Feature Modelling of PCA Difference Vectors for 2D and 3D Face Recognition

Chris McCool, Jamie Cook, Vinod Chandran and Sridha Sridharan
Image and Video Research Laboratory
Queensland University of Technology,
2 George Street, Brisbane, Australia, 4001, GPO 2434
{c.mccool, j.cook, v.chandran, s.sridharan}@qut.edu.au

Abstract

This paper examines the the effectiveness of feature modelling to conduct 2D and 3D face recognition. In particular, PCA difference vectors are modelled using Gaussian Mixture Models (GMMs) which describe Intra-Personal (IP) and Extra-Personal (EP) variations. Two classifiers, an IP and IPEP classifier, are formed using these GMMs and their performance is compared to that of the Mahalanobis cosine metric (MahCosine). The best results for the 2D and 3D face modalities are obtained with the IP and IPEP classifiers respectively. The multi-modal fusion of these two systems provided consistent performance improvement across the FRGC database v2.0.

1. Introduction

Multi-modal face recognition is a growing area of research due to the fact that it is a non-intrusive biometric. A plethora of work has been conducted in both 2D and 3D face recognition. This work concentrates on the use of distance metrics such as L_1, L_2 and cosine; for example in PCA research [11]. However, limited work has been conducted in the area of feature modelling.

Research into 2D face recognition has been conducted for over 20 years. Several methods have been proposed to conduct 2D face recognition ranging from Principal Component Analysis (PCA) [13], free parts 2D Discrete Cosine Transform (2D-DCT) [10] through to morphable models [1]. The *eigenfaces* technique, a *de facto* standard of face recognition, was introduced by Turk and Pentland in [13]. In this work a vector space is created using PCA and feature vectors are obtained by projecting a face image into this vector space. The free parts 2D-DCT based technique was proposed in [10] and extracts multiple 2D-DCT feature vectors by decomposing the face into free parts; the information about the location of these parts is not retained. These 2D-DCT feature vectors are modelled using Gaussian Mixture Models (GMMs) where a background GMM is derived and then used to generate client dependent models; verification is then performed using the log-likelihood ratio of the background and adapted GMMs.

Face recognition using the 3D modality has received greater interest recently due to the decreasing costs in accurate capture devices such as laser range scanners and structured light scanners. Proposed methods for conducting 3D face recognition include the use of point signatures [15], isometric transformations [3] and PCA [4]. Some recent work has explored the modelling of 3D data. In [5] the registration error from the Iterative Closest Point algorithm was modelled with GMMs. For a more in depth review of 3D face recognition readers are referred to [2].

Fusion of the 2D and 3D modalities for face recognition is receiving greater attention, as it has thus far proved to be a more effective recognition method than either modality on its own. The complementary nature of the 2D and 3D modalities means that the weaknesses from any one modality can be reduced. Combining the two modalities, 2D and 3D, can be broadly divided into data fusion, feature fusion and classifier fusion. This work and much of the current work into fused 2D and 3D face recognition explores methods for classifier fusion. In [4] separate 2D and 3D classifiers were formed using PCA-based features, a hybrid 2D+3D classifier was then formed by fusing the individual classifiers with linear weights based on the Rank scores for identification. In [6] log-Gabor features from 2D and 3D part face images were reduced with PCA and a 2D+3D classifier was formed using weighted fusion.

The work presented in this paper examines the effectiveness of feature modelling for the task of face recognition. In particular the effectiveness of feature modelling across two modalities, image data (2D) and structural data (3D), is compared to the Mahalanobis cosine metric (MahCosine); the MahCosine metric was shown in [7] to be an effective distance metric. The experiments in this paper demonstrate that feature modelling for the 3D modality yields an improved classifier and that the hybrid 2D+3D classifier with

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

feature modelling yields an improved classifier than either modality alone. The fusion of the two modalities, 2D and 3D, is approached as classifier fusion and is shown to provide consistent performance improvements.

The paper is structured as follows. The methods used for data normalisation are described in Section 2. Section 3 then describes the method for feature modelling of difference vectors. The experimental procedure is outlined in Section 4 and the results are presented in Section 5. Conclusions and future work are then presented and discussed in Section 6.

2. Data Normalisation

Data normalisation provides a common basis for comparison of two images (or signals). It was shown in [4] that low resolution 3D face images significantly hinder recognition performance, however the same is not true of 2D face images. For these experiments the 2D and 3D face images were retained at different resolutions, the 2D face data has a resolution of 64×64 pixels whereas the the 3D face data has a resolution of 128×128 pixels. For both the 2D and 3D face images only part face images (above the mouth region and below the brow region) were used. This was done to reduce the effect of expression variation.

For the 2D face data only in-plane rotations could be recovered and to achieve this the two eye corners were used to perform the in-plane normalisation and cropping of the images. Registration of the 2D images is based on the eye corners; they reside on the same y-axis and are separated by 64 pixels. Illumination normalisation was conducted by applying local mean window normalisation; a technique previously used in [9]. An example 2D face image which is cropped and fully normalised is provided in Figure 1.

The 3D data consisted of point cloud data on a semi-regular x- and y-grid; there is limited variation along the grid lines. Three landmark points are used to fully normalise the 3D data for in-plane and out-of-plane rotations. The three landmark points used are the right eye corner, left eye corner and chin. After the data has been normalised for all rotations the valid data is then interpolated onto a regular grid of 128×128 pixels. Further processing is then conducted to remove erroneous points using a gradient filter. Finally the data is smoothed using a median filter to reduce the effect of noise. An example of the final 3D data is provided in Figure 2.

To provide a consistent basis for comparison the 3D data is 3D registered and range normalised. Registration is achieved by placing the eye corners on the same y-gridline separated by 128 pixels. The 3D data is range normalised so that the maximum value in every image is set to 255, all the other values are adjusted to be relative to this value. When processing the 3D face image the data is treated in the same way as 2D data, thereby treating it as $2\frac{1}{2}$D data, and example of this concept is provided in Figure 3.

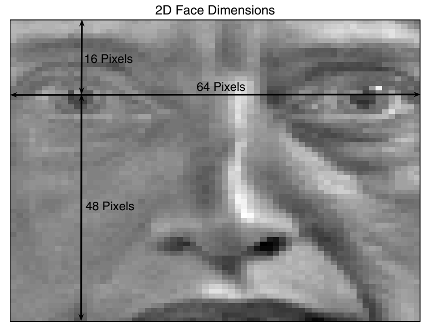

Figure 1. A cropped and fully normalised 2D face image. The image is of a size 64×64 pixels. Included in this image are the dimensions defining the position of the registered eye corners.

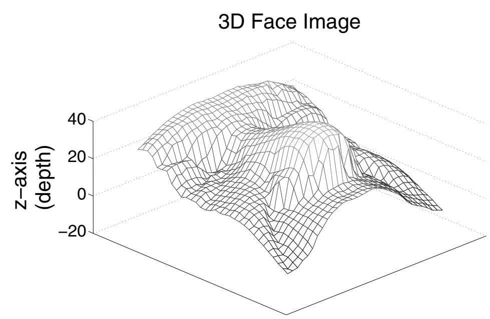

Figure 2. A mesh plot of a interpolated and smoothed 3D face image.

3. Feature Modelling

Feature modelling of global, in this case PCA based, feature vectors is a difficult task due to the limited number of training observations available. In most situations the number of observations available for training is equal to the number of training images,

$$N_{train\_images} = \sum_{i=1}^{D} n_i, \tag{1}$$

where D is the number of individuals (IDs) in the training set and n_i is the number of images available for the i^{th} ID.

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

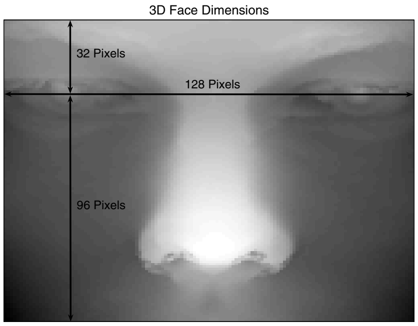

Figure 3. An example of how the 3D data can be treated as 2D data, the 3D data is simply treated as an image. Included in this image are the points used to register the 3D data.

To increase the amount of training data, difference vectors are formed. These difference vectors describe two forms of variation, Intra-Personal (IP) and Extra-Personal (EP).

The IP difference vectors are used to described variation that occurs between images of the same ID. However, practical limitations mean that there is rarely enough data to derive a client specific IP model. In this work a global IP model, Ω_{IP}, is derived, and as such all the IP difference vectors in the training set are able to be used. It's noted that by deriving a global IP model there is an assumption that the IP variation is similar for all individuals.

The EP difference vectors are used to described variation that occurs between images of different people. Were this to be modelled in a global manner this would result in a model that describes noise. This is because it would be attempting to model all variations between all the different individuals. Therefore, the EP difference vectors are formed in a client specific manner.

When forming the IP difference vectors all the permutations of the IP difference vectors are used. This means there are,

$$N_{IP\_obs} = \sum_{i=1}^{D} {}_{n_i}P_2, \tag{2}$$

observations available for training. The client specific EP model, $\Omega_{i,EP}$, is derived using the difference between the enrolled feature vectors and all the training feature vectors. Therefore, if there are E enrolled feature vectors there will be,

$$N_{EP\_obs} = 2 * N_{images} * E \tag{3}$$

observations to train $\Omega_{i,EP}$. The factor of 2 is introduced because all the permutations are formed.

In order to classify a difference vector using the two models, Ω_{IP} and Ω_{EP}, the log-likelihood ratio (LLR),

$$g(\mathbf{x}) = \ln(p(\mathbf{x} \mid \Omega_{IP})) - \ln(p(\mathbf{x} \mid \Omega_{EP})), \tag{4}$$

is used. The LLR can be viewed as providing score normalisation to Ω_{IP} by using information from Ω_{EP}. The term $p(\mathbf{x} \mid \Omega_{IP})$ is the probability that observation $\mathbf{x}$ belongs to class Ω_{IP} and $p(\mathbf{x} \mid \Omega_{EP})$ is the probability that observation $\mathbf{x}$ belongs to class Ω_{EP}.

The LLR, Equation 4, is a reduced form of the discriminant function,

$$g(\mathbf{x}) = \ln(\frac{p(\mathbf{x} \mid \Omega_{IP})}{p(\mathbf{x} \mid \Omega_{EP})}) + \ln(\frac{P(\Omega_{IP})}{P(\Omega_{EP})}), \tag{5}$$

where $P(\Omega_{IP})$ represents the probability of class Ω_{IP} and $P(\Omega_{EP})$ is the probability of class Ω_{EP}. By considering both classes to be equally likely, $P(\Omega_{IP}) = P(\Omega_{EP}) = 0.5$, then Equation 5 simplifies to Equation 4.

4. Experiments

The experiments were conducted on the validation set of the FRGC v2.0 database [12]. This database was split into two sessions, *Fall* and *Spring*. The *Fall* session consists of 2D data and 3D data collected in Fall 2003 while data for the *Spring* session was collected in Spring 2004; within each session there is at least a 1 week time lag between subsequent image captures. This data was further split into 4 disjoint sets in order to perform cross-validation experiments. These disjoint sets were split between the Train, Tune and Test sets with a 2:1:1 ratio. The Train set was used to derive the PCA space, global IP GMM and for background data to train client specific EP GMMs. The Tune set was used to derive fusion weights and the Test set was used to perform testing; including enrollment. Using these cross-validation splits two sets of experiments were defined.

The first set of experiments examined the effectiveness of the IP and IPEP feature modelling techniques for the 2D and 3D modalities respectively. The second set of experiments explored methods for combining the 2D and 3D classifiers to form a fused 2D+3D classifier. Both sets of experiments examined the effect of same-session and between-session Train and Test conditions. Same-session conditions examined the effect of training and testing on the same session of data (same-session variation); for example train on *Fall* and test on *Fall*. Whereas, between-session conditions examined the effect of training and testing on data from different session (between-session variation); for example train on *Fall* and test on *Spring*.

4.1. Enrollment and Background Data

For these experiments the Train set and enrollment data were used to derive the GMMs [14]. The enrollment process consisted of randomly taking E (for these experiments $E = 4$) images for every ID in the set (Tune or Test) with $E + 1$ images; this left at least 1 image, for each enrolled ID, to be used for testing. The IP GMMs were derived globally, to describe all the IDs rather than an individual one, by forming all the permutations of those IDs in the Train set with E images; this was done to avoid any bias in the training scheme. The EP GMMs were derived for each enrolled ID using the comparison of every enrolled image against all the images used to derive the IP GMM. The number of components for the GMMs range from 4 to 16; these limited sizes were used as there was limited training data available. The E enrollment images meant that for each ID multiple scores were obtained which provide multiple chances for the correct decision to be made. To perform score modelling the best score was chosen as the matching core.

5. Results and Discussion

The experimental results indicate that the IPEP system is an effective classifier for the 3D face modality and the IP system is an effective classifier for the 2D face modality. Furthermore, it is shown that the fusion of 2D IP and 3D IPEP classifiers outperforms the fused 2D and 3D MahCosine classifiers. The feature modelling systems were found to perform best with 16 components and so only these results are presented. An in depth analysis of the results is provided below.

5.1. Feature Modelling - 2D

The results for the 2D modality demonstrate that the best overall recognition system is the IP classifier. The results for the 2D experiments, Table 1, indicate that for same-session conditions the IPEP classifier outperforms all other systems while for the between-session conditions the IP classifier outperforms all other systems. It is noted that although the IP system is the worst performing classifier for the same-session conditions it performs significantly better for the between-session conditions, which is the more difficult test case.

The DET plot provided in Figure 4 highlights that when there are between-session variations the IP recognition system performs significantly better. It can also be seen that the use of the EP model leads to a significant reduction in performance. This reduction in performance is most likely due to the significant time lapse between the *Fall* and *Spring* sessions leading to increased variation in the capture conditions.

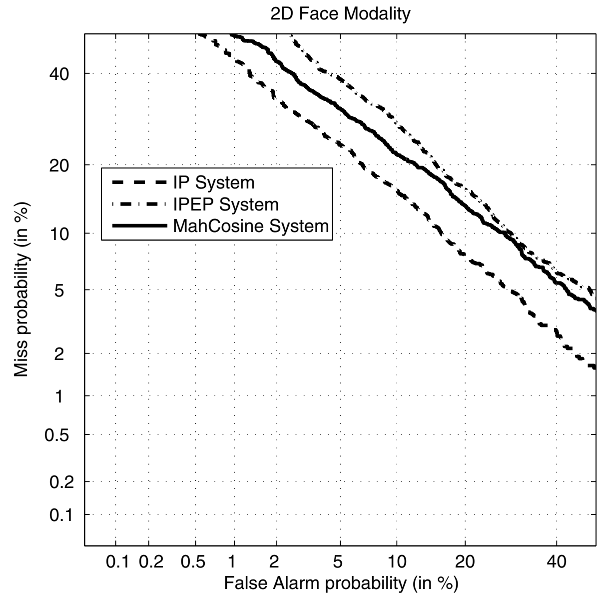

Figure 4. A DET plot for the 2D face modality. Training is conducted on the *Fall* session and testing is conducted on the *Spring* session.

5.2. Feature Modelling - 3D

The results for the 3D modality demonstrate that the best recognition system is the IPEP classifier. This results is consistent for same- and between-session variations, as can be seen in the EERs provided in Table 2. The DET plot provided in Figure 5 shows that the IPEP classifier is superior to both the IP and MahCosine classifiers. It is noted that unlike the 2D modality the inclusion of the EP model provides consistent improved system performance; across both same- and between-session variations. This is most likely due to the fact that variation of the 3D data (structural data) is not as significant across time as variation of the 2D data; for instance 3D data is less sensitive to illumination and pose variations.

5.3. Fused 2D and 3D Classifier

Combining the classifiers from the 2D and 3D modalities was conducted to demonstrate the complementary information was still being captured, even when feature modelling was being conducted. In this work the MahCosine classifiers for both modalities, 2D and 3D, were combined and the various combinations of IP and IPEP classifiers for both modalities were explored. Fusion was conducted using the sum rule which was shown in [8] to be an effective method of combining classifiers. The results for these experiments are provided in Table 3. It can be seen from these results that

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

Train Set	Test Set	MahCosine Classifier	IP GMM Classifier	IPEP GMM Classifier
Fall	*Fall*	7.55%	8.65%	**6.79%**
Spring	*Spring*	8.20%	8.89%	**7.43%**
Spring	*Fall*	14.42%	**11.87%**	16.12%
Fall	*Spring*	16.52%	**12.77%**	17.53%

Table 1. Results for the 2D face modality, the results are presented as EERs. Highlighted is the best classifier for each test case.

Train Set	Test Set	MahCosine Classifier	IP GMM Classifier	IPEP GMM Classifier
Fall	*Fall*	4.33%	6.62%	**2.88%**
Spring	*Spring*	5.02%	7.12%	**3.65%**
Spring	*Fall*	5.51%	6.62%	**4.66%**
Fall	*Spring*	9.09%	9.34%	**7.94%**

Table 2. Results for the 3D face modality, the results are presented as EERs. Highlighted is the best classifier for each test case.

Train Set	Test Set	2D+3D MahCosine Classifier	2DIP + 3DIP GMM Classifier	2DIPEP + 3DIP GMM Classifier	2DIP + 3DIPEP GMM Classifier	2DIPEP + 3DIPEP GMM Classifier
Fall	*Fall*	2.88%	4.24%	3.65%	2.20%	**1.78%**
Spring	*Spring*	3.37%	5.21%	4.51%	2.35%	**2.16%**
Spring	*Fall*	4.33%	4.92%	5.51%	**3.22%**	3.82%
Fall	*Spring*	6.67%	6.61%	8.20%	**5.02%**	6.29%

Table 3. Results for the various combinations of 2D+3D classifiers, the results are presented as EERs. Highlighted is the best classifier for each test case.

for the same-session experiments the best system is the 2D IPEP + 3D IPEP while for the between-session experiments the best system is 2D IP + 3D IPEP; this outcome is consistent with the results obtained from testing the 2D modality. These results are highlighted in the DET plot provided in Figure 6.

6. Conclusions and Future Work

It has been shown that IP and IPEP Feature modelling can provide an improved recognition system than the MahCosine metric. For the 2D modality the best results were obtained using the IP classifier; this classifier performed the best in the most difficult test case of between-session variation. Whereas for the 3D modality the best results were obtained using the IPEP classifier. It is noted that as the number of components are increased the performance is increased. For these experiments only 16 components were used due to the limited amount of training data. It was also demonstrated the the combined 2D IP and 3D IPEP recognition system provided a consistent system for both the same- and between-session experiments.

Further work needs to be conducted to determine the optimal number of components, as there will be a threshold where performance improvements will not occur. In addition, the reason for failure of the EP model in the 2D modality needs to be investigated to determine if this can be addressed by performing more normalisation on the 2D data.

7. Acknowledgements

This research was supported by the Australian Research Council (ARC) through Discovery Grants Scheme, Grant DP0452676,2004-6.

References

[1] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1063–1074, 2003.

[2] K. Bowyer, K. Chang, and P. Flynn. A survey of approaches to three-dimensional face recognition. *Proceed-*

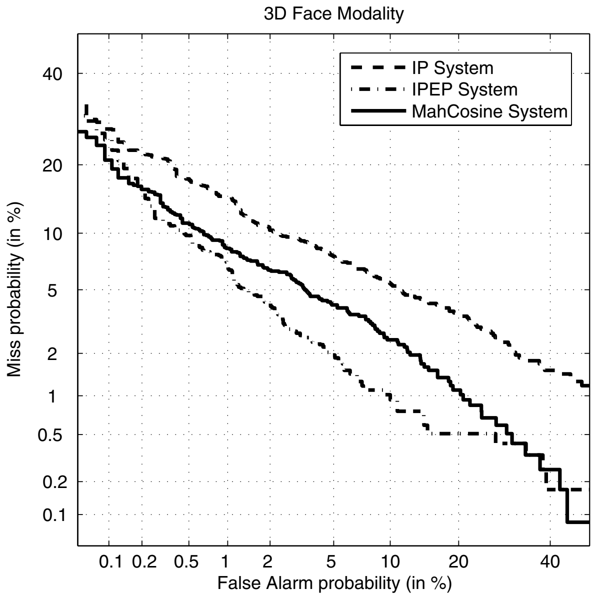

Figure 5. A DET plot for the 3D face modality. Training is conducted on the *Fall* session and testing is conducted on the *Fall* session.

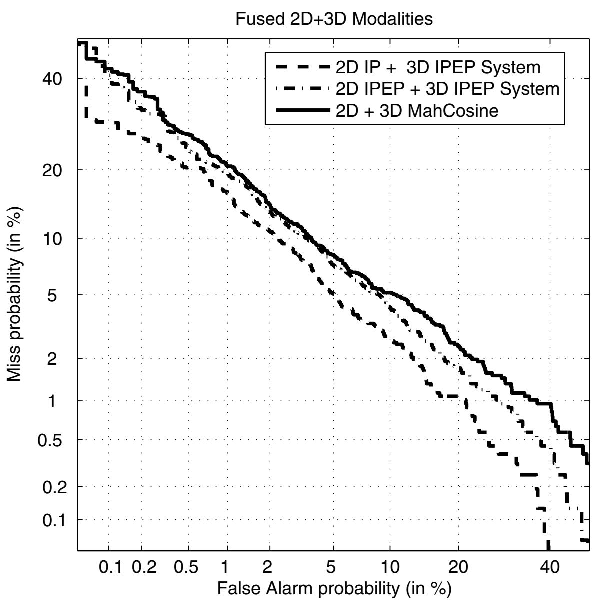

Figure 6. A DET plot for combining the 2D and 3D face modalities. Training is conducted on the *Fall* session and testing is conducted on the *Spring* session. In this case the IPEP classifier performs significantly better.

ings of the 17th International Conference on Pattern Recognition, 1:358–361, 2004.

[3] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Expression-invariant 3d face recognition. *Audio- and Video-Based Person Authentication*, pages 62–70, 2003.

[4] K. I. Chang, K. W. Bowyer, and P. J. Flynn. Face recognition using 2d and 3d facial data. *Workshop in Multimodal User Authentication*, pages 25–32, 2003.

[5] J. Cook, V. Chandran, S. Sridharan, and C. Fookes. Face recognition from 3d data using iterative closest point algorithm and gaussian mixture models. *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission*, pages 502–509, 2004.

[6] J. Cook, V. Chandran, S. Sridharan, and C. Fookes. Gabor filter bank representation for 3d face recognition. *Proceedings of Digital Image Computing: Techniques and Applications*, 2005.

[7] J. R. del Solar and P. Navarrete. Towards a generalized eigenspace-based face recognition framework. *The International Workshop on Statistical Techniques in Pattern Recognition*, 2002.

[8] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:226–239, 1998.

[9] S. Lowther, C. McCool, V. Chandran, and S. Sridharan. Improving face localisation using claimed identity for face recognition. *Proceedings of the 3rd Interational Workshop on the Internet, Telecommunications and Signal Processing*, 2004.

[10] S. Lucey. The symbiotic relationship of parts and monolithic face representations in verification. *2004 Conference of Computer Vision and Pattern Recognition Workshop*, page 89, 2004.

[11] H. Moon and P. J. Phillips. Computational and performance aspects of pca-based face-recognition algorithms. *Perception*, pages 303–321, 2001.

[12] J. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *Proceedings of IEEE Conference of Computer Vision and Pattern Recognition*, 1:947–954, 2005.

[13] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[14] R. Vogt, J. Pelecanos, and S.Sridharan. Dependence of gmm adaptation on feature post-processing. *Eurospeech*, pages 3013–3016, 2003.

[15] Y. Wang, C. Chua, and Y. Ho. Facial feature detection and face recognition from 2d and 3d images. *Pattern Recognition Letters*, 23:1191–1202, 2002.