Robust 3D Face Recognition from Expression Categorisation

Jamie Cook, Mark Cox, Vinod Chandran, and Sridha Sridharan

Speech, Audio, Image and Video Technology (SAIVT) Laboratory,
Queensland University of Technology, Brisbane, Queensland, 4000, Australia
jamie@ieee.org, {md.cox,v.chandran,s.sridharan}@qut.edu.au

Abstract. The task of Face Recognition is often cited as being complicated by the presence of lighting and expression variation. In this article a novel combination of facial expression categorisation and 3D Face Recognition is used to provide enhanced recognition performance. The use of 3D face data alleviates performance issues related to pose and illumination. Part-face decomposition is combined with a novel adaptive weighting scheme to increase robustness to expression variation. By using local features instead of a monolithic approach, this system configuration allows for expression variability to be modelled and aid in the fusion process. The system is tested on the Face Recognition Grand Challenge (FRGC) database, currently the largest available dataset of 3D faces. The sensitivity of the proposed approach is also evaluated in the presence of systematic error in the expression classification stage.

1 Introduction

Biometrics research has enjoyed a recent wave of increased interest fueled by a political climate demanding increased security. Face Recognition has the distinct advantage over other biometric modalities such as fingerprint, DNA and iris recognition, in that the acquisition stage is non-intrusive and can be achieved with readily available equipment. However widespread adoption of Face Recognition Technology (FRT) has been hindered by excessive sensitivity to 3 factors: pose, illumination and expression [1]. The use of 3D data has the potential to overcome issues relating to the first two factors and in this paper a novel system is demonstrated to alleviate performance degradation caused by the third.

Early work in 3D facial recognition emerged in the late 1980's but it wasn't until recently that substantial research databases have become available. The Face Recognition Grand Challenge (FRGC) [2] was created to address this issue and provides both a common dataset and experimental methodologies to enable accurate comparisons of different algorithms. A good summary of the current research in 3D and composite 2D-3D recognition is given in [3].

The task of expression classification is an interesting problem with the potential to further our understanding of inter-personal interaction and to enable sophisticated Human Machine Interfaces (HMI) [4]. There is considerable literature in the psychology community to suggest that in humans, recognition of

S.-W. Lee and S.Z. Li (Eds.): ICB 2007, LNCS 4642, pp. 271–280, 2007.

faces and expression comprehension occur in parallel with information fusion occuring subsequently.

Currently most automated face recognition systems provide robustness to expression variation by either a non-linear normalisation to remove expression [5] or by selecting features which are invariant to changes in expression. In [6], a 3D matching of faces is performed by defining three Regions Of Interest (ROI) around the nose which are deemed to be most stable in the presence of expression. These regions are matched between scans by means of the Iterative Closest Point (ICP) algorithm. Such an approach, however, doesn't make full use of the discriminable information that exists in the entire face. The explicit use of expression classification and categorisation to direct the operation of automated face recognition is a research area yet to be fully explored.

In 2002, Martinez [7] created 6 region and identity specific subspaces from face images in the AR database with the aim of alleviating the problems of occlusion and expression variation. The author posits that expressions do not manifest symmetrically on the human face and demonstrates that happy faces are better recognized by the left side while angry faces have better recognition from the right side. The author then uses a train set to define weights for each of three distinct expressions, which are used to modify the contribution from each of the 6 regions. Further testing on a test set with known expression showed improvements over an unweighted baseline when using weights appropriate to the currently displayed emotion.

In [8] an automated variant of this approach is detailed. A front end expression classification system is used to select from multiple classification systems. The authors posit the use of six expression categories, namely happiness, sadness, anger, fear, surprise and disgust. The core concept is demonstrated using a small database of 30 subjects and a "happy face" recognition system.

An important aspect which is not covered in either of the previous approaches is the so called "front end effect" [9]. This refers to the cascading effect of errors in the expression classification module (front end) to subsequent stages of the system. In the proposed approach, expression strength categorisation is combined with a part-face recognition system using an adaptive weighting scheme which is tolerant to expression misclassification. A discussion on automatic expression classification is given in Section 2 and details of the proposed system are presented in Section 3. Experimentation on the FRGC database and error analysis can be found in Section 4.

2 Expression Classification

Researchers in the field of facial expression and emotion analysis use the Facial Action Coding Scheme (FACS) as a method to encode the current state of the face as a combination of atomic facial actions. Subtlety of expression in the face is captured using an additional intensity parameter. After a face has been parameterised in this fashion, it can then be used to infer the underlying emotion or detect a particular facial expression [10].

For face recognition, comprehension of facial expression is of less use than a mapping of facial deformation. Such a mapping could allow the recognition algorithm to compensate for expression variation during the recognition process. The contribution of spatial regions exhibiting significant deformation can be adaptively de-emphasised while the contribution from portions of the face unaffected can be increased. Currently there exists no ground truth data for 3D face databases which provides a detailed annotation (FACS or other) of expression, and constructing such annotations is time consuming and expensive.

Instead, the manual annotations of the FRGC 3D data set by researchers at Geometrix [11] shall instead be used to demonstrate the proposed system. Each of the acquired images is allocated, based on the displayed expression, to one of the three classes: Neutral, Small (slight expression) and Large (highly expressive). A sample from each of the classes can be seen in Figure 1.

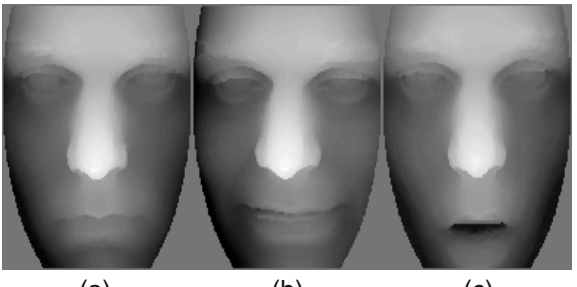


Fig. 1. Examples of FRGC 3D data for three classes of expression strength (a) Neutral (b) Small Expression and (c) Large Expression

The Geometrix annotations contain global rather than local deformation information, as such they can not be used to construct detailed mappings of which spatial regions to use in the recognition process. Instead, three generic weighting schemes shall be defined, correposponding to three classes of expression, details of this are given in Section 3.2. Given that knowledge of the expression displayed in the 3D scans being compared can be acquired, an appropriate weighting scheme can then be selected which emphasises regions that are more resilient to expression changes. It is expected that the extraction of FACS parameters would allow more detailed deformation maps to be constructed and hence allow more flexibilitiy in constructing weight vectors. Future work should utilise automated FACS annotations systems such as that detailed in [12].

3 Face Recognition

3.1 Part-Face Methodology

Face Verification techniques typically employ a monolithic representation of the face during recognition, however, approaches which decompose the face into

sub-regions have shown considerable promise. Many authors [13,14] have shown superior performance by adopting a modular representation of the face provided that face localisation is performed accurately [14].

The recognition system used in the following experiments is an extension of previous work in component face recognition [15], a block diagram of the matching process is shown in Figure 2. In this approach face images are decomposed into multiple regions which are classified independently using subspace projection. PCA with a Mahalanobis Cosine distance metric was chosen due to the [-1,1] bounded nature of the output. Alternate subspaces projection methods such as Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA) are also equally applicable. Late fusion is then used to recombine the classifier scores into a single classification decision.

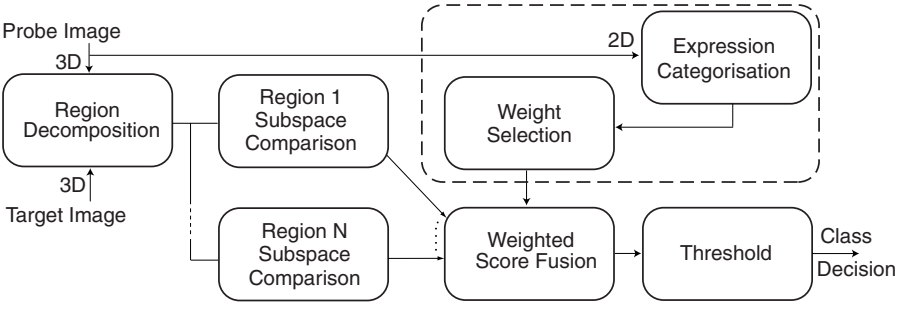


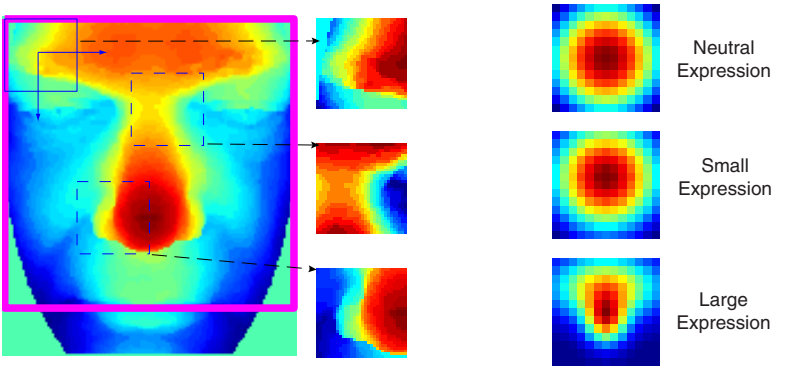
Fig. 2. Block Diagram of the proposed system (extension to previous work enclosed in dashed rectangle)

Regions decomposition is achieved by using a 32x32 pixel sliding window to extract a 13×13 grid of regions [demonstrated in Figure 3(a)]. The choice of window size is an important consideration for optimal performance and there is no single choice which will perform *best*. The selected window and step sizes were chosen so as to balance the conflicting needs to both accurately localise features and to encapsulate sufficient local information to enable discrimination.

3.2 Adaptive Weighting

Most existing recognition algorithms have satisfactory performance when the facial expression of gallery and probe images is similar. The goal then of expression classification in a Face Recognition context, should be to identify expression mismatch between the target and probe images and allow for graceful handling of such situations. *The novel contribution of this paper comes from the inclusion of the adaptive weighting scheme which modifies the behaviour of the fusion stage based on the detection of expression mismatch.*

Facial expression is a non-rigid distortion and as such can not be compensated for using standard global face normalisation techniques. When either the gallery



(a) (b)

Fig. 3. (a) Example 3D face with 32x32 pixel sliding window with 8 pixel shift. (b) Three expression specific weighting functions which are selected based on the currently displayed expression for use in the weighted score fusion process.

or probe image contains significant expression the assumption of correspondence between the two images breaks down. However due to the physical characteristics of the human face, this break down is not uniform; for example the nose and cheek bone areas are less prone to expression distortion than the mouth region [6].

By identifying images in which expression mismatch is present, the decision process can place more emphasis on those regions which are least affected. By looking at the distortions on a region by region basis, this system configuration can scale to any number of facial expressions without requiring the use of expression specific recognition systems as in [8]. This also has the benefit over systems such as in [6], that when expression mismatch does not occur information from the entire face can be utilised.

Previous work has demonstrated that discriminative information in 3D faces is distributed towards the center of the face [15]. This corroborates the generally held belief of many researchers that the nose region is most invariant to expression variations [6,3]. The weighting models chosen for experimentation are therefore modeled using Gaussian Mixture Models which naturally emphasise the central regions. The three weighting schemes corresponding to expression categories are illustrated in Figure 3(b).

In the Neutral case, the model has a single mixture which encompases a significant portion of the face. In the case of mild expression, the model is shifted higher towards the nasal bridge, reducing the contribution of the lips and cheeks which are posited to be more variable under expression variation. Finally, for large expressions, the dominant mixture is tightened to further exclude the cheeks and mouth and a second mixture is added to retain contributions from the eyes and brow area. These weight vectors, w, are all then normalised such that $\sum w_i = 1$.

4 Experimental Results

The experiments described in this section were conducted using 3D data provided as part of the Face Recognition Grand Challenge [2]. The FRGC dataset, which contains 4007 registered texture and shape images of 466 subjects, is currently the largest publicly available database of 3D face images. The data was collected by the Computer Vision Research Laboratory at the University of Notre Dame (UND) over 3 semesters using a Minolta Vivid 900 range finder. Although the following experimentation is limited to 3D faces, the proposed methodology is not inherently 3D based and as such can be easily transposed to the processing of traditional intensity images.

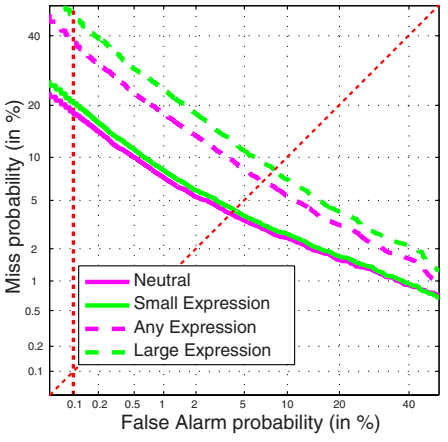
The 466 subjects in the database are broken into training and testing groups according to the specification of FRGC Experiment 3. There are 943 images in the training set, and of the 4007 images in the test set 59% are captured with a neutral expression while the remainder are evenly distributed between mild and severe distortions [11]. These annotations are used in place of an automated system to demonstrate the efficacy of the proposed adaptive weighting scheme.

4.1 Baseline Results

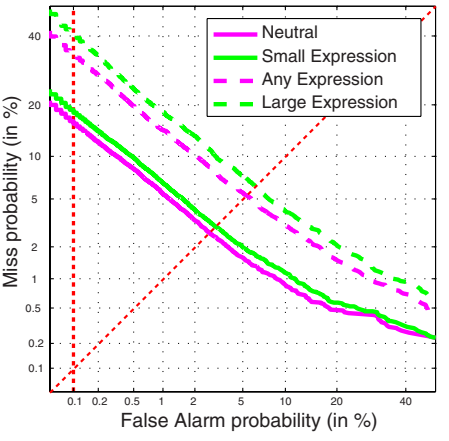
In order to demonstrate the advantage of the proposed system a baseline is required for comparison. In keeping with previously published results, the standard monolithic PCA algorithm is used to provide a benchmark against which other researchers can measure. In Figure 4 Detection Error Tradeoff (DET) curves are presented for the monolithic system and for the corresponding part face approach utilising unweighted summation. These results compare a gallery of neutral faces against progressively more expressive probe sets. As can easily be seen, the inclusion of highly expressive faces significantly degrades the performance of both the monolithic and part face methods. The effects of expression however, do not affect all regions evenly, and the parts face approach is ideally suited to provide a mapping of how expression mismatch manifests as performance degradation. To visualise this, the performance of each individual face region is calculated using neutral gallery and probe sets. The resulting EER is then compared against the EER obtained using the same region with a highly expressive probe set.

The performance differential can then be measured as change in EER and this is reformed in Figure 5 into a viewable image. This demonstrates the non-linear nature of degradation effects caused by expression variation. As was postulated earlier, regions in the cheeks and around the corners of the mouth are the most effected by expression variation. In these regions the EER drops an average of around 14% for expressive faces compared to neutral faces. The upper portion of the face appears to be much more stable, in particular regions which encompass any significant portion of the nose appear to have a significantly greater resilience to these effects.

These results validate the significant de-emphasis of the cheek regions in the chosen weight vectors. They also add credence to the position that the upper part of the face contains discriminable information both in human recognition



(a)



(b)

	Equal Error Rate (EER)		Accuracy at FAR=0.1%	
	Mono	Parts	Mono	Parts
Neutral	3.94%	2.67%	81.98%	84.19%
Small	4.20%	3.00%	79.50%	81.83%
Any	6.87%	5.32%	62.32%	66.52%
Large	8.23%	6.16%	53.87%	59.63%

(c)

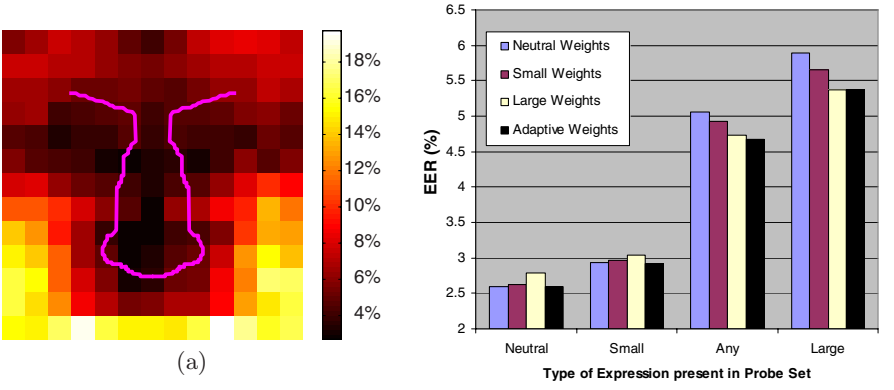
Fig. 4. Baseline DET curves for (a) Monolithic PCA and (b) Part face PCA with unweighted summation. Four curves show performance for four levels of expression containined in the probe set. (c) Tabulated results from (a) and (b) are presented for the two indicated operating points (dashed red lines).

[16] and in automated systems [17]. In humans it is plausible that this extra weighting has been given to the upper portions of the face due to their stability in the presence of expression variation, given that such variations are encountered so often in everyday life.

4.2 Adaptive Weighting

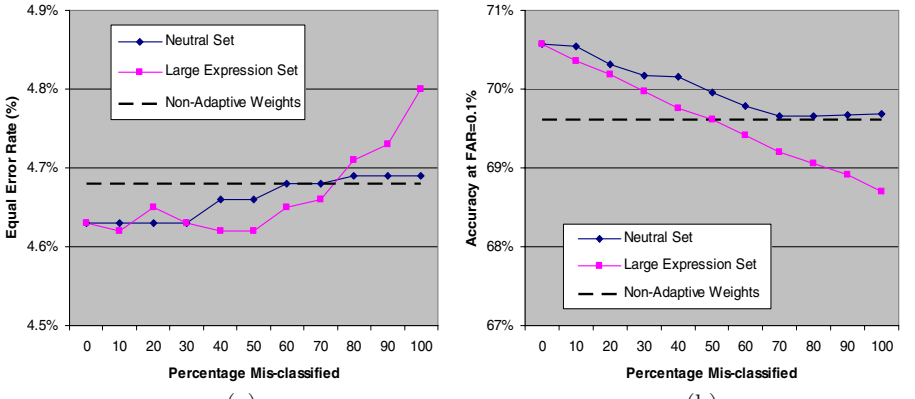
Using the weight vectors defined in Section 3, the regions around the face are adaptively combined using the hand labeled expression data. These results, shown in Figure 5, show that the proposed method achieves the best performance across all expression categories. In the case of highly expressive faces the EER is only 5.37% compared against 6.16% for unweighted summation and 8.23% for the monolithic system. The use of hand labelled data makes this a best-case scenario, however it serves to demonstrate the performance that can be gained when using a reliable expression categorisation system.

In practice no system works at 100% accuracy and errors in the expression classification stage should not cause catastrophic failure of the system. In order to demonstrate the effects of misclassification, the two most serious types of error



(a) (b)

Fig. 5. (a) Difference in Accuracy (EER) between Neutral and Expressive Probe Sets on region-by-region basis (nose outline added to provide spatial landmark to reader). (b) Comparison of three fixed weighting schemes to the proposed adaptive weighting scheme, results presented as Equal Error Rates.



(a) (b)

Fig. 6. Performance of adaptive weighting scheme as a function of introduced error in the indicated expression category. Results calculated using ROC III and expressed as (a) Equal Error Rate (b) Accuracy at a False Alarm Rate of 0.1%.

are considered: Neutral Expression misclassified as Large Expression and vice versa. Firstly system performance is evaluated while increasing the percentage of neutral images misclassified as expressive. The same process is then replicated to introduce errors to the set of images containing large expression. In this experiment performance is evaluated using the FRGC ROC III protocol, which incorporates time lapse of at least 1 semester between target and probe.

Results presented in Figure 6 show the EER and accuracy at a False Accept Rate (FAR) of 0.1% for both types of error. The non-adaptive weighting scheme,

shown as a dashed black line in these plots, is the baseline achieved when using the 'Large' weight vector. This weight vector is chosen as the baseline because it is the most robust to expression variation and therefore the most logical choice when no knowledge of facial expression is assumed.

As can be expected, the mistaken use of a neutral weight vector when an expressive face is present gives the greatest degradation to recognition accuracy. The alternate type of error, i.e. using the "Large" weighting mask for neutral images, has a less significant impact upon performance. The robustness of the proposed system can be observed with performance improvements in the presence of up to 50% error of either variety.

5 Conclusion

In this paper a novel adaptive weighting scheme has been proposed which increases the robustness of parts based face recognition. The proposed system makes use of expression strength information to increase or decrease the contribution of regions susceptible to expression distortion. Testing is conducted using the 3D face component of the Face Recognition Grand Challenge dataset which is currently the largest publicly available. The resilience of the system to the "front end effect" is evaluated and shows robust performance in the presence of up to 50% error in the expression classification stage.

Acknowledgement

This research was supported by the Australian Research Council (ARC) through Discovery Grants Scheme, Grant DP0452676 and Linkage Grants Scheme, Grant LP0562101.

References

1. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys (CSUR) 35(4), 399–458 (2003)
2. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the Face Recognition Grand Challenge. In: CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Washington, DC, USA, vol. 1, pp. 947–954. IEEE Computer Society Press, Los Alamitos (2005)
3. Bowyer, K.W., Chang, K., Flynn, P.: A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. Computer Vision and Image Understanding 101(1), 1–15 (2006)
4. Pantic, M., Rothkrantz, L.: Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE 91(9), 1370–1390 (2003)
5. Li, X., Mori, G., Zhang, H.: Expression-Invariant Face Recognition with Expression Classification. In: The 3rd Canadian Conference on Computer and Robot Vision (CRV'06), p. 77 (2006)

6. Chang, K.I., Bowyer, K., Flynn, P.: Adaptive rigid multi-region selection for handling expression variation in 3d face recognition. In: Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference, vol. 3, p. 157. IEEE Computer Society Press, Los Alamitos (2005)
7. Martinez, A.: Recognizing imprecisely localized, partially occluded, andexpression variant faces from a single sample per class. Pattern Analysis and Machine Intelligence, IEEE Transactions 24(6), 748–763 (2002)
8. Li, C., Barreto, A.: An integrated 3D face-expression recognition approach. In: Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference, vol. 3, pp. III-1132–III-1135. IEEE Computer Society Press, Los Alamitos (2006)
9. Lucey, S., Sridharan, S., Chandran, V.: Improved Facial-Feature Detection for AVSP via Unsupervised Clustering and Discriminant Analysis. EURASIP Journal on Applied Signal Processing 2003(3), 264–275 (2003)
10. Donato, G., Bartlett, M., Hager, J., Ekman, P., Sejnowski, T.: Classifying Facial Actions. Pattern Analysis and Machine Intelligence, IEEE Transactions 21(10), 974–989 (1999)
11. Maurer, T., Guigonis, D., Maslov, I., Pesenti, B., Tsaregorodtsev, A., West, D., Medioni, G.: Performance of Geometrix ActiveID^{TM}3D Face Recognition Engine on the FRGC Data. In: CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Washington, DC, USA, IEEE Computer Society Press, Los Alamitos (2005)
12. Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Recognizing facial expression: machine learning and application to spontaneous behavior. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005., vol. 2, pp. 568–573. IEEE Computer Society Press, Los Alamitos (2005)
13. Brunelli, R., Poggio, T.: Face Recognition: Features versus Templates. IEEE Trans. Pattern Anal. Mach. Intell. 15(10), 1042–1052 (1993)
14. Lucey, S., Chen, T.: Face recognition through mismatch driven representations of the face. In: Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop, pp. 193–199. IEEE Computer Society Press, Los Alamitos (2005)
15. Cook, J., Chandran, V., Fookes, C.: 3D Face Recognition using Log-Gabor Templates. British Machine Vision Conference (BMVC) (September 2006)
16. Shepherd, J., Davies, G., Ellis, H.: Studies of cue saliency. Perceiving and Remembering Faces, 105–131 (1981)
17. Gokberk, B., Akarun, L., Alpaydin, E.: Feature selection for pose invariant face recognition. In: Proceedings of the 16th International Conference on Pattern Recognition (2002)