Removing Pose from Face Images

Seán Begley John Mallon Paul F. Whelan

Vision Systems Group
Dublin City University
Dublin 9, Ireland

Abstract. This paper proposes a novel approach to pose removal from face images based on the inherent symmetry that is present in faces. In order for face recognition systems and expression classification systems to operate optimally, subjects must look directly into the camera. The removal of pose from face images after their capture removes this restriction. To obtain a pose-removed face image, the frequency components at each position of the face image, obtained through a wavelet transformation, are examined. A cost function based on the symmetry of this wavelet transformed face image is minimized to achieve pose removal. Experimental results are presented that demonstrate that the proposed algorithm improves upon existing techniques in the literature.

1 Introduction

Biometrics continue to grow in their common everyday usage. Examples include fingerprint scanners on laptops to secure access, face recognition in security systems, and forensic dental analysis for identification. The increased prominence of biometrics may be attributed to the need for better security measures, and to the advancement in the technology that is used. Faces in particular provide an attractive biometric feature that may be used for person verification or recognition. Faces exhibit large variations from person to person which makes computer recognition of faces an attractive approach [1]. Another motivation for using faces is that the techniques involved are non-intrusive and user friendly. Other approaches require the user to have a significant role in the process. However, in order for face recognition systems to operate optimally, subjects must maintain a neutral expression and look directly into the camera. Evidence of this is found in the literature where only images portraying fronto-parallel neutral faces are used in experimental verification. Indeed, it is noted in [2] that "pose variations degrade the recognition rate of all the three investigated methods." Variations in illumination, pose, and expression are not well tolerated and so must be removed. This becomes an issue for the user, as slight variations in pose are natural, and neutral expressions are difficult to portray. Efforts have to be made by the system creators to remove expression and pose, and allow the user to be imaged in near frontal positions with any expression. This will allow greater accessability and ease of use for the subjects. Expression classification and removal has been carried out using similar techniques to those found in face recognition systems [3]. It is pose removal that will be considered in this paper.

Since faces are inherently symmetrical, we propose to use symmetry to restore the pose of faces back to fronto-parallel. This paper presents a novel approach to pose removal from face images. Each image is rectified by calculating a symmetry based cost function on each view of a model texture mapped with the face image. The cost function is minimized in an iterative optimisation process.

This paper is organized as follows. Section 2 gives a brief critique of current methods in the literature including highlighting aspects of the techniques that restrict their application. Section 3 describes the proposed approach to pose removal from faces. Section 4 describes the proposed cost function to be used and highlights its advantages over other similar cost functions. Experimentation is demonstrated in Section 5 with both quantitative and subjective results shown. Finally, Section 6 gives a conclusion and direction for future work.

2 Background

Much research is being conducted to improve recognition rates under varying pose, illumination and expression conditions. Varying pose in particular significantly decreases recognition and verification rates [2]. Many approaches can be applied to improve facial recognition, the majority of which are categorized into three major streams; multiple appearance techniques; 3D-model based techniques, and generic appearance based methods. The latter two streams aim to remove pose, while the first attempts to operate on the posed face images. Recognition is typically carried out on frontally imaged faces, however, with multiple appearance techniques, recognition is carried out between posed images and other similarly posed images. 3D model based algorithms require the construction of explicit 3D models for each face through a learning process within which multiple views of each subject are utilized. Generic techniques on the other hand, do not create specific models and typically remove pose in feature space.

A middle-ground between 3D-model based methods and generic methods, multiple appearance based techniques use multiple views of subjects to create multiple recognition spaces, one for each view. These subspaces are created by selecting face images exhibiting the same pose and performing dimensionality reduction on this subset of images. Different subspaces are created for each pose onto which test images may be projected for vector discrimination [4]. Typically, the system user determines the pose that best matches the presented face and chooses the most relevant subspace to use for recognition.

As an alternative to learning individual subspaces for each view, face pose removal is more desirable. Pose removal is generally less restrictive, since a continuous range of poses may be handled. The majority of the research in the field attempts to remove facial pose as a pre-requisite to face recognition. As such, face pose removal algorithms remove non-frontal view information in the feature space where subject discrimination occurs and not directly on the image itself.

3D-model based techniques tend to be restrictive in that multiple views of each subject are generally required for the systems to be trained. In [5], 3 views at known positions of each subject are used to construct a 2D+3D Active Appearance Model. This is a cumbersome and time consuming approach and yields classification rates lower than using frontally imaged faces. Similarly in [6], generic 3D face models are used which are morphed based on the locations of features detected from faces imaged in arbitrary positions. A drawback of these approaches, is that multiple images are required to build the face models.

Contrary to these two approaches where multiple views are required, in [7] and [8] only a single view of each face is required to build the 3D face model. In [7], features are extracted from each frontally imaged face, the locations of which are used to linearly weight the shape vectors that describe the 3D face model. The obtained face model is texture mapped with the imaged face and may be re-rendered in a different view. The obvious restriction of this method is that an initial fronto-parallel image of each subject is required, which is generally not available. In [8], a learned 3D face model is morphed in an optimisation process so that the distance between projected 3D points and the corresponding 2D points is minimised in an iterative process. There are two significant hindrances to using this method. Firstly, some initial manual feature selection is required to start the iteration process. And secondly, due to the high dimension of the parameter space that defines the shape, illumination and color of the face model to be morphed, the algorithm takes a very long time to execute.

Generic appearance based techniques do not explicitly create a 3D model with which to render new views of the face. Instead, relations between views are derived in the feature space within which subject discrimination between is carried out. For example, in [9] a linear transformation between the basis functions of face images in one view and the basis function representation of the frontally imaged faces is calculated. This allows them to project the basis functions of a presented test image into its fronto-parallel view using the learned transformation matrix. This method is however restrictive in that for each viewing angle, a separate transformation matrix has to be calculated at the training stage, and also the pose of the individual has to be known to the user prior to pose removal.

Similar methods exist that operate on the positions of facial features and are the most prominently researched techniques in this field. Active shape models (ASM) use statistical learning methods to build up a reduced dimension subspace of the facial feature vectors [10, 11]. Faces portraying varying expressions and in different poses are used to train the system. The subspaces onto which the feature vectors are projected are spanned by vectors that control pose, expression and inter-person variation. Removing components that control pose and expression leaves the user with feature vectors that only contain identification information.

The two appearance based techniques mentioned above use the assumption that the pose components are entirely or almost entirely independent of the inter-person variation vectors. This however is not the case. So through the removal of the primary pose components, some of the essential variations that distinguish subjects' identities may also be removed. This is due to the pose and identity components being weakly correlated.

3D model based approaches provide the most realistic pose removal systems since once the model is known any view of the face my be synthesized by rotating the three dimensional model and re-rendering the image with the new view. There are however some significant drawbacks to the algorithms employed. An issue that occurs with both the generic 3D model and the learned 3D model is that of registering the image and the model. Accurate detection of feature points is vital, a feat that is difficult to achieve. Inaccurate texture mapping of the model could have a significant impact on subsequent expression classification and person recognition processes.

For these reasons, a new approach to face pose removal is proposed. Each face will be examined individually and directly operated on with no training database required. Also, no feature extraction and correspondence matching is required. This will reduce many of the restrictions that are currently associated with the existing techniques, and remove sources of error.

3 Proposed Pose Removal Algorithm

Pose removal will be carried out using a technique where the face image will be re-rendered in a new view based on a measure of the symmetry. Three different techniques for generating the new view to be rendered are considered. The face images will be texture-mapped onto a planar surface, a cylinder, and an ellipsoid respectively to form the basis for the three different view creation techniques. A cost function based on a symmetry measure is optimized to determine the optimum view to remove the pose variation.

The algorithm operates as follows. The face is segmented from the background of the image and the face region is cropped. The cropped face image is then texture mapped to one of the three surface types. In the case of the cylinder and ellipsoid techniques, the cropped image is texture mapped to one half of the surface with the other half being left blank. The texture-mapped surface is then rotated to a new viewing position and the projection of the surface onto the image plane is taken. The continuous wavelet transform of the image is taken to obtain the frequency response of the new view at each position in the image. At this point, the cost function based on the symmetry of the re-rendered view is calculated. A non-linear optimisation algorithm is employed to minimize the symmetry cost of the wavelet transformed views and obtain the rectified image. The algorithm is outlined in Fig. 1.

4 Symmetry Measure

Faces exhibit strong symmetry which can easily be identified by a human observer. Computers on the other hand, require a strong description of what symmetry is. Although symmetry is typically considered as a true or false metric, there are levels of "how symmetric" an object appears to be. This notion was first proposed in [12] where the minimum mean squared distance that feature-points need to be moved to ensure symmetry, was used as the continuous symmetry measure. However, that symmetry measure was initially designed to measure the symmetry of molecules and therefore is only suited to strong geometric shapes.

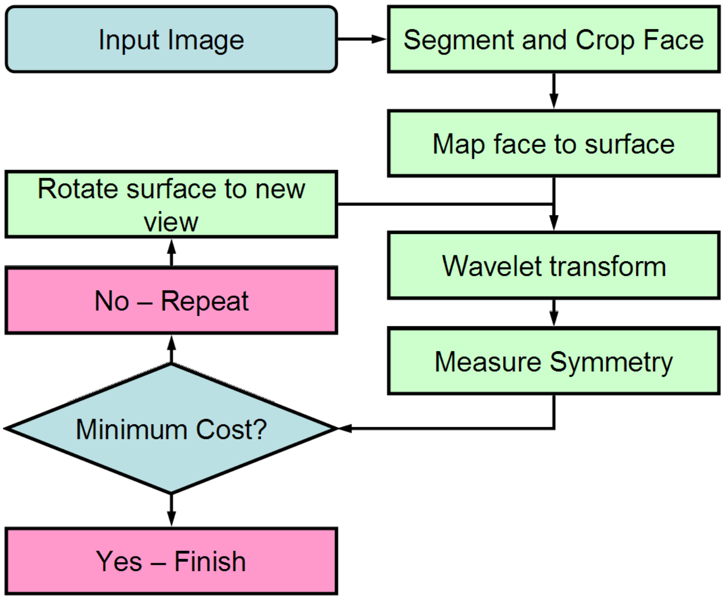

Fig. 1: Flow Diagram representation of Pose removal algorithm

A more suitable symmetry measure is proposed in [13] where a ratio of the symmetric and antisymmetric components of a function may be used to obtain a continuous symmetry measure. Elementary algebra tells us that functions are even if $f(x) = f(-x)$ and are odd if $f(x) = -f(-x)$, which are directly equivalent to symmetric and antisymmetric representations respectively. It is the proportion of the overall signal that is symmetric that determines the overall measure of symmetry. Each image is decomposed into its symmetric and antisymmetric components about the y axis as follows:

$$f_{sym}(x,y) = (f(x,y) + f(-x,y))/2 \tag{1}$$
$$f_{asym}(x,y) = (f(x,y) - f(-x,y))/2 \tag{2}$$

The continuous symmetry measure about the vertical y axis is calculated as:

$$S\{f(x,y)\} = \frac{\|f_{sym}\|^2}{\|f_{sym}\|^2 + \|f_{asym}\|^2} \tag{3}$$

This symmetry measure will achieve its maximum value of 1 when the image in question is perfectly symmetric and will achieve its minimum value of 0 when the image is perfectly asymmetric about the vertical axis. However, due to the nature of the optimisation algorithm that is employed, the level of asymmetry in the image will be minimized which equates to maximising the symmetry. The function for calculating the antisymmetric coefficient uses the antisymmetric component of the image in the numerator of Eqn. 3.

To achieve optimum pose normalisation performance, the cost function being used should form a smooth continuous curve with as few, and preferably zero, local minima. This should also be true in the case where minor levels of antisymmetry such as image noise, minor distortions and illumination variations are present. Because the symmetry cost function of [13] operates directly on the grayscale pixel values of the image, which we call dense-matching, the symmetry coefficient is slightly prone to noise. One solution to this drawback is to use Gaussian filtered images in the cost function, but these can not overcome lighting variations. An alternative approach is required, where the frequency response at each position in the image is examined. This allows higher frequency noise components to be neglected from the estimation and also lighting imbalances are removed as it is the frequency content that is being examined. These criteria can be achieved with the use of a wavelet transformation of the image space.

The wavelet transformation of a one-dimensional function is given as:

$$W(a,b) = \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{|a|}} \psi^* \left(\frac{x-b}{a} \right) dx \tag{4}$$

where $W(a,b)$ is the wavelet decomposition of the function $f(x)$ at position b using the mother wavelet ψ scaled in width by parameter a. An infinite range of values for a exists, but in practice only a very small set of values are chosen. A similar expression may be used to describe the wavelet decomposition of an image using a two dimensional wavelet decomposition in which case 3 parameters are used, one for scale and two for position. It is possible to have two different scales for the mother wavelet in the x and y directions separately for two dimensional decompositions, but in practice this is not done.

To demonstrate the advantages of using the wavelet transformed image over directly using the raw grayscale values or Gaussian filtered values, a simple experiment was conducted. A fronto parallel image was selected, noise added to it, and mapped to a plane. This plane was then synthetically rotated about its vertical axis. The symmetry was measured in each position using the proposed wavelet based cost, the dense-matching cost, and the gaussian-filtered image cost. Fig. 2 shows the results of the experiment which are scaled to have the same mean for display purposes. The cost function based directly on the grayscale values exhibits two local minima, neither of which is correct. The gaussian filtered image achieved its minimum at -5 degrees while the wavelet transformed data cost was the only method to achieve its minimum at the correct location. For these reasons, the wavelet transformed image cost function will be utilized in the experiments, with the other two methods used for comparison.

5 Experiments

In order to validate the proposed algorithm, a database of imaged subjects portraying a multitude of expressions in various poses was captured. The database contained 5 subjects each portraying 5 different expressions in 20 orientations

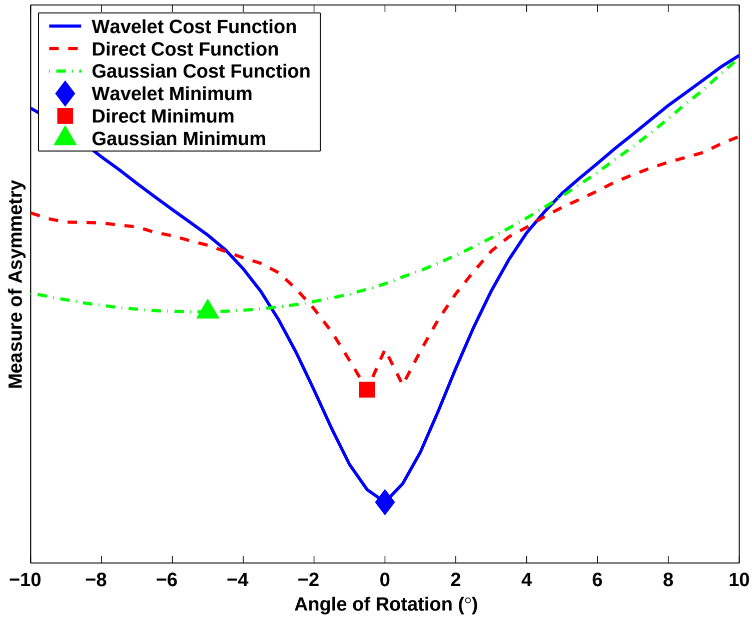

Fig. 2: Symmetry-cost measured on three different input spaces as the image plane is rotated about its yaw axis. The costs are scaled for display purposes.

relative to the camera, 500 images in total. Each subject was required to sit maintaining their expression while a computerized rotation stage and camera stepped through rotations from -10 to +10 degrees about the vertical axis.

Three experiments were carried out, each of which were compared directly to the dense matching and gaussian dense matching methods. The first experiment used a planar approximation to the face images in the rectification process. In the second and third experiments, the face images were mapped to a cylinder and an ellipsoid respectively. In each case, the algorithm was run on each image and for the second and third experiments the angle of rotation of the mapped shape was recorded. No rotation angle for the planar approximation method could be derived, since the camera matrix was unknown. In each experiment, the comparison methods were also employed, and it should be noted that with the coarse to fine gaussian filtered approach a number of iterations of the algorithm were required, with less gaussian blurring on each successive iteration. Examples of the rectified images obtained using the proposed algorithm are shown in Fig. 3

5.1 Experiment 1 - Planar Approximation

Initially, each image is processed to segment and crop the face from the background. At this point, the face is treated as a plane with planar transformations being applied for the remainder of the experiment. Since the motion of the subjects is restricted to rotations about the yaw axis, solutions are restricted to homographies that deviate from I_{3x3} in only the x perspective parameter. The symmetry cost function is minimized in a non-linear least squares optimisation scheme. Each time the algorithm converges to a solution, the resulting image is stored. Due to a lack of priori camera information the rotation angle can not be extracted from the retrieved perspective transformation matrix. only subjective results are presented in column four of Fig. 3.

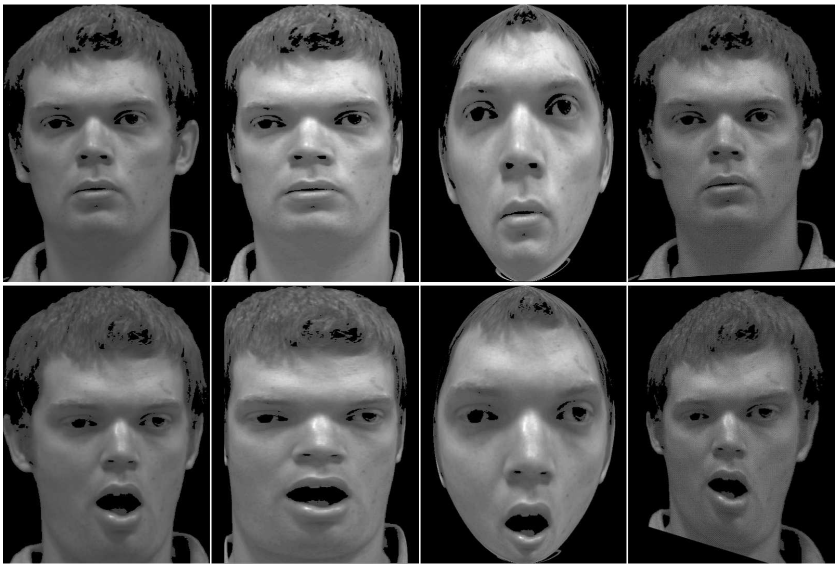

Fig. 3: Rectifying face images. Column 1 shows the cropped images. Column 2 displays images rectified with the cylinder mapping. And finally, Column 3 displays rectified faces using the ellipsoid texture-mapping. Column 4 displays planar rectified images.

5.2 Experiment 2 - Cylinder Mapping

Each face is segmented and cropped from the images and mapped to one half of a cylinder, the other half is left blank. The cylinder is rotated about it's vertical axis and the projection of the mapped cylinder onto the image plane is then used to compute the cost functions. An iterative optimisation scheme is implemented using the symmetry cost to obtain a rectification with each of the algorithms. The coarse to fine approach is run 5 times, with progressively less gaussian blur at each stage and the previous result being used to initialize the next phase. For each of the algorithms used, the angle of rotation is recorded for each result obtained. There are 25 images of subjects in each orientation that are rectified, from which each error is computed. The error at each orientation is computed as the mean absolute error of the 25 images. The results are presented in Fig. 4

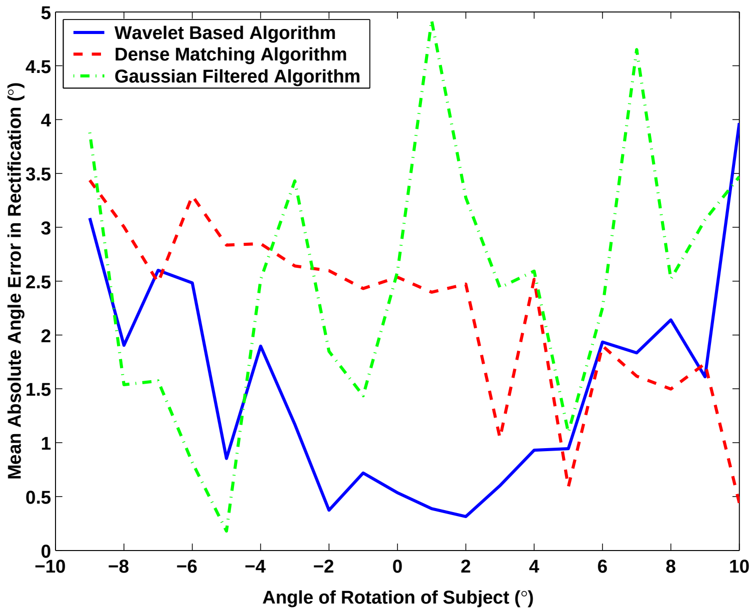

Fig. 4: Mean absolute angle error in rectifying faces with the texture-mapped cylinder

From Fig. 4, it can be seen that the proposed algorithm achieves better rectification results than the two comparison methods for rotations of the subject between −4 and +4 degrees. Also, the errors obtained with the proposed wavelet rectification algorithm increase more smoothly with increasing rotation angle than the comparison methods. A bias in both the dense matching and gaussian blurred algorithms is visible in Fig. 4 resulting in a perceived improved performance over the proposed algorithm, however, the proposed algorithm is more stable and is also without bias.

5.3 Experiment 3 - Ellipsoid Mapping

Once the faces have been segmented and cropped from the images, they are texture-mapped to one half of an ellipsoid with a vertical diameter 1.5 times the two horizontal diameters which are equal, which is the approximate shape of a human head. Each image is then rotated on the ellipsoid about its vertical axis and the projection of the texture mapped ellipsoid onto the image plane is taken. The employed cost functions are then computed on the projection of the ellipsoid onto the image plane. These costs are minimized by obtaining the angle that provides the minimum cost in an optimisation scheme. There are 25 images of subjects in each orientation that are rectified, from which each error is computed. The error at each orientation is computed as the mean absolute error of the 25 images. A graph of the results are shown in Fig. 5

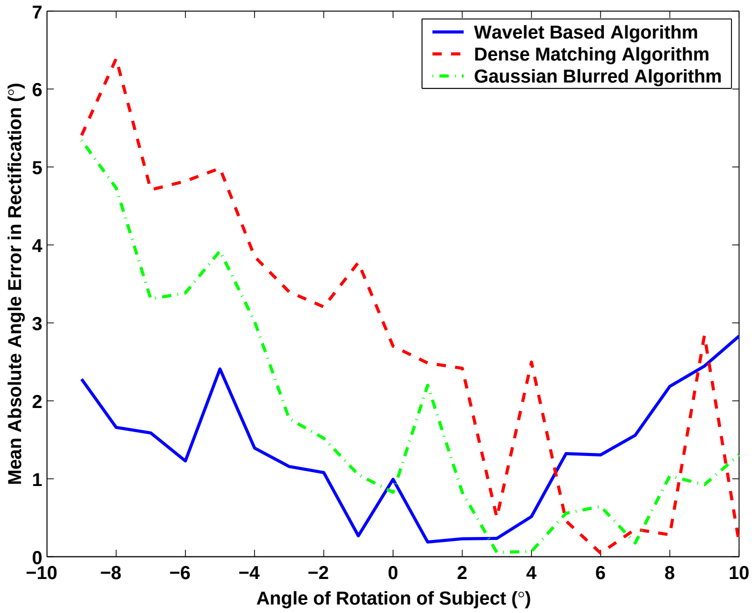

Fig. 5: Mean absolute angle error in rectifying faces with the texture-mapped ellipsoid.

As it can be seen from Fig. 5, the wavelet based function performs better than the comparison methods for the majority of the presented angles. The algorithm is more stable and displays a more smooth error curve, which provides the user with a higher confidence in the algorithm. For the larger positive angles of rotation, the comparison methods demonstrate superior performance compared to the proposed wavelet-based technique. This can be attributed to the bias that is evident for each of the comparison techniques, which may have resulted from unbalanced lighting conditions. The proposed technique is more robust to unbalanced lighting conditions and as such shows no bias.

6 Conclusions and Future Work

A novel method for facial pose removal was presented. No 3D face model was required to be trained and no statistical learning process was used, the algorithm operates on each image individually. It was shown that the continuous symmetry measure proposed in [13] could be improved by first transforming the image into the space-frequency domain using a wavelet transformation. Results were presented highlighting the improvement in the symmetry cost. Finally, experiments were carried out on a database of real face images to remove pose. The results demonstrate improved rectification over the comparison methods. From both the subjective and numerical results, it can be seen that the ellipsoid rectification algorithm provides the most realistic results with little distortion.

To further highlight the advantages of using the wavelet based rectification algorithm, it would be desirable to benchmark the accuracy rate of a person recognition system before and after pose removal. Also, improved efficiency in the computation of wavelet transformations would be desirable since the algorithm can take up to 30 seconds to remove pose from each image. Other possible future applications of the space-frequency domain based rectification methods would be in the areas of planar pose estimation for the purposes of camera calibration.

Acknowledgements

The authors would like to thank Science Foundation Ireland for financially supporting the research presented.

References

1. Goldstein, A.J., Harmon, L.D., Lesk, A.B.: Identification of human faces. Proceedings of the IEEE **59** (1971) 748–760
2. Gao, Y., Leung, M.: Face recognition using line edge map. Pattern Analysis and Machine Intelligence, IEEE Transactions on **24** (2002) 764–779
3. Ghent, J., McDonald, J.: Generating a mapping function from one expression to another using a statistical model of facial shape. In: Proceedings of the 7th Irish machine vision and image processing conference. (2003)
4. Huang, F.J., Zhou, Z., Zhang, H.J., Chen, T.: Pose invariant face recognition. Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on **1** (2000) 245–250
5. Sung, J., Kim, D.: Pose-robust facial expression recognition using view-based 2d + 3d aam. Systems, Man and Cybernetics, Part A, IEEE Transactions on **38** (2008) 852–866
6. Zhang, C., Cohen, F.S.: 3-d face structure extraction and recognition from images using 3-d morphing and distance mapping. IEEE Transactions on Image Processing **11** (2002) 1249–1259
7. Jiang, G., Tsui, H.T., Quan, L., Zisserman, A.: Single axis geometry by fitting conics. In: ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I, London, UK, Springer-Verlag (2002) 537–550
8. Blanz, V., Vetter, T.: Face recognition based on fitting a 3d morphable model. Pattern Analysis and Machine Intelligence, IEEE Transactions on **25** (2003) 1063–1074
9. Lee, H.S., Kim, D.: Generating frontal view face image for pose invariant face recognition. Pattern Recogn. Lett. **27** (2006) 747–754
10. Lanitis, A., Taylor, C.J., Cootes, T.F.: Automatic interpretation and coding of face images using flexible models. Pattern Analysis and Machine Intelligence, IEEE Transactions on **19** (1997) 743–756
11. Edwards, G., Lanitis, A., Taylor, C., Cootes, T.: Statistical models of face images: Improving specificity (1996)
12. Zabrodsky, H., Peleg, S., Avnir, D.: Symmetry as a continuous feature. Pattern Analysis and Machine Intelligence, IEEE Transactions on **17** (1995) 1154–1166
13. Gofman, Y., Kiryati, N.: Detecting symmetry in grey level images: the global optimization approach. Pattern Recognition, 1996., Proceedings of the 13th International Conference on **1** (1996) 889–894