On Intelligent Surveillance Systems and Face Recognition for Mass Transport Security

Brian C. Lovell *† Shaokang Chen *†, Abbas Bigdeli *†, Erik Berglund *†, Conrad Sanderson *†,
* NICTA, PO Box 6020, St Lucia, QLD 4067, Australia
† School of ITEE, The University of Queensland, Brisbane, QLD 4072, Australia

***Abstract*—We describe a project to trial and develop enhanced surveillance technologies for public safety. A key technology is robust recognition of faces from low-resolution CCTV footage where there may be as few as 12 pixels between the eyes. Current commercial face recognition systems require 60-90 pixels between the eyes as well as tightly controlled image capture conditions. Our group has thus concentrated on fundamental face recognition issues such as robustness to low resolution and image capture conditions as required for uncontrolled CCTV surveillance. In this paper, we propose a fast multi-class pattern classification approach to enhance PCA and FLD methods for 2D face recognition under changes in pose, illumination, and expression. The method first finds the optimal weights of features pairwise and constructs a feature chain in order to determine the weights for all features. Computational load of the proposed approach is extremely low by design, in order to facilitate usage in automated surveillance. The method is evaluated on PIE, FERET, and Asian Face databases, with the results showing that the method performs remarkably well compared to several benchmark appearance-based methods. Moreover, the method can reliably recognise faces with large pose angles from just one gallery image.**

***Index Terms*—surveillance, mass transport, CCTV, face recognition, security**

I. Introduction

For isolated crimes such as assault and robbery, it is well-known that video surveillance is highly effective in helping to find and successfully prosecute the perpetrators. Moreover, electronic surveillance has been shown to act as a significant deterrent to crime. Cost is mitigated by recording most of the camera feeds without any human monitoring — if an event is reported to security, the relevant video is manually extracted and reviewed.

However, in recent times the game has changed due to the human and political cost of successful terrorist attacks on soft targets such as mass transport systems. Traditional forensic analysis of recorded video after the event is simply not an adequate response from government and large business. This seachange in the security sector is due to the fact that in the case of suicide attacks there is simply no possibility of prosecution after the event, so simply recording surveillance video provides no terrorism deterrent. Video of successful attacks may indeed add impact to the political message of the perpetrators by highlighting the failure of Western governments to protect their populace. A pressing need is emerging to detect events and persons of interest using video surveillance before such harmful actions can occur. This means that cameras must be monitored at all times.

The problem is that human monitoring of surveillance systems requires a large number of personnel, resulting in high ongoing costs and questionable reliability due to the attention span of humans decreasing rapidly when performing such tedious tasks. A solution may be found in advanced surveillance systems employing computer monitoring of all video feeds, delivering the alerts to human responders for triage. Indeed such systems may assist in maintaining the high level of vigilance required over many years to detect the rare events associated with terrorism — a well-designed computer system is never caught off-guard.

In 2006, NICTA was awarded a research grant to conduct long term trials of Intelligent CCTV (ICCTV) technologies in important and sensitive public spaces such as major ports and railway stations [1]. One such advanced technology is a system that projects all the CCTV video feeds on to a 3D model of the environment providing rapid situational assessment facilitating a rapid response to situations arising as shown in Figure 1. The trial will highlight operational and capability deficiencies in current ICCTV systems and will focus NICTA's research on capability gaps. The project is thus a vertically integrated collaboration of researchers, vendors, and user agencies aimed at delivering advances in computer vision and pattern recognition for human activity recognition.

The potential of intelligent security systems is huge and this fact is just starting to be recognised by the industry.

> I can see in the next 20 years everything will become automated. Once the camera is sophisticated enough, it will profile people that we don't really need human beings apart from to check it out and analyse it

— Angus Hamilton, Director, Corporate Security, Shangri-La Hotels and Resorts, former assistant commissioner of Hong Kong Police [2].

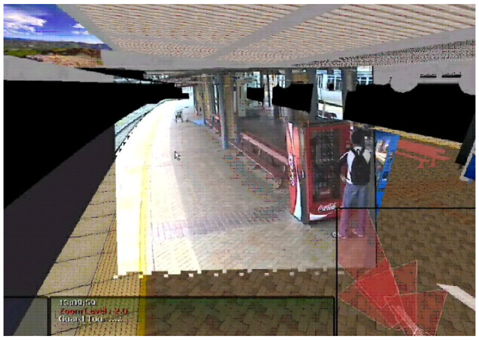

Fig. 1. Immersive 3D Visual Presentation of Camera View and 3D model of the railway platform.

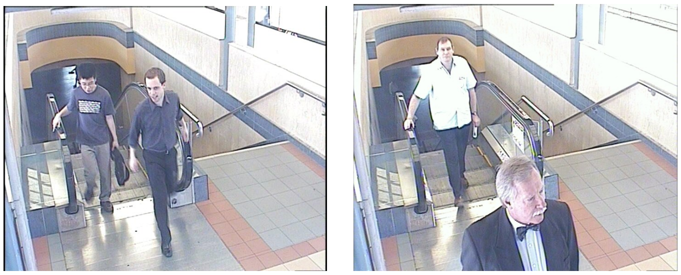

Fig. 2. Examples of typical face pose under surveillance conditions.

One of the “test-beds” we are using for our advanced surveillance field trials is a railway station in Brisbane (Australia), which provides us with implementation and installation issues that can be expected to arise in similar mass-transport facilities. Capturing the camera feeds in a real-world situation can be problematic, as there must be no disruption in operational capability of existing security systems. The optimal approach would be to simply use IP camera feeds. However, in many existing surveillance systems the cameras are analog and often their streams are fed to relatively old analog or digital recording equipment. Limitations of such systems may include low resolution, recording only a few frames per second, non-uniform time delay between frames, and proprietary codecs. To avoid disruption while at the same time obtaining video streams which are more suitable for an intelligent surveillance system, it is useful to tap directly into the analog video feeds and process them via dedicated analog-to-digital video matrix switches.

A key technology being developed within our group for prevention of crime and terrorism is the reliable detection of “persons of interest” through face recognition. While automatic face recognition of cooperative subjects has achieved good results in controlled applications such as passport control, CCTV conditions are considerably more challenging. Examples of real life CCTV conditions captured at the railway station are shown in Figure 2.

Nuisance factors such as varying pose, illumination, and expression (PIE) can greatly affect recognition performance. According to Phillips *et al*. head pose is believed to be the hardest factor to model [3]. In mass transport systems, surveillance cameras are often mounted in the ceiling in places such as railway platforms and passenger trains. Since the subjects are generally not posing for the camera, it is rare to obtain a true frontal face image. As it is infeasible to consider remounting all the cameras (in our case more than 6000) to improve face recognition performance, any practical recognition system must have highly effective pose compensation.

A further complication is that in many practical situations there is generally only have one frontal gallery image of each person of interest (*e.g*. a passport photograph or a mugshot). In addition to robustness and accuracy, scalability and fast performance are of prime importance for surveillance. A face recognition system should be able to handle large volumes of people (*e.g*. peak hour at a railway station), possibly processing hundreds of video streams. While it is possible to setup elaborate parallel computation machines, there are always cost considerations limiting the number of CPUs available for processing. In this context, a face recognition algorithm should be able to run in real-time or better, which necessarily limits complexity.

We note that while true 3D based approaches in theory allow face matching at various poses, current 3D sensing hardware has too many limitations [4] including cost and range. Moreover unlike 2D recognition, 3D technology cannot be retrofitted to existing surveillance systems.

Certainly 2D recognition presents much greater technical challenges due to difficulties presented by illumination and shadow effects as was famously noted by the great Leonardo da Vinci (1452-1519):

> After painting comes Sculpture, a very noble art, but one that does not in the execution require the same supreme ingenuity as the art of painting, since in two most important and difficult particulars, in foreshortening and in light and shade, for which the painter has to invent a process, sculpture is helped by nature.

We continue the paper as follows. An overview of previous work on robust face recognition is given in Section II. We propose a new robust face recognition method, dubbed *Chained Weighted Feature Pairs*, in Section III. An empirical evaluation of our method on three public face databases is given in Section IV. We draw our conclusions and describe future directions for the project in Section V.

II. Previous Approaches

For dealing with illumination variation, two main approaches have been proposed. One is to represent images with features that are less sensitive to illumination change [5], [6] such as the edge maps of the image. Another approach is to construct a low dimensional linear subspace for images of faces taken under different lighting conditions [7], [8]. The former approach suffers from the fact that features generated from shadows are related to illumination change and may have an impact on recognition, while the latter is based on an assumption that images of a convex Lambertian object under variable illumination form a convex cone in the space of all possible images [8]. Note that around 3 to 9 gallery images are needed to construct the convex cone. However, it is hard for these methods to deal with cast shadows due to the fact that the surface of human faces is not truly Lambertian reflected nor convex.

To deal with expression changes, Black *et al*. [9] suggested that images be morphed to be the same expression as the one used for training. But not all images can be morphed correctly, for example an image with closed eyes cannot be morphed to a neutral image because of the lack of texture inside the eyes. Liu *et al*. [10] proposed using optical flow for face recognition with expression variations. However, it is hard to learn the local motions within the feature space to determine the expression changes of each face, since the way one person express a certain emotion is normally somewhat different from another. Martinez proposed a weighting scheme to deal with facial expressions in [11]. An image is divided into several local areas and those that are less sensitive to expression change are chosen and weighted accordingly.

Pose variability is usually considered to be the most challenging problem. There are three main approaches developed for 2D based pose invariant face recognition. Wiskott *et al.* proposed Elastic Bunch Graph Matching [12], while Sankaran and Asari [13] proposed multiple-view templates to represent faces with different poses. Multiple view approaches require several gallery images per person under controlled viewing conditions to identify the face, which prevents its application when only one gallery image per person is available. Face synthesis methods have emerged in an attempt to overcome this issue. In [14], Gao *et al.* constructed a Face-Specific Subspace by synthesising novel views from a single image. In [15] a method for direct synthesis of face model parameters is proposed. In [16], an Active Appearance Model (AAM) based face synthesis method is applied for face recognition subject to relatively small pose variations. A recurring problem with AAM based synthesis and multi-view methods is the need to reliably locate facial features to determine the pose angle for pose compensation — this turns out to be difficult task in its own right.

The above methods can handle certain kinds of face image variation successfully, but drawbacks still restrict their application. It may be risky to rely heavily on choosing invariant features [5], [11], [12], [6], such as using edge maps of the image or choosing expression insensitive regions. This is because features insensitive to one variation may be highly sensitive to other variations and it is very difficult to abstract features that are completely immune to all kinds of variation [17]. Some approaches attempt to construct face specific models to describe possible variations under changes in lighting or pose [7], [14], [8]. Such methods require multiple images per person taken under controlled conditions to construct a specific subspace for each person for the face representation. This leads to expensive image capture processes, poor scalability of the face model, and does not permit applications where only one gallery image is available per person.

Other approaches divide the range of variation into several subranges (e.g., low, medium, and high pose angles) and construct multiple face spaces to describe face variations lying in the corresponding subrange [13]. These approaches require us to register several images representing different variations per person into the corresponding variation models so that matching can be done in each interval individually. Once again, acquiring multiple images per person under specific conditions is often very difficult, if not impossible, in practice.

III. Chained Weighted Feature Pairs

We propose an appearance based approach for reliable face recognition under pose, illumination, and expression changes. We develop a learning method for finding the optimal weights within feature pairs, which are then placed in a chain in order to obtain the weights for all features. From a classifier combination point of view, a classifier using each feature pair can be considered as a base classifier and the feature chain is equivalent to the combined classifier. We call this approach Chained Weighted Feature Pairs (CWFP). The technique is used to enhance both Principal Component Analysis (PCA) and Fisher's Linear Discriminant (FLD) based techniques (benchmark methods), yielding 'PCA+CWFP' and 'FLD+CWFP' methods.

It must be noted that compared to other recent approaches for dealing with pose variations (e.g. [18], [19], [20]) we have deliberately developed a low-complexity technique in order to facilitate usage in real-time video surveillance. In such situations there is often a glut of video data (e.g., at a mass transit centre there are multiple video streams covering many people) coupled with constrained processing power (due to cost limitations). The complexity of the proposed method is similar to standard PCA, while achieving considerably better performance as demonstrated on several public databases.

Researchers have previously developed various methods to improve PCA or FLD by whitening [21], [22], [23] to compensate for the overweighting of the leading features, based on the observation that not all features have the same importance in recognition. However, normal whitening may excessively enhance minor features which leads to over-fitting to the training data. It is difficult to assign appropriate weights to all features in the high dimensional space at the same time. We thus design a learning method to weight features pairwise.

Consider two features a and r from the m dimensional space. We assign weight $\eta_{a,r} \in [0,1]$ to feature a, and weight $\sqrt{1-\eta_{a,r}^2}$ to feature r (the choice is explained later). Now we define the difference of two face images $I_{j,k}$ and $I_{j'k'}$ lying in the subspace defined by features a and r as the Euclidean distance of their transformed vectors $\tilde{s}_{j,k}$ and $\tilde{s}_{j',k'}$ in rotated face space as follows:

$$d_{jk,j'k'} = ||M(\eta_{a,r})\tilde{s}_{j,k} - M(\eta_{a,r})\tilde{s}_{j',k'}||_2 \tag{1}$$

where $M(\eta_{a,r})$ is an $m \times m$ square matrix with elements $M_{a,a} = \eta_{a,r}$ and $M_{r,r} = \sqrt{1-\eta_{a,r}^2}$, with all other elements being zeros. We define a continuous cost function Λ to search the one dimensional space to determine the optimal value for $\eta_{a,r}$ as follows:

$$\begin{aligned} \Lambda(\eta_{a,r}) \quad = \quad & \sum\nolimits_{j=1}^{N} \sum\nolimits_{k=1}^{K_j} \sum\nolimits_{n} \left(\frac{d_{jk,j0}}{d_{jk,n0}} \right) \\ & \forall n \in d_{jk,n0} < d_{jk,j0}, n \in [1 \cdots N] \end{aligned} \tag{2}$$

where $d_{jk,j0}$ is the within-class difference between the sample $I_{j,k}$ and its corresponding reference image $I_{j,0}$ in class S_j. Note that the condition $d_{jk,n0} < d_{jk,j0}$ is only true when there is a misclassification error. The optimal weight for feature pairs a and b is found with:

$$\eta_{a,r} = \arg\min_{\widehat{\eta}_{a,r}} \Lambda(\widehat{\eta}_{a,r}) \tag{3}$$

We assign the weight $\sqrt{1-\eta_{a,r}^2}$ to feature r so that $\eta_{a,r}^2 + \left(\sqrt{1-\eta_{a,r}^2}\right)^2 = 1$. This ensures that Eqn. (1) is comparable across different feature pairs.

We empirically found that the shape of most Λ curves tends to be approximately concave, and hence elected to use a straightforward golden section search [24] for the minimisation.

At this stage we have the optimal weights for using a pair of features for classification. We now find the weights for all features, as follows. First, a reference feature r is chosen from the m available features. Second, feature pair weighting is found for feature r paired with each of the remaining $m-1$ features. Consequently, we have a set of η values: $(\eta_{1,r},\ \eta_{2,r},\ \cdots,\ \eta_{r-1,r},\ \eta_{r+1,r},\ \cdots,\ \eta_{m,r})$. The weights of all features are found in a chain, by updating each feature's weight in relation to the reference feature and updating the weights of the preceding features in the chain. The weights for each feature present in the chain must satisfy the following constraints:

$$\frac{w_a}{w_r} = \frac{\eta_{a,r}}{\sqrt{1-\eta_{a,r}^2}} \tag{4}$$

$$\sum\nolimits_{i\in\Psi} w_i^2 = 1 \tag{5}$$

where w_r is the weight for the reference feature, w_a is the weight for an arbitrary feature (excluding the reference feature r) and Ψ is the set of features present in the chain. The constraints ensure that the ratio between weights of an arbitrary feature and the reference feature is equivalent to the ratio of the weights in the corresponding feature pair.

As an example, let us assume there are only two features in the chain, w_r and w_f. Following constraints (4) and (5) leads to $w_r = \sqrt{1-\eta_{f,r}^2}$ and $w_f = \eta_{f,r}$. If a feature g is added to the chain, the following weights are obtained:

$$w_r = \frac{1}{\sqrt{\frac{\eta_{f,r}^2}{1-\eta_{f,r}^2} * \frac{\eta_{g,r}^2}{1-\eta_{g,r}^2} + 1}} \tag{6}$$

$$w_f = \frac{\eta_{f,r}}{\sqrt{1-\eta_{f,r}^2}} * \frac{1}{\sqrt{\frac{\eta_{f,r}^2}{1-\eta_{f,r}^2} * \frac{\eta_{g,r}^2}{1-\eta_{g,r}^2} + 1}} \tag{7}$$

$$w_g = \frac{\eta_{g,r}}{\sqrt{1-\eta_{g,r}^2}} * \frac{1}{\sqrt{\frac{\eta_{f,r}^2}{1-\eta_{f,r}^2} * \frac{\eta_{g,r}^2}{1-\eta_{g,r}^2} + 1}} \tag{8}$$

(9)

When dealing with face data, we have found that the following approximate relationship tends to occur:

$$\frac{w_f}{w_g} \simeq \frac{\eta_{f,g}}{\sqrt{1-\eta_{f,g}^2}} \tag{10}$$

which suggests that the weight ratio between two arbitrary features and the ratio of the weights in the corresponding feature pair is approximately maintained.

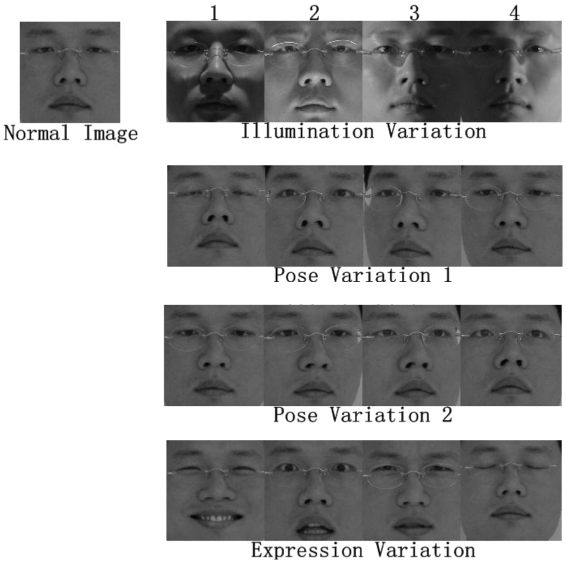

Fig. 3. Sample images from the Asian Face Database.

IV. Evaluation

As we are currently in the process of creating a suitable dataset for face classification in CCTV conditions (part of a separately funded project), here we compare the performance (in terms of recognition accuracy) of the CWFP method on three publicly available databases: Asian Face Database [25], PIE [26], and FERET [27]. The performance is compared against five techniques: standard PCA and FLD (on all databases), Synthesis+PCA [28] (on PIE and FERET databases), Pose-Robust Features [28] (on PIE and FERET databases), and Eigen Light-Fields [29] (on the FERET database). In the Synthesis+PCA method, an Active Appearance Model (AAM) [30] is fit to a given non-frontal face, followed by transformation of the AAM's parameters to represent the frontal view. The frontal face is then synthesised and fed to a standard PCA based recognition system. In the Pose-Robust Features method, the synthesis step is skipped and the transformed AAM parameters are used directly for recognition [28].

For all trials, we divide the corresponding data set into three equal-sized disjoint partitions with different subjects. We then choose images from one of the partitions for training and the remaining two partitions for testing. In each case the training set is used to construct the face space and weight feature pairs. The test set contains images of unseen subjects. For testing, we only register one neutral normally lit frontal image per subject as gallery and use the remainder of the images as probe. All the results are the average of three-fold cross validation using three different partitions of the datasets.

Figure 3 shows some images from the Asian Face Database, which contains 103 persons. Each person has 17 images including 1 normal face, 4 illumination variations, 8 pose variations (each about 15 degrees), and 4 expression variations. All images are grayscale of size of 125 × 125 pixels and are aligned according to their eye positions. We only use one normal image (top-left in Figure 3) in the test dataset as the gallery image and use the remainder as the probe images.

Table I shows the results on the Asian Face Database. Here, PCA+CWFP performs better than PCA by a considerable margin — an average of 77.6% correct recognition vs 60%, respectively. The largest difference occurs for illumination changes, where PCA+CWFP is about three times better than PCA, due to PCA's sensitivity to within-class changes. Pose variations have less influence on PCA than illumination variations, with an average accuracy of 72.7% compared to 80.3% for PCA+CWFP. The performance of FLD is somewhat improved, with an average accuracy of 82.7% for FLD+CWFP and 80.9% for FLD.

FLD+CWFP performs slightly better than FLD in pose and expression variations, while FLD is a little better under lighting changes. All four methods are sensitive to expression changes with relatively lower accuracy. We conjecture that this is due to different people expressing the same expression somewhat differently to others, which makes expression changes harder to model.

The worst recognition rate of 60.6% for FLD+CWFP is for expression change with eyes closed (the 4th one in the last row in Figure 3), which also affects PCA and FLD substantially as they achieve only 50.3% and 54.9% respectively. The reason is that the alignment of face images relies heavily on the eyes – with the eyes closed, the alignment is less accurate, leading to differences in scale.

Overall, CWFP can noticeably improve the performance of both PCA and FLD. FLD+CWFP is more robust to illumination, expression, and pose variations than other methods with relatively little change in accuracy across all three variations.

For comparison with the Synthesis+PCA and Pose-Robust Features methods, the pose variation subsets of the PIE and FERET databases are used. On the PIE database, three poses are used: ±22.5° and 0°. On the FERET database, nine poses are used: ±60°, ±40°, ±25°, ±15° and 0° (i.e. the 'b' subset). For each person, the frontal face image was the gallery image and the remaining images were the probe images. All the images were horizontally scaled and aligned according to their eye positions. This normalisation is an approximation of the 3-point normalisation used in [29].

Tables II and III show the results[1] obtained on the PIE and FERET databases, respectively. The CWFP method remarkably improves the performance of PCA – on the PIE and FERET databases the average improvement is approximately 25 and 21 percentage points, respectively. It also increases the average accuracy of FLD by approximately 43 and 10 percentage points, respectively. This effect is more significant for poses with angles greater than ±40°. Out of of the six methods, FLD+CWFP is the best performer across all pose angles.

Table IV shows the comparison with the Eigen Light-Fields [29] method. For consistency with the results presented in [29], we report the average recognition accuracy across all poses, using each pose angle separately for gallery images. The result of standard PCA method in our test is 40.6%, comparable to 39.4% in [29], which implies that our image normalisation is a close approximation of the 3 point normalisation. From Table IV, we observe that when CWFP is applied, the accuracy of PCA and FLD increases by approximately 29 and 10 percentage points, respectively. Moreover, FLD+CWFP outperforms Eigen Light-Fields remarkably. We note that in CWFP we do not need to determine the pose angles of the images, while in Eigen Light-Fields method camera intrinsics and relative orientation of the camera to the object should be acquired beforehand. This is often difficult or impossible in some situations.

TABLE I
RECOGNITION ACCURACY ON THE ASIAN FACE DATABASE

Variation Type	Method	Database subset				Average
		1	2	3	4	
Illumination	PCA	16.9	39.4	26.8	32.4	28.9
	PCA+CWFP	**80.3**	90.1	66.2	85.9	80.6
	FLD	73.2	**94.4**	**91.5**	95.8	**88.7**
	FLD+CWFP	70.4	93.0	88.7	**97.2**	87.3
Pose 1	PCA	84.5	77.5	62.0	74.6	74.7
	PCA+CWFP	90.1	81.7	69.0	84.5	81.3
	FLD	93.0	87.3	74.6	88.7	85.9
	FLD+CWFP	**93.1**	**87.3**	**80.3**	**90.1**	**87.7**
Pose 2	PCA	83.1	70.4	60.6	69.0	70.8
	PCA+CWFP	85.9	78.9	73.2	78.9	79.2
	FLD	88.7	77.5	73.2	81.7	80.3
	FLD+CWFP	**88.7**	**78.9**	**76.1**	**87.3**	**82.7**
Expression	PCA	80.3	67.6	64.8	50.7	65.9
	PCA+CWFP	85.9	**74.6**	66.2	49.3	69.0
	FLD	88.7	69.0	62.0	54.9	68.7
	FLD+CWFP	**91.5**	71.8	**67.6**	**60.6**	**72.9**

TABLE II
RECOGNITION ACCURACY ON THE PIE DATABASE. RESULTS FOR SYNTHESIS+PCA AND POSE-ROBUST FEATURES ARE FROM [28].

Pose	PCA	FLD	Synthesis + PCA	Pose-Robust Features	PCA+ CWFP	FLD+ CWFP
-22.5°	30.2	62.3	60.0	83.3	67.9	**94.3**
-22.5°	13.2	37.7	56.0	80.6	24.5	**90.6**

TABLE III
RECOGNITION ACCURACY ON THE FERET DATABASE. RESULTS FOR SYNTHESIS+PCA AND POSE-ROBUST FEATURES ARE FROM [28].

Pose	PCA	FLD	Synthesis + PCA	Pose-Robust Features	PCA+ CWFP	FLD+ CWFP
−60°	23.3	62.4	-	-	45.9	**75.9**
−40°	36.8	71.4	-	-	56.4	**85.0**
−25°	53.4	78.2	50.0	85.6	75.9	**88.0**
−15°	79.7	84.2	71.0	88.2	81.2	**91.0**
+15°	66.1	85.7	67.4	88.1	80.5	**92.5**
+25°	46.6	81.2	42.0	66.8	76.7	**91.0**
+40°	35.3	75.9	-	-	66.2	**86.5**
+60°	28.6	69.2	-	-	55.6	**77.4**

TABLE IV
RECOGNITION ACCURACY ON THE THE FERET DATABASE. RESULTS FOR EIGEN LIGHT-FIELDS ARE FROM [29].

Method	PCA	FLD	Eigen Light-Fields	PCA+ CWFP	FLD+ CWFP
Avg. Accuracy	40.6	76.0	75.0	69.2	**86.3**

[1] Our results for PCA is somewhat different from [28] as our PCA space is constructed from sample images in the PIE database (which were not used as gallery or probe images), while the PCA space in [28] was constructed from the Asian Face database.

V. Conclusions and Future Directions

In this paper we have proposed a fast appearance-based method, dubbed *Chained Weighted Feature Pairs (CWFP)*, for robust face recognition in conditions that can be present in surveillance applications (i.e. changes in pose, illumination, and expression). CWFP consists of two main steps: (1) feature pair weighting to assign optimal weights to features; and (2) a feature chain construction to combine feature pairs in order to find the weights for all features. The method was designed to be of low-complexity in order to facilitate use in real-time surveillance applications. Empirical comparisons on three publicly available databases show that CWFP can significantly improve the recognition performance of both PCA and FLD. Moreover, FLD+CWFP provides considerably improved recognition performance against three recent appearance-based recognition methods: Synthesis+PCA, Pose-Robust Features, and Eigen Light-Fields. However, being a holistic method, CWFP is still sensitive to geometric transformations such as scale changes and translation; we note that the technique presented in [31] can be adopted to overcome these drawbacks.

The natural next step in our surveillance project is extended trials of the proposed algorithm with real-life surveillance data from mass transport public spaces, which we are currently in the process of collecting. Prior to being able to collect the data, we encountered several non-technical issues. Privacy laws or policies at the national, state, municipal or organisational level may prevent surveillance footage being used for research even if the video is already being used for security monitoring – the primary purpose of the data collection is the main issue here. Moreover, without careful consultation and/or explanation, privacy groups as well as the general public can become uncomfortable with security research. Plaques and warning signs indicating that surveillance recordings are being gathered for research purposes may allow people to consciously avoid monitored areas, possibly invalidating results. Nevertheless, it is our experience that it is possible to negotiate a satisfying legal framework within which real-life trials of intelligent surveillance systems can proceed.

Acknowledgements

This project is supported by a grant from the Australian Government *Department of the Prime Minister and Cabinet* as well as the Australian Research Council through the Research Network for Securing Australia. NICTA is funded by the Australian Government as represented by the *Department of Broadband, Communications and the Digital Economy* and the Australian Research Council through the ICT Centre of Excellence program.

References

[1] A. Bigdeli, B. Lovell, and C. Sanderson, "Vision processing in intelligent CCTV for mass transport security," in *Proc. of SAFE 2007: Workshop on Signal Processing Applications for Public Security and Forensics*, 2007.

[2] A. Hamilton, "Building scalable situational intelligence," *Asian Security Review*, pp. 12–13, August 2007.

[3] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and M. Bone, "Face recognition vendor test 2002," in *Proc. Analysis and Modeling of Faces and Gestures*, 2003, p. 44.

[4] K. Bowyer, K. Chang, and P. Flynn., "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition." *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.

[5] Y. Gao and M. K. Leung, "Face recognition using line edge map," *IEEE Trans. PAMI*, vol. 24, no. 6, pp. 764–779, 2002.

[6] A. Yilmaz and M. Gokmen, "Eigenhill vs. eigenface and eigenedge," in *Proc. of Intl Conf. on Pattern Recognition*, 2000.

[7] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. PAMI*, vol. 25, no. 2, pp. 218–233, 2003.

[8] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. PAMI*, vol. 23, no. 6, pp. 643–660, 2001.

[9] M. J. Black, D. J. Fleet, and Y. Yacoob, "Robustly estimating changes in image appearance," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 8–31, 2000.

[10] X. Liu, T. Chen, and B. V. Kumar, "Face authentication for multiple subjects using eigenflow," *Pattern Recognition*, vol. 36, pp. 313–328, 2003.

[11] A. M. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class," *IEEE Trans. PAMI*, vol. 24, no. 6, pp. 748–763, 2002.

[12] L. Wiskott, J. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. PAMI*, vol. 7, no. 19, pp. 775–779, 1997.

[13] P. Sankaran and V. Asari, "A multi-view approach on modular PCA for illumination and pose invariant face recognition," in *Proc. of Applied Imagery Pattern Recognition Workshop*, 2004.

[14] W. Gao, S. Shan, X. Chai, and X. Fu, "Virtual face image generation for illumination and pose insensitive face recognition," in *Proc. of Intl Conf. on Multimedia and Expo*, 2003.

[15] C. Sanderson, S. Bengio, and Y. Gao, "On transforming statistical models for non-frontal face verification," *Pattern Recognition*, vol. 39, no. 2, pp. 288–302, 2006.

[16] T. Shan, B. C. Lovell, and S. Chen, "Face recognition robust to head pose from one sample image," in *Proc. of Intl Conf. on Pattern Recognition*, 2006.

[17] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensation for changes in illumination direction," *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 721–732, 1997.

[18] V. Blanz, K. Scherbaum, and H.-P. Seidel, "Fitting a morphable model to 3D scans of faces," in *IEEE 11th International Conference on Computer Vision*, 2007.

[19] C. D. Castillo and D. W. Jacobs, "Using stereo matching for 2D face recognition across pose," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2007.

[20] A. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Trans. PAMI*, vol. 29, no. 11, pp. 1927–1943, 2007.

[21] S. Chen and B. C. Lovell, "Illumination and expression invariant face recognition with one sample image," in *Proc. of Intl Conf. on Pattern Recognition*, 2004.

[22] C. Liu and H. Wechsler, "Enhanced fisher linear discriminant models for face recognition," in *Proc. of Int.l Conf. on Pattern Recognition*, 1998.

[23] ——, "Evolutionary pursuit and its application to face recognition," *IEEE Trans. PAMI*, vol. 22, no. 6, pp. 570–582, 2000.

[24] J. Kiefer, "Sequential minimax search for a maximum," *Proceedings of the American Mathematical Society*, vol. 4, pp. 502 – 506, 1953.

[25] Asian Face Image Database PF01, "Intelligent Multimedia Lab, Pohang University of Science and Technology," http://nova.postech.ac.kr/.

[26] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. PAMI*, vol. 25, no. 12, pp. 1615 – 1618, 2003.

[27] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. PAMI*, vol. 20, no. 10, pp. 1090–1104, 2000.

[28] C. Sanderson, T. Shan, and B. C. Lovell, "Towards pose-invariant 2D face classification for surveillance," in *Analysis and Modeling of Faces and Gestures (AMFG), Lecture Notes in Computer Science (LNCS)*, vol. 4778, 2007, pp. 276–289.

[29] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. PAMI*, vol. 26, no. 4, pp. 449– 465, 2004.

[30] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[31] H. Bischof and A. Leonardis, "Robust recognition of scaled eigenimages through a hierarchical approach," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 1998.