different cameras. Finally, the third layer compares the output of the second layer with the training set, which is manually clustered using multi-view videos. Though multi- view data is used to deal with occlusions when more than one subject is present, pose variations are not effectively addressed in this work [5].

D. Video processing in multi-camera networks:

Camera networks have been extensively used for surveillance and security applications. Research in this field has been focused on distributed tracking, resource allocation, activity recognition and active sensing. They adapt the feature correspondence computations by modelling the long term dependencies between them and then obtain statistically optimal paths for each subject [6].

E. Spherical harmonics (SH) in machine vision:

To estimate the SH basis images for a face at a fixed pose from a single 2D image based on statistical learning. When the 3D shape of the face is available, the SH basis images can be estimated for test images with different poses [7]. As a result, they require a 3D face model and face pose estimation to infer the face appearance. An SH-based feature to directly model face appearance rather than the reflectance function is used, and hence do not require a 3D face surface model or a pose estimation step.

III. PROPOSED WORK

For a given set of multi-view video sequences, first use a particle filter to track the 3D location of the head using multi-view information. At each time instant or video frame, build the texture map associated with the face under the spherical model for the face. Given that the 3D location of the head from the tracking algorithm, back-project the image intensity values from each of the views onto the surface of the spherical model, and construct a texture map for the whole face. Then compute a Spherical Harmonic (SH) transform of the texture map, and construct a robust feature that is based on the properties of the SH projection.

For recognition with videos, the feature similarity is measured by the limiting Bhattacharyya distance of features in the Reproducing Kernel Hilbert Space.

The proposed approach outperforms traditional features and algorithms on a multi-view video database collected using a camera network. Building rotational tolerances into this feature completely bypasses the pose estimation step.

The proposed approach of the Multi-view Face Recognition Algorithm is defined as follows.

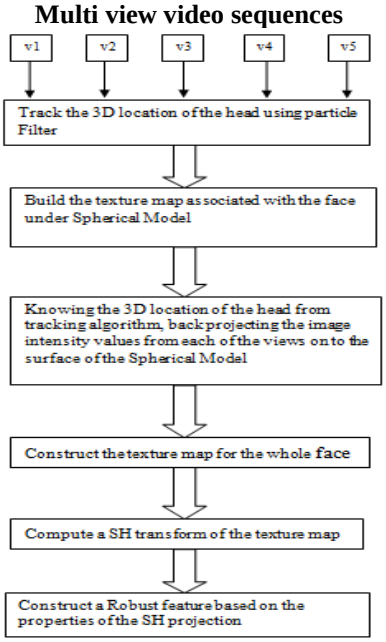

Fig 1: Flow diagram of Multi-view Face Recognition Algorithm

Robust feature:

The robust feature is based on the theory of spherical harmonics. Spherical harmonics are a set of orthonormal basis functions defined over the unit sphere, and can be used to linearly expand any square- integrable function on S^2 as:

$$f(\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l} f_{lm}\,Y_{lm}(\theta,\phi).$$

Where $Y_{lm}(\cdot,\cdot)$ defines the SH basis function of degree $l \geq 0$ and order $m \in (-l,-l+1,\ldots,l-1,l)$. f_{lm} is the coefficient associated with the basis function Y_{lm} for the function f .The SH basis function for degree l and order m has the following form:

$$Y_{lm}(\theta,\phi)=K_{lm}\,P_l^m(\cos\theta)e^{im\phi}$$

where K_{lm} denotes a normalization constant such that:

$$\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} Y_{lm}Y_{lm}^{*}\,d\phi d\theta = 1$$

Here, $P^m{}_l\,(x)$ is the associated Legendre functions.

In this paper, we are interested in modelling real-valued functions (eg. texture maps) and thus, we are more interested in the real Spherical Harmonics which are defined as:

$$Y_l^m(\theta,\phi)=\begin{cases} Y_{l0} & \text{if } m=0\\ \frac{1}{\sqrt{2}}(Y_{lm}+(-1)^m\,Y_{l,-m}) & \text{if } m>0\\ \frac{1}{\sqrt{2}i}(Y_{l,-m}-(-1)^m\,Y_{lm}) & \text{if } m<0\end{cases}$$

The real SHs are also orthonormal and they share most of the important properties of the general Spherical Harmonics. We visualize the SH for degree l = 0, 1, 2.

As with Fourier expansion, the SH expansion coefficients $f^m{}_l$ can be computed as:

$$f_l^m=\int_{\theta}\int_{\phi} f(\theta,\phi)Y_l^m(\theta,\phi)\,d\theta d\phi$$

The expansion coefficients have a very important property which is directly related to our "pose free" face recognition application.

A robust multi-view tracking algorithm based on Sequential Importance Resampling (SIR) (particle filtering). Tracking is an essential stage in camera-network-based video processing. It automates the localization of the face and has direct impact on the performance of the recognition algorithm.

Multi-View Tracking: It is well known that higher the dimensionality of the state space is the harder the tracking problem becomes. This is especially true for search-algorithms like SIR since the number of particles typically grows dramatically for high-dimensional state spaces. However, given that our eventual recognition framework is built on the robust feature derived using SH representation under the diffuse lighting assumption, it suffices that we track only the location of the head in 3D. Hence, the state space for tracking **s** = *(x, y, z)* represents only the position of a sphere's centre, disregarding any orientation information [8].

Histogram: A normalized 3D histogram in RGB space is built from this image region. Its difference with the template, which is set up at the first frame through the same procedure and subject to adaptive update thereafter, is measured by the Bhattacharyya distance. This defines the first cue matching function.

Gradient map:

The magnitude of the image gradient response and its direction to be perpendicular to the tangent directions, [9]. Consequently, the second cue matching score formulated as:

$$\varphi(O_t,\mathbf{s}_t^i)=\frac{1}{r_j^i}\sum_{m=1}^{M}|\mathbf{n}_m\cdot\nabla\mathbf{I}_m|,$$

Where $r^i{}_j$ is the radius of $E^i{}_j$ measured in number of pixels ,$\mathbf{n}_m$ is the normal vector of the *m*-th pixel on the arc, and **I***m* is the image gradient at this pixel.

Texture Mapping: Once, the texture map of the head center is obtained. First, the sphere's surface is sampled according to the following procedure:

1) Uniformly sample within the range $[-R, R]$, where R is the radius of the sphere, to get z_n , $n = 1, 2, .\ N$.

2) Uniformly sample α_n within the range $[0, 2\pi]$, and independent of z_n.

3) $\mathrm{xn}=\sqrt{\mathrm{R}^2-\mathrm{Zn}^2}\cos\alpha\mathrm{n},\ \mathrm{yn}=\sqrt{\mathrm{R}^2-\mathrm{Zn}^2}\sin\alpha\mathrm{n}.$

Then, a coordinate transformation for these sample points is performed.

IV. CONCLUSIONS

A multi-view face recognition algorithm does not require any pose estimation or model registration step. A multi-view video tracking algorithm is presented to automate the feature acquisition in a camera network setting. The video-based recognition problem can be modelled as one of measuring ensemble similarities in Reproducing Kernel Hilbert Space (RKHS). The performance of this method can be demonstrated on a relatively uncontrolled multi-view video database.

REFERENCES

[1] Ming Du, "Robust face recognition from multi-view videos," in IEEE trans on image processing,Aswin C. Sankaranarayanan, Member,IEEE,and Rama Chellappa,Fellow ,vol.23,No 3,March 2014 .

[2] "Video-based Face Recognition: A Survey", Huafeng Wang, Yunhong Wang, And Yuan Cao World Academy of Science, Engineering and Technology Vol-3 2009-12-25.

[3] F. Schroff, T. Treibitz, D. Kriegman, and S. Belongie, "Pose, illumination and expression invariant pairwise face-similarity measure via Doppelgänger list comparison," in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 2494–2501.

[4] I. Kotsia, N. Nikolaidis, and I. Pitas, "Frontal view recognition in multiview video sequences," in Proc. Int. Conf. Multimedia Exposit., Jun. 2009, pp. 702–705.

[5] M. Nishiyama, M. Yuasa, T. Shibata, T. Wakasugi, T. Kawahara, and O. Yamaguchi, "Recognizing faces of moving people by hierarchical image-set matching," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2007, pp. 1–8.

[6] A. C. Sankaranarayanan, A. Veeraraghavan, and R. Chellappa, "Object detection, tracking and recognition for multiple smart cameras," Proc. IEEE, vol. 96, no. 10, pp. 1606–1624, Oct. 2008.

[7] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 2, pp. 218–233, Feb. 2003.

[8] K. C. Lee, J. Ho, M. H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., vol. 1. Jun. 2003, pp. 313–320.

[9] O. Arandjelovic and R. Cipolla, "A pose-wise linear illumination manifold model for face recognition using video," Comput. Vis. Image Understand, IEEE transaction, vol. 113, no. 1, pp. 113–125, 2009.

