

3D Face Recognition using Log-Gabor Templates

Jamie Cook, Vinod Chandran and Clinton Fookes
Image and Video Research Laboratory
Queensland University of Technology
GPO Box 2434, Brisbane Qld 4000, Australia
{j.cook, v.chandran, c.fookes}@qut.edu.au

Abstract

The use of Three Dimensional (3D) data allows new facial recognition algorithms to overcome factors such as pose and illumination variations which have plagued traditional 2D Face Recognition. In this paper a new method for providing insensitivity to expression variation in range images based on Log-Gabor Templates is presented. By decomposing a single image of a subject into 147 observations the reliance of the algorithm upon any particular part of the face is relaxed allowing high accuracy even in the presence of occlusions, distortions and facial expressions. Using the 3D database collected by University of Notre Dame for the Face Recognition Grand Challenge (FRGC), benchmarking results are presented showing superior performance of the proposed method. Comparisons showing the relative strength of the algorithm against two commercial and two academic 3D face recognition algorithms are also presented.

1 Introduction

Face as a biometric has the distinct advantage over other modalities such as fingerprint, DNA and iris recognition, in that the acquisition stage is non-intrusive and can be achieved with readily available equipment. 3D representations of the human face have the potential to overcome many of the obstacles such as pose and illumination sensitivity, which have prevented the widespread adoption of Face Recognition Technology (FRT).

Early work in 3D facial recognition emerged in the late 1980's and for the main part used surface curvature information. However, such approaches required high resolution facial scans obtained from laser range finders and typically only had a small number of test subjects [1]. Indeed until very recently most research in the field has only been evaluated on small customized databases, recently however the Face Recognition Grand Challenge program [2] has provided an extensive dataset for benchmarking of 3D face recognition algorithms.

In [1] the authors present a good summary of the current research in 3D and composite 2D-3D recognition, in particular they note that while it is accepted that a combination of 2D and 3D gives greater performance, it is still unclear which modality performs better in isolation. In this paper the focus is applied to 3D recognition with the knowledge that late-fusion of results with most 2D recognition algorithms will improve performance.

In general, approaches to 3D recognition fall into 3 main categories [1]: those that use 3D correspondence matching explicitly to provide discrimination [3, 4]; those that

extract 3D features such as curvature directly from the face; and those that treat the range image as a 2D image in order to extract features [5]. The latter has the advantage that a considerable number of well tested image processing algorithms can be directly applied.

Gabor filters are one such method which have been demonstrated to achieve high recognition rates in traditional 2D face recognition tasks [6, 7] and have been shown in [8] to exhibit robustness to misalignment. Techniques such as Hierarchical Graph Matching (HGM) and Elastic Bundle Graph Matching (EBGM) enhances this resilience further by adding a degree of freedom into the localisation of feature points [5].

In this paper a new method for achieving robust face matching called Log-Gabor Templates (LGT) is presented. It is established that the use of multiple observations improves biometric performance; LGT exploits this fact by breaking a single range image of a subject into multiple observations in both the spatial and frequency domains. These observations are each classified individually and then combined at the score level. This provides a distributed approach which is resilient to local distortions such as expression variation and minor occlusions such as from glasses, scarves and facial hair.

2 Gabor and Log-Gabor Filters

The Gabor family of wavelets first started gaining popularity in the field of image processing in 1980 when Daugmann first showed that the kernels exhibit many properties common to mammalian cortical simple cells. Properties such as spatial localisation, orientation selectivity and spatial frequency characterisation. They have enjoyed much attention in the field of 2D face recognition [6, 9] and associated fields as researchers attempt to emulate and surpass the face recognition capabilities of human beings.

The Gabor filter is composed of two main components, the complex sinusoidal carrier, $s(\mathbf{x})$, and the gaussian envelope, $w_r(\mathbf{x})$. In general N-dimensional terms these two components are combined as,

$$\begin{aligned} g(\mathbf{x}) &= s(\mathbf{x}) \cdot w_r(\mathbf{x}) \\ &= e^{2\pi j \mathbf{k}^T \mathbf{x}} \cdot K e^{-\pi \mathbf{y}^T \text{diag}(\alpha) \mathbf{y}} \\ &= K e^{2\pi j \mathbf{k}^T \mathbf{x} - \pi \mathbf{y}^T \text{diag}(\alpha) \mathbf{y}}, \end{aligned} \quad (1)$$

where $\mathbf{k} = [u_0 \ v_0 \ w_0 \ \dots]^T$ defines the frequency of the complex valued plane wave, $\alpha = [\sigma_1^2 \ \sigma_2^2 \ \dots \ \sigma_N^2]^T$ is the variance of the gaussian and \mathbf{y} is defined as $R_\theta(\mathbf{x} - \mathbf{x}_0)$. \mathbf{x}_0 is the location of the peak of the gaussian and R_θ is a rotation of the gaussian about this peak.

In many approaches these two fundamental aspects are also combined with a DC compensation component which prevents the filters from having a DC response [6].

2.1 Log-Gabor Filters

In [10] Field proposes an alternate method to perform both the DC compensation and to overcome the bandwidth limitation of a traditional Gabor filter. The Log-Gabor filter has a response that is Gaussian when viewed on a logarithmic frequency scale instead of a linear one. This allows more information to be captured in the high frequency areas and also has desirable high pass characteristics. Field [10] defines the frequency response of

a Log-Gabor filter as,

$$G(\mathbf{f}) = \exp - \frac{\log(\mathbf{f}/\mathbf{k})}{2\log(\sigma/\mathbf{k})}, \quad (2)$$

where $\mathbf{k} = [u_0 \ v_0 \ w_0 \ \dots]^T$ is once again the centre frequency of the sinusoid and σ is a scaling factor of the bandwidth. In order to maintain constant shape ratio filters, the ratio of σ/\mathbf{k} should be maintained constant. It is worth noting that a Log-Gabor filter with a 3 octave bandwidth has the same spatial width as a 1 octave Gabor filter, demonstrating the ability of the filters to capture broad spectral information with a compact spatial filter.

In order to cover the frequency spectrum effectively, a range of both scales and orientations of the Gabor filters must be considered. The overall aim is to provide an even coverage of the frequency components of interest while maintaining a minimum of overlap between filters so as to achieve a measure of independence between the extracted co-efficients. For the following experiments the shape parameter, σ/\mathbf{k} , was chosen such that each filter had a bandwidth of approximately 2 octaves and the filter bank was constructed with a total of 6 orientations and 3 scales.

3 Face Verification

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 [11, 12] have shown superior performance by adopting a modular representation of the face provided that face localisation is performed accurately [12]. The LGT method improves upon previous approaches by using Log-Gabor filter responses to reduce the sensitivity to pose variation [8]. In the following section an array of classifiers are constructed from sub-regions in both the spatial and frequency domains, the combination of which outperforms a single classifier using the entire face.

3.1 Log-Gabor Templates

It is well established that using multiple probe images aids recognition performance, the same effect can be obtained by breaking a single face into multiple observations. After application of the 18 Log-Gabor filters, the face is broken into 49 square windows arranged in a 7x7 grid with 50% overlap in both the horizontal and vertical directions. These regions are then further decomposed by 3 scales of filter to generate 147 subregions which are considered individually at the dimensionality reduction stage; an overview of the process can be seen in Figure 1.

Principal Component Analysis (PCA) is then applied to the Log-Gabor filter responses in each of the 147 subregions. In each region only the top 100 eigen-vectors in the feature subspace were retained, thus each face is finally represented as 147 feature vectors each comprising 100 dimensions.

3.2 Distance Measure

The original Eigenfaces approach of Turk and Pentland used a simple Euclidean Based classifier, however, experimentation with a wide variety of distance measures has shown

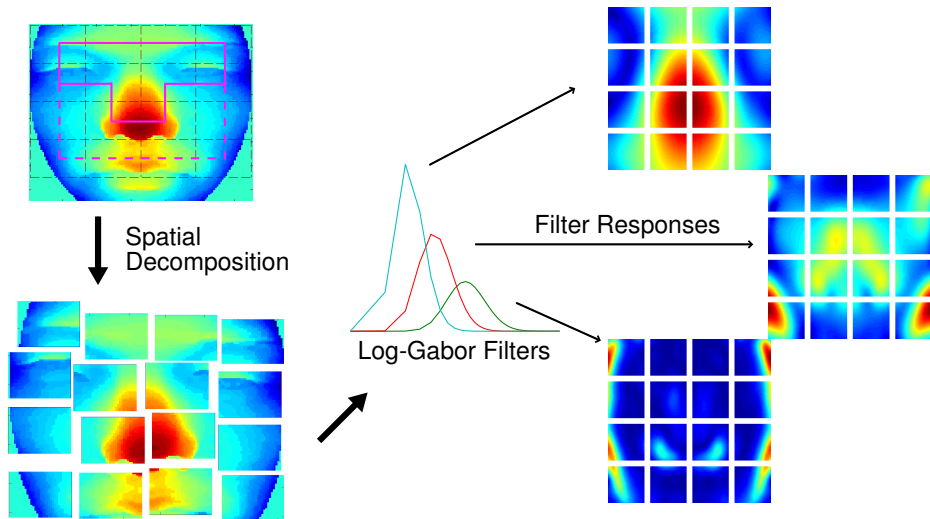


Figure 1: Decomposition of the face into subregions

that the Mahalanobis Cosine distance measure provides better performance [13]. This measure is defined as,

$$\begin{aligned}
 D_{MahCosine} &= -\frac{|m| |n| \cos \theta_{mn}}{|m| |n|} \\
 &= -\frac{m \cdot n}{|m| |n|},
 \end{aligned} \tag{3}$$

where m and n are two feature vectors transformed into the Mahalanobis space, this is subtly different from the Eigen-space transformation as it also involves a whitening stage.

3.3 Classifier Fusion

There have been a large variety of methods proposed for combining the outputs of multiple classifiers. They range from simple voting schemes and weighted summation to more complex Support Vector Machines and Neural Networks. However it has been shown that combination of multiple classifiers using unweighted summation has a negligible performance degradation over more intricate schemes and has the benefit of not requiring a separate tuning phase [12]. In this paper the process of score fusion refers to this method of combining classifiers.

4 Dataset Description

The Face Recognition Grand Challenge (FRGC) [2] was created to push the state of the art in face recognition and to provide a common dataset for benchmarking of algorithms. The experiments described in this article were conducted using 3D data provided as part of the Challenge.

The 3D dataset provided with the FRGC, 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 which uses a Structured Light Scanning (SLS) technique to capture shape information.

The 466 subjects in the database were broken into training and testing groups according to the specification of FRGC Experiment 3, however only the range images are used for the reasons stated in Section 1. Within Experiment 3 there are 3 sub-experiments of increasing difficulty, all results quoted in this paper were evaluated on the hardest of these (Mask III) which is comprised of target/query pairs which are captured in different semesters. Unless otherwise stated all results are quoted as true acceptance rates at a False Acceptance Rate (FAR) of 0.1%.

Of the 4007 images 59% are captured with a neutral expression while the remainder are captured variously with expressions of surprise, happiness, sadness and disgust. Manual classification by researchers at Geometrix [4] shows that these non-neutral images are evenly distributed between mild and severe distortions. Examples of range images under various expressions are shown in Figure 2. The FRGC specifications of Experiment 3 also

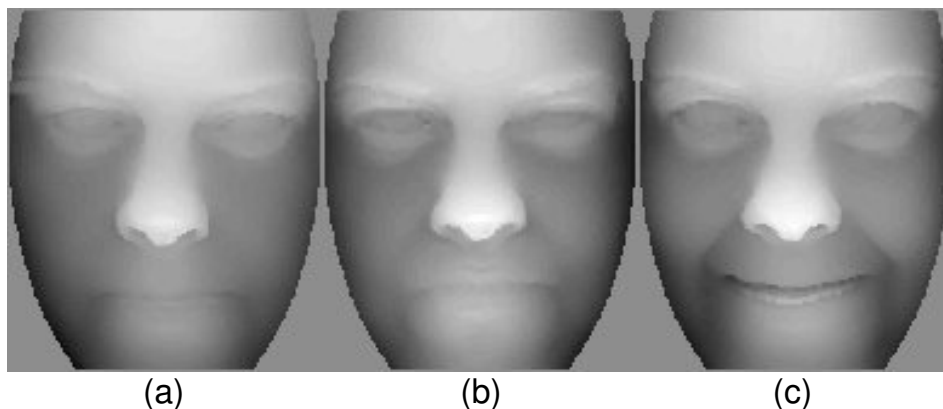


Figure 2: Examples of FRGC range images with (a) neutral expression (b) slight distortion and (c) large distortion

utilize a training set of 943 images which is used in the training of Eigen-spaces for the various subregions considered in the following experiments.

5 Experimentation

To test the hypothesis that a parts based representation is capable of providing superior verification accuracy over a monolithic approach, the performance shall be evaluated in both the spatial and frequency domains separately and then in a combination of the two. These shall be compared against monolithic representations constructed from both the range image and its Log-Gabor response.

5.1 Monolithic

When constructing the monolithic Log-Gabor baseline it was found that the feature fusion of all Log-Gabor responses across the entire face provided feature vectors of sufficient length that physical memory became an issue. Thus various rates of pixel decimation were applied to the vectors before input to PCA to determine the effects of feature vector length on the Equal Error Rate (EER), results of which are shown in Figure 3a. A decimation level of 1:10 was found to provide negligible performance degradation and yields a feature vector of much more tractable length and was thus used to construct the Log-Gabor baseline.

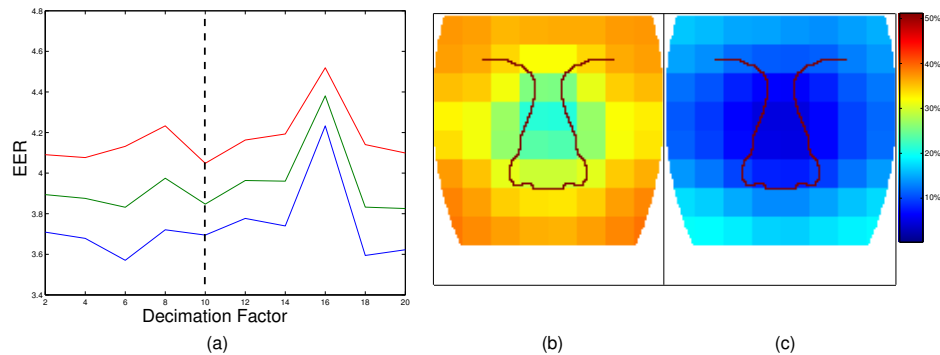


Figure 3: Equal Error Rate as a function of: (a) feature space decimation (b) spatial location in range image (c) spatial location in Log-Gabor transformation

5.2 Log-Gabor Decomposition

After dividing the face into 49 overlapping regions and calculating recognition performance in isolation, a picture emerges of the distribution of discriminatory information. In Figures 3b and 3c it can be seen that the EER deteriorate with distance from the image center for both range and Log-Gabor based representations, however this degradation is more abrupt in regions below the nose. Best individual performance is observed around the upper nose and interestingly, both plots show that higher recognition rates appear to follow the cheek bones across the face.

After consideration of the centralised nature of this distribution three sets of regions were considered for score level fusion: All regions, central regions and the nose/eye regions (which are generally held to be more invariant to expression variation). The bounding boxes can be seen overlaid on the face in Figure 1 and results from each are given in Table 1.

Firstly it is important to note that across all representations, accuracy appears to be adversely affected by the inclusion of the pixels from outer areas. This is expected behaviour as these pixels generally comprise information from the mouth (very deformable) and the forehead and cheeks (very bland).

These results also show that use of spatial decomposition significantly boosts the performance of monolithic representations. This is especially true when considering the

entire face, thus illustrating the power of this method to decrease the impact of non-discriminatory information, such as that provided by outer regions. In the larger regions we also note that adding overlapping blocks improves performance, despite the significant degree of redundant information provided by these regions.

Each of these regions are also broken in subregions based on the scale of the Log-Gabor filters, the right hand side of Table 1 shows the performance of individual subbands across the specified regions. As can be seen the combination of classifier scores from frequency bands again provides better performance than a single representation, however the improvement is not as drastic as was seen with spatial decomposition. This can be attributed to the smaller number of classifiers, the presence of shared information in bands caused by overlap of the filters and an inability to deemphasise information from poor performing spatial regions.

| | PCA (Range) | PCA (Log-Gabor) | Spatial Fusion | Frequency (cycles/pixel) | | | Freq. Fusion |
|----------------------------------|----------------|--------------------|-------------------|--------------------------|-------------|-------------|-----------------|
| | | | | 0.023 | 0.047 | 0.094 | |
| Nose/Eye Regions with overlap | 38.90% - | 78.83% - | 87.29% 86.21% | 68.94% - | 73.68% - | 64.99% - | 83.59% - |
| Center Regions with overlap | 60.25% - | 77.57% - | 89.75% 91.67% | 60.23% - | 72.86% - | 66.39% - | 79.40% - |
| All Regions With overlap | 51.44% - | 42.16% - | 83.21% 87.27% | 47.93% - | 34.90% - | 40.88% - | 49.76% - |

Table 1: Recognition rates for experiment baselines and decomposition in both spatial and frequency domain.

These results show that decomposition of the face and simple score fusion in both the spatial and frequency domains provides better accuracy than a single classifier. These concepts are then combined to give a full decomposition of the face in both domains. Due to computational constraints only the central regions are considered, however in both cases this further decomposition again increased the performance of the overall classifier. The nose/eye region recognition rate improved to 87.84% when using 4 non-overlapping sub-regions across 3 frequency bands. A best recognition rate of 92.01% was achieved by using 3 bands in each of the central 25 overlapping sub-regions for a total of 75 of the 147 available classifiers.

5.3 Expression Variation

In [4], the authors manually divide the FRGC 3D corpus into three categories based on strength of expression variation which they provide as an appendix to their publication. In order to test the robustness of the presented approach in the presence of expression variation the neutral images are used as the gallery and compared against each of the three classes. The Detection-Error Tradeoff for each comparison is shown in Figure 4. Comparisons with neutral probes gave a recognition rate of 98.25% and even when comparing the gallery against a probe set comprising severe expression variation the LGT approach achieved a rate of 90.75%. Compare this to traditional Eigenfaces which drops from 75% recognition to 54% for the same comparisons. This demonstrates the ability of the LGT method to provide resilience to and graceful performance degradation in the presence of expression variation.

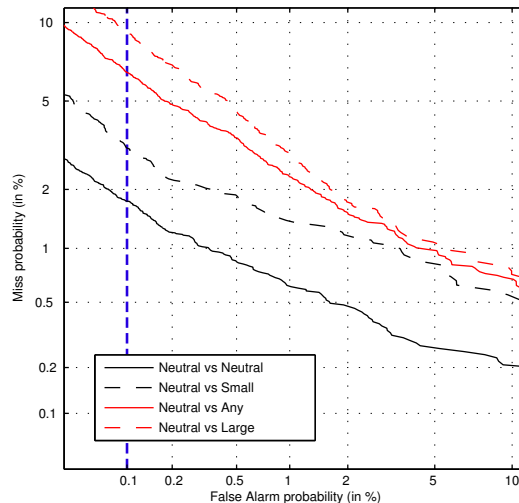


Figure 4: Detection Error Tradeoff for varying degrees of expression

6 Comparison to other methods

In order to provide a comparison with other methods, additional experiments were run to provide results directly comparable to other researchers using the FRGC dataset. Four research groups were identified in [1] as using the second version of the FRGC corpus; a summary of results from the most recent publications from these groups are given in Table 2. All figures quoted are verification rates at a FAR of 0.1% with the exception of last row which shows Rank 1 recognition, where numerical values are not given in publications results have been estimated from plots and rounded *up* to the nearest 1%.

| | Maurer [4] | | Husken [5] | | Passalis [14] | Chang [3] | LGT |
|-----------------|------------|-------|------------|-------|---------------|--------------|------------------------|
| | 3D | 2D-3D | 3D | 2D-3D | | | |
| All vs. All | 87% | 93.5% | - | - | - | | 92.31% |
| Neutral vs. All | 92% | 95.8% | - | - | - | | 95.81% |
| ROC I | - | - | 92% | - | 89% | | 93.71% |
| ROC II | - | - | - | - | 88% | | 92.91% |
| ROC III | - | - | 89.5% | 97.3% | 86% | | 92.01% |
| Rank One | | | | | 89.5% | 92.3% | 92.93% / 94.63% |

Table 2: Comparison of LGT to other methods tested on FRGC dataset (Italicised entries are estimated from tables and bold entries are from a gallery of neutral expression)

Maurer [4] use their manual categorization of images to generate recognition rates for 4 combinations of gallery/probe images using a commercial ICP based recognition engine. In the two most difficult combinations they consider the case of a probe set containing the full set of expressions compared against gallery sets with and without expression variation. Results for LGT, which use only shape information, compare favourably with

their combined 2D-3D recogniser and easily outperform the shape based classifier.

Passalis [14] use an Annotated Deformable Model to conduct experiments with respect to expression variation. Their approach appears to work well even when the gallery set comprises large expression changes. The exact gallery/probe sets used for these experiments are not published and so results from the standard three sub-experiments are listed instead. On average verification rates using LGT have a 3-5% absolute improvement over the UR3D method of [14].

In [3], the authors combine multiple sub-regions around the nose region using the product rule to improve the robustness of an Iterative Closest Point (ICP) based classifier. They report a rank one rate of 91.9% which rises to 92.3% when manual landmark identification is used, this compares to a rate of 94.63% using LGT. An important distinction to note between the rank one recognition rates of Chang and Passalis is that Chang uses a gallery set containing **only** neutral expressions whereas Passalis simply uses the first session acquired for each subject. The LGT algorithm was tested under both conditions and achieves superior performance in both cases (the rates for a neutral gallery are shown in bold in the above table).

Finally, in [5] Husken presents a commercial recognition system which utilises the HGM approach mentioned earlier on both the shape and texture channels. The results presented are exceptional and no evidence can be found to refute the authors claims that their system outperforms all others previously tested on the FRGC 3D dataset in either modality or their combination. Nevertheless, recognition based on LGT provides a 23% relative improvement at an FAR of 0.1% when considering only shape information.

7 Conclusions

In this paper a novel and robust 3D face recognition algorithm is presented. The Log-Gabor Templates method exploits the multitude of information available in the human visage to construct multiple observations of a subject which are classified independently and combined with score fusion. Analysis of the spatial distribution of discriminable information has shown that best results are achieved by using only the central regions of the face and in particular the nose and eye regions.

The proposed method has been evaluated on the largest publicly available 3D face database. Results have shown that the parts based methodology adopted reduces the error rate by over 60% over an equivalent monolithic approach (at a FAR of 0.1%). Comparisons of the LGT method with several leading 3D recognition engines have also been presented and the enhanced resilience to expression variation offered by the use of LGT has been demonstrated.

Acknowledgements

This research was supported by the Australian Research Council (ARC) through Discovery Grant Scheme, Project ID DP452676, 2004-6.

References

- [1] Kevin W. Bowyer, Kyong Chang, and Patrick 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, TY - JOUR.
- [2] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek, "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) - Volume 1*, Washington, DC, USA, 2005, pp. 947–954, IEEE Computer Society.
- [3] Kyong I. Chang, K.W. Bowyer, and P.J. Flynn, "Adaptive rigid multi-region selection for handling expression variation in 3d face recognition," in *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, 2005, vol. 3, p. 157.
- [4] Thomas Maurer, David Guigonis, Igor Maslov, Bastien Pesenti, Alexei Tsaregorodtsev, David West, and Gerard Medioni, "Performance of Geometrix ActiveID™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, 2005, IEEE Computer Society.
- [5] Michael Husken, Michael Brauckmann, Stefan Gehlen, and Christoph Von der Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," in *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, Washington, DC, USA, 2005, p. 174, IEEE Computer Society.
- [6] B. Duc, S. Fischer, and J. Bigun, "Face Authentication with gabor Information on Deformable Graphs," *IEEE Transactions on Image Processing*, vol. 8, no. 4, pp. 504–516, April 1999.
- [7] R. Chellapa, C. L. Wilson, S.Sirohey, and C. S. Barnes, "Human and Machine Recognition of Faces: A Survey," Tech. Rep., University of Maryland and NIST, August 1994.
- [8] Shiguang Shan, Wen Gao, Yizheng Chang, Bo Cao, and Pang Yang, "Review the strength of gabor features for face recognition from the angle of its robustness to mis-alignment.," in *ICPR (1)*, 2004, pp. 338–341.
- [9] Ki-Chung Chung; Seok Cheol Kee; Sang Ryong Kim, "Face recognition using principal component analysis of gabor filter responses," in *Proceedings of Intl. Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, September 1999, pp. 53–57.
- [10] D. Fields, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of Optical Society of America*, vol. 4, no. 12, pp. 2379–2394, 1987.
- [11] Roberto Brunelli and Tomaso Poggio, "Face Recognition: Features versus Templates.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, 1993.
- [12] S. Lucey and Tsuhan Chen, "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 on*, 2005, pp. 193–199.
- [13] M. Teixeira D. Bolme, R. Beveridge and B. Draper, "The CSU Face Identification Evaluation System: Its Purpose, Features and Structure," in *International Conference on Vision Systems*. 2003, pp. 304–311, Springer-Verlag.
- [14] G. Passalis, I.A. Kakadiaris, T. Theoharis, G. Toderici, and N. Murtuza, "Evaluation of 3D Face Recognition in the presence of facial expressions: an Annotated Deformable Model approach," in *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, 2005, vol. 3, p. 171.