Reliable Face Recognition for Intelligent CCTV

Ting Shan[‡†], Brian C. Lovell[‡†], Shaokang Chen[†] and Abbas Bigdeli[‡]

[‡]*National ICT Australia (NICTA)*
and
[†]*Intelligent Real-Time Imaging and Sensing Group*
EMI, School of ITEE, The University of Queensland

ABSTRACT: Face Recognition is generating enormous interest due to government concerns about identity management and global terrorist activity. One ambition of Intelligent CCTV is to help prevent terrorism and a key technology is reliable face recognition. With passport quality photographs, current face recognition technologies can approach 95% recognition accuracy. Yet trials show that performance drops to only 10 to 20% when there are significant changes in lighting, pose, and facial expression. We describe research by the authors to address these issues and provide reliable face recognition performance in real-time. Our system has three major components comprising: 1) a Viola-Jones face detection module based on cascaded simple binary features to rapidly detect and locate multiple faces from the input still image or video sequences, 2) a Pose Normalization Module to estimate facial pose and compensate for extreme pose angles, and 3) Adaptive Principal Component Analysis to recognize the normalized faces. Experimental results show that our approach can achieve good recognition rates on face images across a wide range of head poses with different lighting and expressions.

Introduction

The Prime Minister of Australia, John Howard, returned from London after the July 2005 suicide bombings and said (Howard J 2005), "The biggest thing that I have learnt by a country mile out of my visit, particularly to Britain, is the extraordinary value of surveillance cameras." The integrated security camera system deployed in Britain was primarily installed to reduce the incidence of assault and property damage. Yet it was able to help identify and discover the movements of the four suicide bombers within just 24 hours of the bombing. The effectiveness of this particular security system comes about because of the saturation coverage of the British people by CCTV. In Britain, CCTV systems cover cities, public transport and motorways, while in many other countries the coverage is quite haphazard. It was public demand for security in public places that led to this pervasiveness. Moreover, the adoption of centralised digital video databases, largely to reduce management and monitoring costs, has also resulted in an extraordinary co-ordination of the CCTV resources.

An emerging testbed for intelligent CCTV is the emerging experimental planetary scale sensor web, IrisNet (Gibbons P B, Karp B, Ke Y, Nath S and Sehan S 2003). IrisNet uses internet connected desktop PCs and inexpensive, off-the-shelf sensors such as webcams, microphones, temperature, and motion sensors deployed globally to provide a wide-area sensor network. IrisNet is deployed as a service on PlanetLab (www.planet-lab.org), a worldwide collaborative network environment for prototyping next generation internet services initiated by Intel Research and Princeton University. Gibbons *et al.* (Gibbons P B, Karp B, Ke Y, Nath S and Sehan S 2003)

envisage a worldwide sensor web in which many users can query, as a single unit, vast quantities of data from thousands or even millions of planetary sensors.

One ambition of Intelligent CCTV is to help prevent terrorism rather than just recording the events leading up to an attack and a key technology is reliable face recognition. We have just begun a major project to field trial these technologies in transport centres accessed by large numbers of people on a daily basis. Now we will focus on some of the crucial technologies underpinning such intelligent CCTV services — automatically detecting, normalizing and recognizing faces in image and video databases.

Face Detection

Face detection is a necessary first-step in a face recognition system to locate a face or faces from cluttered backgrounds. Many techniques for face detection have been developed including: feature-based approaches (Govindaraju V 1996, McKenna S, Gong S and Liddell H 1995), template matching (Jeng S H, Liao H Y M, Liu Y T and Chen M Y 1998, Gunn S R and Nixon M S 1994) and image-based approaches (Turk M and Pentland A 1991, Sung K K and Poggio T 1998). In 2001, Viola and Jones (Viola P and Jones M 2001) proposed an image-based face detection system which can achieve remarkably good performance in real-time. Our face detection module is based on the Viola-Jones approach using our own training sets.

The cascade face detector uses a sequence of binary classifiers which discard non-face regions and only send likely face candidates to the next level of classifier. Simple classifiers can be constructed which reject the majority of non-face regions at the very early stage of detection, before the use of more complex classifiers with higher discriminative capability. In this way, the more discriminating but more complex classifiers concentrate their processing time on face-like regions.

Haar-like wavelet features (Mallat S G 1989) extracted from the image sub-windows are an image representation method which characterises the texture similarities between different regions by computing the sum of pixel values in different regions. It can be computed rapidly from the integral image (Crow F 1984) which maintains a running sum of pixel values in two dimensions.

The Adaboost learning algorithm selects the best rectangle features and linearly combines these features into a classifier. Adaboost is a boosting learning algorithm, which can fuse many weak classifiers into a single more precise classifier. The main idea of Adaboost is as follows. At the beginning of training, all training examples are assigned equal weight. During the process of boosting, the weak classifier with the lowest classification error is selected and the weights of the samples which are wrongly classified by the weak classifier increase. The final classifier is a linear combination of the weak classifiers from all rounds, where classifiers with lower classification error have a higher weighting. Figure1 shows examples of face detection using our implementation.

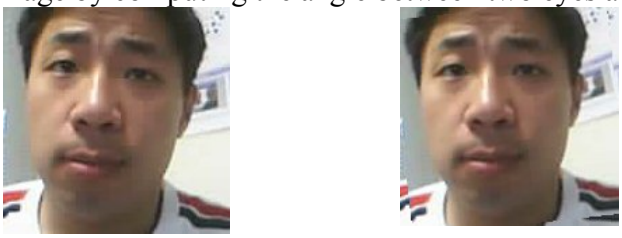
Fig.1: Detection results on several photographs.

Pose Normalization

Pose normalization refers to compensating for the pitch, roll, and yaw motions of the head to allow for non-frontal viewing conditions. It acts as a bridge between the face detection and face recognition modules.

In-Plane Rotation

A Viola-Jones eye detector is used to locate the eyes, and the face image is rotated to a vertical frontal face image by computing the angle between two eyes and the horizontal baseline (Fig. 2).



Before After

Fig. 2: A face sample before and after in-plane image rotation

Out-of-Plane Rotation

Active Appearance Models are a powerful tool to describe deformable object images. Given a collection of training images for a certain object class where the feature points have been manually marked, a shape and texture can be represented by applying PCA to the sample as:

$$x = \bar{x} + P_s c$$
$$g = \bar{g} + P_g c$$

where $\bar{x}$ is the mean shape, $\bar{g}$ is the mean texture and P_s, P_g are matrices describing the shape and texture variations. The parameter ***c*** is a vector of model parameters controlling both shape and texture of the model.

We assume that this model parameters ***c*** is related to the viewing angle, θ, approximately by:

$$c = c_0 + c_c \cos(\theta) + c_s \sin(\theta)$$

where c_0, c_c and c_s are vectors which are learned from the training data.

Given a new face image with model parameters ***c***, we can estimate orientation and then synthesize new views. Let c_{res} be the residual vector not explained by the rotation model,

$$c_{res} = c - (c_0 + c_s \cos(\theta) + c_s \sin(\theta))$$

To calculate the model parameters $c(\alpha)$ at a new angle, α, we simply use the equation:

$$c(\alpha) = c_0 + c_c \cos(\alpha) + c_s \sin(\alpha) + c_{res}$$

Here α is 0, so the equation will be:

$$c(0) = c_0 + c_c + c_{res}$$

Then frontal face images can be synthesized (Fig. 3).

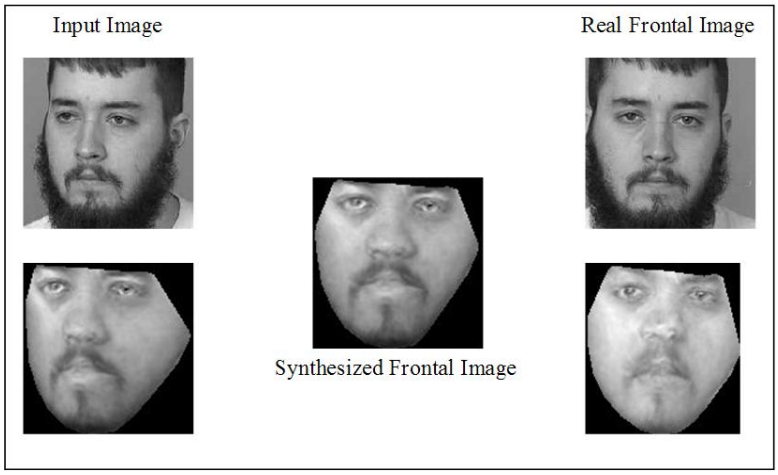


Fig.3: Some synthesized frontal face images from the Feret Face Database.

Face Recognition

Recent research on face recognition has been focused on diminishing the impact of nuisance factors on face recognition. Many approaches have been proposed for illumination invariant recognition (Yilmaz A and Gokmen M 2000, Gao Y S and Leung M K H 2002) and expression invariant recognition (Beymer D and Poggio T 1995, Black M J, Fleet D J and Yacoob Y 2000).

But these methods suffer from the need to have large numbers of example images for training, Chen and Lovell (Chen S K and Lovell B 2004) developed Adaptive Principal Component Analysis (APCA) to compensate for illumination and facial expression variations. It inherits merits from both PCA (eigenfaces) and FLD (Fisher Linear Discriminant or fisherfaces) by warping the face subspace according to the within- and between-class covariance. The method which we may call "eigen-fisherfaces" consists of four steps:

- Subspace Projection: Apply PCA to project face images into the face subspace to generate the m-dimensional feature vectors
- Whitening Transformation: The subspace is whitened according to the eigenvalues of the subspace with a whitening power p

$$\text{cov} = diag\{\lambda_1^{-2p}, \lambda_2^{-2p}, \ldots \lambda_m^{-2p}\} \quad (1)$$

- Filtering the Eigenfaces: Eigen-features are weighted according to the identification-to-variation value ITV with a filtering power q.

$$\gamma = diag\{ITV_1^q, ITV_2^q, \ldots ITV_m^q\} \quad (2)$$

- Optimizing the cost function: Minimize the cost function according to the combination of error rate and the ratio of between-class distance and within-class distance:.

$$OPT = \sum_{j=1}^{M} \sum_{k=1}^{K} \sum_{m} \frac{d_{jj,k_0}}{d_{jm,k_0}}, \forall m \in d_{jm,k_0} < d_{jj,k_0}, m \in [1 \cdots m] \quad (3)$$

More details about APCA can be found in [4]. Experiments show this technique performs well under changes in lighting conditions and facial expression (Fig. 4).

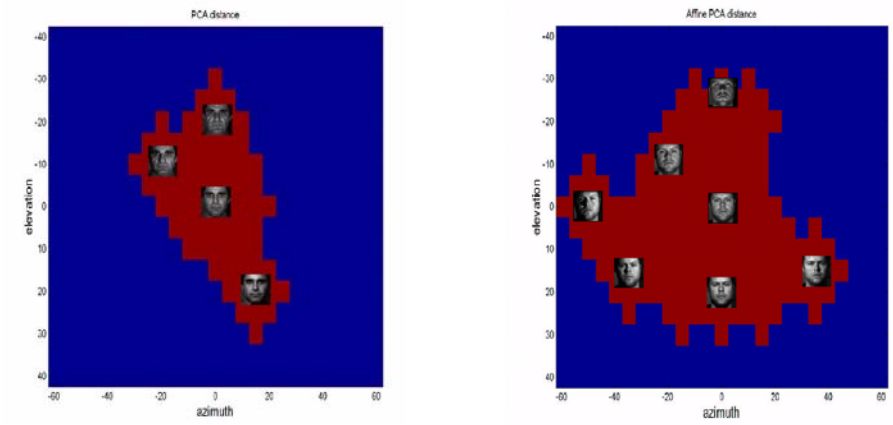


Fig.4: Contours of 95% recognition performance for the original PCA and the proposed APCA method against lighting elevation and azimuth.

Tracking

A recognition confidence output is calculated from the angle θ which measures how closely the input face image subspace matches the recognized face subspace. Only the face sub-window with high confidence will be registered for tracking (Fig.5).

Fig. 5: Some selected frames (from left to right, up to down) showing different recognition confidence levels using different colour rectangles.

Conclusion

In this paper we describe some technologies underpinning the pattern recognition engine of a system for locating persons using Intelligent CCTV. A fully automated system must be highly robust to nuisance problems such as lighting, expression change, pose, and camera variation. Our system has three components: 1) a Viola-Jones face detection module based on cascaded simple binary features to rapidly detect and locate multiple faces from the input still image or video sequences, 2) view-based Active Appearance Models to estimate facial pose and compensate for extreme pose angles, and 3) Adaptive Principal Component Analysis, or eigen-fisher faces, to recognize the faces since the method is robust to poor lighting and extreme facial expressions. We plan to report on the results of the public field trials at a later date.

Acknowlegement

This project is supported by a grant from the Australian Government Department of the Prime Minister and Cabinet. National ICT Australia is funded by the Australian Government's Backing Australia's Ability initiative and the Queensland Government, in part through the Australian Research Council.

References

Beymer D and Poggio T (1995): Face Recognition from One Example View. Proc. Int'l Conf. of Comp. Vision, 500-507

Black M J, Fleet D J and Yacoob Y (2000): Robustly estimating Changes in Image Appearance. In: Computer Vision and Image Understanding, 78(1), 8-31.

Chen S K and Lovell B (2004): Illumination and Expression Invariant Face Recognition with One Sample Image per Class. In: 17th International Conference on Pattern Recognition (ICPR' 04) – Volume 1 pp. 300-303

Cootes T F, Walker L and Taylor C J (2000): View-Based Active Appearance Models. 4th International Conference on Automatic Face and Gesture Recognition, pp: 227-232, March.

Crow F (1984): Summed-are tables for texture mapping. Proc of SIGGRAPH, Volume 18(3), pages 207-212.

Feret (2005): http://www.itl.nist.gov/iad/humanid/feret/ [last visited 23-Nov-2005]

Gao Y S and Leung M K H (2002): Face Recognition Using Line Edge Map. In IEEE PAMI. 24(6), June, 764-779

Gibbons P B, Karp B, Ke Y, Nath S and Sehan S (2003), IrisNet: An Architecture for a Worldwide Sensor Web. In: Pervasive Computing, 2(4), 22-23, Oct – Dec

Govindaraju V (1996) Locating human faces in photographs. In: International Journal of Computer Vision, Volume 19, Issue 2, August, pp: 129-146.

Gunn S R and Nixon M S (1994): A dual active contour for head boundary extraction. In: IEE Colloquium on Image Processing for Biometric Measurement, pp: 6/1 – 6/4, 20 Apr.

Horward J (2005): http://smh.com.au/news/national/howard-backs-more-security-cameras/2005/07/24/1122143730105.html [last visited 23-Nov-2005]

Jeng S H, Liao H Y M, Liu Y T and Chen M Y (1998): An efficient approach for facial feature detection using geometrical face model. Proc: 13th International Conference on Pattern Recognition, Volume 3, pp: 426-430, 25-29 Aug

Mallat S G (1989): A theory for multi-resolution signal decomposition: The wavelet representation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7): 674-693.

McKenna S, Gong S and Liddell H (1995) Real-time tracking for an integrated face recognition system. In 2nd Workshop on Parallel Modelling of Neural Operators, Faro, Portugal, Nov

Sung K K and Poggio T (1998): Example-based learning for view-based human face detection. ITEE Transaction on Pattern Analysis and Machine Intelligence.

Turk M and Pentland A (1991): Face recognition using eigenfaces. Proc. Computer Vision and Pattern Recognition, pp: 586-591.

Yilmaz A and Gokmen M (2000): Eigenhill vs. eigenface and eigenedge. In Procs of International Conference Pattern Recognition, Barcelona, Spain, 827-830

BIOGRAPHY:

Ting Shan: PhD student of University of Queensland, his research interest is in face detection, multi-view face recognition and real-time face recognition system.

Brian Lovell: Research Leader of the SAFE Sensors group in National ICT Australia and Professor in the School of ITEE, The University of Queensland. His main interest is the analysis of video streams to recognize human activities.

Shaokang Chen:

Abbas Bigdeli: