3D face recognition from multiple images: a shape-from-motion approach

Manuel Marques and João Costeira
Institute for Systems and Robotics (ISR) - Instituto Superior Técnico(IST)
Av. Rovisco Pais - 1049-001 Lisboa PORTUGAL
{manuel,jpc}@isr.ist.utl.pt

Abstract

In this article we explore the use of methodologies for 3D reconstruction from multiple images to recognize faces. We try to devise a strategy to tackle the problem of recognizing faces from images exhibiting strong pose (rotation and occlusion) and without prior knowledge (uncalibrated cameras, images from different sources). We do so by framing the recognition in the context of 3D structure from motion with missing data problems. In fact,recently, there has been a strong trend towards using 3D information to verify and recognize faces. However most of the state of the art works are developed over 3D sensors (3D range finders, stereo).Here we propose to do recognition by measuring the likelihood that one or more images were generated by a given 3D shape. In other words, given the subject's 3D shape, we compute the required transformation to match the probe images and measure its deviation from rigidity. This process is not straightforward due to the strong pose: face points are not visible in all images so there is a "missing data" problem intrinsic to the formulation. One key assumption is that all images (and shape) correspond to neutral expression. Though limited in scale due to lack of large databases, a set of tests demonstrate the adequacy and good performance of the approach.

1. Introduction

Face recognition has attracted significant attention in commercial and research communities due to an explosion of applications in security related issues multimedia, human-computer interaction and others. Summarizing the research area in the most classic framework, one can say that "face recognition" is used in many contexts such as recognition from video streams [2] or person identification by a single or a set of images (known as "probes"), knowing *a priori* a face database (referred as "gallery"). Currently, this last context is split into two different tasks: "authentication" or "verification" and "recognition" or "identification".

To achieve this goal, the matching process (the comparison between the probe set and the gallery sets) is one-to-one in authentication scenario and is one-to-many in another one. In this paper, the proposed approach is applied in the identification case (more complex) but it can be used in the verification as well. In recent years, the effort into building robust and user-friendly face recognition system, was concentrated on dealing with the following main issues: illumination changes, head pose variation and different facial expressions.

Early works [3, 21] use 2D images as the input data and can be classified as template-based/apearence-based methods [11]. These algorithms are holistic but its performance decrease significantly with small pose changes. To handle pose variations, geometric-model based methods [10] use 2D deformable models. Before calculating the similarity between the probe face image and the gallery ones, these methods build face models and fit the model to the given image. Another 2D strategy for dealing with the pose variation problem is using images with multiple views of the same face in the gallery [12,14] or synthesizing new images from a given image and 3D face prior knowledge [4, 22].

In recent years, 3D "images" are used as input data for many algorithms because new 3D sensing devices allow better recognition performance than 2D [5, 18]. This fact is true because the 3D data overcome difficulties due to pose and illumination variations. When the 3D data is available, these methods can cope with pose and illumination changes because it is possible to estimate the face surface curvature [15, 23]. Due to the isomorphism between the face's surface and $\mathbb{R}^2$, the method proposed in [6] deals with the face expression problem, by measuring the intrinsic distance between facial feature points. 3D scans of the face can also be used to built detailed face models [1]. The 3D data can be obtained through 3D sensors or two calibrated cameras [17]. This alternative way allows us to obtain a 3D

This work was partially supported by the Fundacao para a Ciencia e Tecnologia (ISR/IST pluriannual funding) through the POSC Program that includes FEDER funds, and J. Costeira partially funded by project PTDC/EEA-ACR/72201/2006

978-1-4244-2154-1/08/$25.00 ©2008 IE

dense reconstruction by a stereo preprocessing.

In this paper we want to take advantages of modeling the human face as a 3D object using 2D images, such as [9,17], but avoiding expensive 3D sensors. Unlike these two works, the method proposed in this paper needs a few feature points and can be used with uncalibrated cameras. Since we use point features and do not require calibration our method can use images of different sources (like cameras, press and Internet) and avoids computationally heavy matching preprocess. Also, in general feature points are more reliable to illumination changes and finally, since we can cope with partial data (missing data), unreliable points can be eliminated (declared invisible). In summary, we propose here a face recognition method with the following main characteristics:

- uses 2D images from uncalibrated cameras
- images can display large pose variations
- models explicitly scaling factors allowing images captured at different distances from the subject.

These main features must be understood in a constrained context. In other words, there are some underlying assumptions:

- face expression is close to neutral.
- images are well modeled by orthographic projection.
- at least 3 images for training.
- position of feature points are known (matching is solved).

Our methodology hinges on the rigid (3D) relation between different views of the same object. By using a 3D reconstruction algorithm that can handle missing data, we compute the 3D shape (3D location up to scale) of a set of face points (training). With the available test images (one or more) recognition/verification is performed by verifying the deviation of these images from a rigid transformation followed by a projection.

2. Problem and Notation

The human face is represented by a set of point features which image position we consider known. In our notation, the human face is thus represented by P points viewed in F frames. We can define a *data matrix* built by stacking the image coordinates as follows:

$$\mathbf{W} = \begin{bmatrix} u_1^1 & \dots & u_1^P \\ v_1^1 & \dots & v_1^P \\ \vdots & \ddots & \vdots \\ u_F^1 & \dots & u_F^P \\ v_F^1 & \dots & v_F^P \end{bmatrix} \quad (1)$$

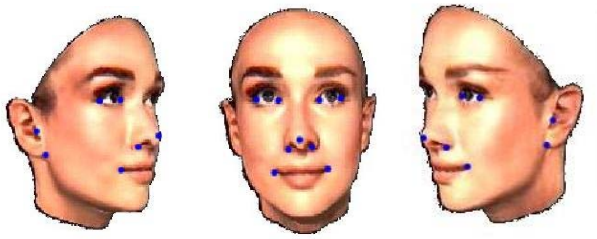

Figure 1. The select features

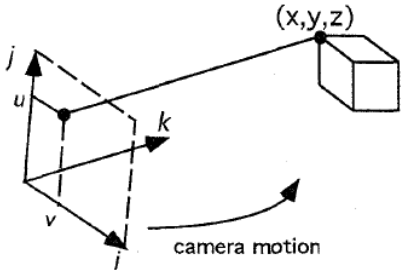

Figure 2. Orthographic projection

where u_f^p and v_f^p are the p point projection in frame f. In this paper we will consider 13 points shown in figure 1[1]): **eyes corners** (4 points), **mouth corners** (2 points), **tip of the nose and 2 points in the nose base** and **targus and lobule in the ear** (4 points).

Of course, the approach is independent of point type and its number. These were chosen on one hand due to the its near-rigidity (except the mouth) and on the other because they are the most identifiable feature points, providing good discrimination [24]. There are no restrictions to adding other feature points.

If the size of the head is small relative to its distance from the camera, as in [9], we can assume the scaled-orthographic camera model for the data. Then, $\mathbf{W}$ can be written according to the following equation (see figure 2):

$$\mathbf{W} = \begin{bmatrix} \mathbf{i_x^1} & \mathbf{i_y^1} & \mathbf{i_z^1} \\ \mathbf{j_x^1} & \mathbf{j_y^1} & \mathbf{j_z^1} \\ \dots & \dots & \dots \\ \mathbf{i_x^f} & \mathbf{i_y^f} & \mathbf{i_z^f} \\ \mathbf{j_x^f} & \mathbf{j_y^f} & \mathbf{j_z^f} \end{bmatrix} \begin{bmatrix} \mathbf{X_1} \dots \mathbf{X_P} \\ \mathbf{Y_1} \dots \mathbf{Y_P} \\ \mathbf{Z_1} \dots \mathbf{Z_P} \end{bmatrix} + \begin{bmatrix} t_u^1 \\ t_v^1 \\ \dots \\ t_u^f \\ t_v^f \end{bmatrix} \quad (2)$$

$$\mathbf{W} = \mathbf{MS} + \mathbf{t}^T\mathbf{1} \quad (3)$$

where $\mathbf{M}_{[2F\times 3]}$ is the *motion matrix*, $\mathbf{S}_{[3\times P]}$ the *shape matrix* and $\mathbf{t} = \begin{bmatrix} t_u^1 t_v^1 \dots t_u^F t_v^F \end{bmatrix}^T$ the *translation vector*. Rows $2i-1$ and $2i$ of $\mathbf{M}$ are orthogonal.

Note that vectors i^f and j^f, which represent the camera axis in image f, are not unit-norm vectors. Their norm is

[1] http://www.kyb.tuebingen.mpg.de/bu/people/volker/audrey.MPG

$||\mathbf{i}^f||^2 = ||\mathbf{j}^f||^2 = \alpha^f$. This models the image scale factor. In fact faces can be obtained at any distance from the camera, and if orthography is adequate, this scale factor relates directly to depth. Each pair of stacked vectors i^f and j^f form a Stiefel matrix (two rows of a 3x3 rotation matrix) multiplied by the scale factor.

If all point projections are known in all f frames, the *shape matrix* (3) can be calculated in closed-form [19]. However this is not the case if images have strong pose. In almost any images there will be a subset of points that are not visible (occluded), as figures 1 and 3 show.

Mathematically, we do not know all the *data matrix* values (1) but only a set of them given by the *known data matrix* $\mathbf{Z}$, written as:

$$\mathbf{Z} = \mathbf{W} \odot \mathbf{D} \tag{4}$$

where $\odot$ represents the Schur (elementwise) product and $\mathbf{D}$ is the *mask matrix*. This matrix is a binary matrix identifying the known data with 1 and unknown data with 0.

3. Recognizing using 3D Shape

The whole face recognition process hinges on 3D shape computation. In other words, there are two main steps in the whole process: First we must compute the 3D shape of the feature points from incomplete data (missing points due to pose). Matrix S (shape) thus defines uniquely each subject. Second, the goal in the recognition process amounts to finding the shape that after some rotation, translation and projection is more likely to explain the 2D data (one or more test images). This process is also done with some data missing.

In summary, we recognize by seeking the best rigid explanation for the data. It is clear now that estimating shape from incomplete data is a key procedure in the whole process, so next sections will be devoted to this task.

3.1. The missing data problem and shape estimation

In this section, we explain how to estimate the face's shape knowing the 2D projections of 13 feature points in F images with a partial view (4). Considering the human face (with neutral expression, in this case) as a rigid object and verifying the orthographic model equation (3), we can easily conclude that $\mathbf{W}$ has at most rank 4 [19]. We can estimate $\mathbf{Z}$'s unknown values in the following way: Among all rank 4 matrices choose the one that best approximates (in the LSE sense) the known data. In other words, solve the following problem:

Problem 1

$$(\widehat{\mathbf{Z}})^* = arg\min_{\widehat{\mathbf{Z}}} \quad \left|\left|\left(\mathbf{W} - \widehat{\mathbf{Z}}\right) \odot \mathbf{D}\right|\right|_F^2$$
$$s.t. \quad \widehat{\mathbf{Z}} \in \mathcal{S}_4$$

Figure 3. Left - Frontal pose: the viewed features are represented by the green squares Right - Cyan triangles are the the missing features, estimated by a state of the art rank-based algorithm.

Figure 4. Visible features are represented by green squares. Estimates of occluded features using [7] are represented by cyan triangles. Features represented by red circles were computed using our rigid factorization algorithm. Left - The cyan triangles correspond to features of the right ear and, in this case, data is not degenerate. Right - Frontal pose: a typical degenerate case.

where $\mathcal{S}_4$ is the space of matrices with rank equal to 4.

There are several iterative algorithms [7, 13] to solve the missing data problem with rank constraint, such as Problem 1. After solving it through this approach - the rank-based approach -, the Tomasi-Kanade algorithm [19] can be used to estimate shape and motion. However, the image stream has one single degenerate image (an image where the 3D points of the known projections belong to a 1D or 2D subspace), problem 1 becomes an unconstraint problem [16]. This situation is quite usual in face recognition since frontal face images (figures 3 and 4) are fairly common in databases. The wrong results described before happen because the viewed 3D points (the eyes and mouth corners, for example) belong to the same 2D subspace (lye on a 3D plane). Due to this, the rank-based approaches do not obtain a correct face shape estimate.

To use frontal images in the gallery set and, at the same time, obtain a correct face, we must redefine the optimization problem according to [16], replacing problem 1 by

Problem 2

$$(\widehat{\mathbf{M}}, \widehat{\mathbf{S}})^* = arg\min_{\widehat{\mathbf{M}}\widehat{\mathbf{S}}} \sum_f \left|\left|\left(\mathbf{W}_f - \widehat{\mathbf{M}}_f\widehat{\mathbf{S}} + \widehat{t}_f\mathbf{1}_{[1,P]}\right) \odot \mathbf{D}\right|\right|_F^2$$
$$\widehat{\mathbf{M}_1}\widehat{\mathbf{M}_1}^T = \alpha^1 I$$
$$s.t. \quad \vdots$$
$$\widehat{\mathbf{M}_F}\widehat{\mathbf{M}_F}^T = \alpha^F I$$
$$\alpha^f \in \mathbb{R}^+, \quad \forall_f$$

where $\widehat{\mathbf{M}} = \left[\widehat{\mathbf{M}_1}^{\mathrm{T}} ... \widehat{\mathbf{M}_F}^{\mathrm{T}}\right]^{\mathrm{T}}$ (3) and α^f is the scale factor of frame f.

The main difference between this strategy and the previous one is in the constraints: the constraints in problem 2 are the orthogonality constraints. Due to this, the orthographic camera model (3) is imposed to the data. To find a solution for problem 2, we use the following iterative algorithm:

ALG1
1. Initializations: $\widehat{\mathbf{Z}}_0 = \widehat{\mathbf{W}}, \quad k = 0$
2.Estimate translation (centroid).
$\widehat{\mathbf{t}}_k = \left[\frac{1}{P}\sum_i \widehat{\mathbf{Z}_{1ik}} \quad \dots \quad \frac{1}{P}\sum_i \widehat{\mathbf{Z}_{[2f,i]_k}}\right]$
$\widehat{\mathbf{Z}_{ck}} = \widehat{\mathbf{Z}}_k - \widehat{\mathbf{t}}_k$ Remove translation
$k = k + 1$
3. Estimate $\widehat{\mathbf{M}}_k$ e $\widehat{\mathbf{S}}_k$ (run ALG2 with $\widehat{\mathbf{Z}_{ck}}$)
4. Update data matrix
$\widehat{Z}_k = \underbrace{(\widehat{\mathbf{M}}_k\widehat{\mathbf{S}}_k + \widehat{\mathbf{t}}_{k-1}\mathbf{1}_{[1,P]}) \odot \bar{\mathbf{D}}}_{\text{Missing data estimate}} + \underbrace{\mathbf{Z} \odot \mathbf{D}}_{\text{Known data}}$
$\bar{\mathbf{D}}$ - 2's complement of $\mathbf{D}$ i.e. $\bar{\mathbf{D}} = \mathbf{1}_{[2F,P]} - \mathbf{D}$
5. Verify if $||\widehat{\mathbf{Z}}_k - \widehat{\mathbf{Z}}_{k-1}|| < \epsilon$.
If not verify go to step 2 and $k = k + 1$.

Each matrix $\widehat{\mathbf{M}}_k$, computed in step 3, must satisfy the orthogonality constraints. Due to this, we use a rigid factorization method given by

ALG2 Args: $\mathbf{Z}_c$
1. Initial factorization:
factorize $\mathbf{Z}_c$ using any factorization (e.g. SVD)
$\mathbf{Z}_c = \mathbf{AB}, \quad \mathbf{R} = \mathbf{A}, \quad \widehat{\mathbf{M}^0} = \mathbf{A}, \quad \widehat{\mathbf{S}^0} = \mathbf{B}$
$k = 1$
2. Project R into the manifold of motion matrices
$\widehat{\mathbf{M}^k} = \arg\min_{\mathbf{X}} \sum_f ||\mathbf{R}_f - \mathbf{X}_f||_F^2$
s. t. $\mathbf{X_f}\mathbf{X}_f^T = \alpha_f \mathrm{I}_{2x2} \quad \forall f$
$\alpha \in \mathbb{R}^+$
3. $\widehat{\mathbf{S}^k} = \widehat{\mathbf{M}^k}^+ Z_c, \quad \widehat{\mathbf{M}^k}^+$ - Moore-Penrose pseudoinverse
4. $\mathbf{R} = \mathbf{Z}_c\widehat{\mathbf{S}}_k^+$
5. Verify if $||\widehat{\mathbf{M}}_k - \widehat{\mathbf{M}}_{k-1}|| < \epsilon$.
If not, go to step 2 and $k = k + 1$.
6. $\widehat{\mathbf{M}} = \widehat{\mathbf{M}}_k$ and $\widehat{\mathbf{S}} = \widehat{\mathbf{S}}_k$

In step 2, it is important to note that matrices $\widehat{\mathbf{M}^k}$ are obtained in closed-form - *Procustes problem*. To see the full details about this algorithm and its performance see [16]. In figure 4, we can verify that this algorithm provides accurate and reliable locations to the occluded features. In the left image, the cyan triangles represent the estimated position of the right ear points done with Buchanan& Fitzgibbon's algorithm [7], considered the best performing matrix completion algorithm. Estimates of the above algorithm are represented by the red dots (the points of the left ear are projected over the nose, where it should be). The right image of the same figure just show a frontal image (degenerate data - points on a planar surface). Note that one single degenerate image produces such error, regardless of the number of "well-behaved" images (with non-coplanar points).

So, this means that shape and motion estimates are correct and thus allows our face recognition methodology dealing simultaneously with 3D shape and strong pose and scale factor. In other words the correct results can be obtained only if the orthogonality constraints are verified. Any other method (say [7, 13]) would fail unless all data are non-degenerate (no frontal images, for example).

3.2. The classification task

The probe images are represented the same way as the gallery images: by those face features mentioned in section 2. It is also important to note that the pose of face in a probe image is not *a priori* specified.

The main idea about evaluating the similarity degree between two different data sets is to use the same criterion used to estimate 3D shape (in section 3.1) - rigidity. This means that if a probe and gallery sets match, given a known 3D shape of the gallery subject, the 2D projections of the probe set are in accordance to the orthographic camera model (3). To compute the compatibility degree between these two sets, we solve a similar problem to Problem 2 but with one frame, only, that is, computing the best Stiefel matrix $\widehat{\mathbf{M}}_i$ that adjust gallery shape to probe:

Problem 3
$$\epsilon_{ij} = \min_{\widehat{\mathbf{M}}_i} \sum_k^F \left|\left|\mathbf{W}_{ik} - \widehat{\mathbf{M}}_i\mathbf{S}_j\right|\right|^2$$
$$\textit{s.t.} \quad \widehat{\mathbf{M}}_i\widehat{\mathbf{M}}_i^T = \alpha I$$

where $\mathbf{W}_{ik}$ is composed by the 2D projections of image k of probe i and $\mathbf{S}_j$ the shape of face j. Note that, in this case, $\mathbf{W}_{ik}$ and $\mathbf{S}_j$ are known (no missing elements). Then, the classification of the face of probe image k is obvious: it is the same face of gallery 3D shape j if ϵ_{ij} is below of a threshold.

The above optimization is quite straightforward and efficiently done: Knowing the Shape an initial estimate can be computed from the Least-Square Error estimate ($\widehat{\mathbf{M}}_i = \mathbf{W}^i\mathbf{S}_j^T(\mathbf{S}_j\mathbf{S}_j^T)^{-1}$) and for high precision, three steps of a Newton algorithm are performed.

4. Experiments

To evaluate if 3D shape and rigidity are good cues to face classification with strong pose, we put forward a small experimental setup consisting of 110 images belonging to 12 subjects.

Figure 5. Pose variation of typical gallery set. Each set was formed by 5 images.

Figure 6. Typical images for the probe set.Some sets included 2 images per subject

The type of images used in this test are shown in figure 5 and 6. Images were obtained by a regular digital camera (resolution 3072×2048 pixels), with strong pose variations. There was no constraints on the viewing angle and distance between the camera and the subjects (scale factor) also varied. Feature points were selected manually.

The results presented in figures 7 and 8 were obtained from 130 experiments, consisting on randomly selecting probe and gallery images from the whole set of images. After the random choice of the gallery set, each shape face is calculated from 5 2D images of a same person. In figure 5 we show 4 images from a typical gallery set. On the other hand, one element of the probe set is composed by 1 or 2 images. In the 2 image case, they are also selected randomly. In figure 6, the 4 images belong to 4 different probe sets, composed by one single image.

In figure 7 we show the error histograms for the 130 experiments with one image (left graph) and 130 with two images (right graph). This histogram represents the percentage of correct cases (blue dashed curves) where shape and data belong to the same subject and the percentage of wrong cases (red curves), as a function of error ϵ_{ij}. As it is clear, both classes can be easily separated. This separation is even deeper if we increase the size of probe images as shown in the right graph of figure 7 and more evident in the ROC curve shown in figure 8. In this figure, the solid line represents the ROC curve for the one probe image experiment and the dashed line the ROC curve for the 2 image experiment. At 0% false positive rate we can still achieve 85% of correct classification.

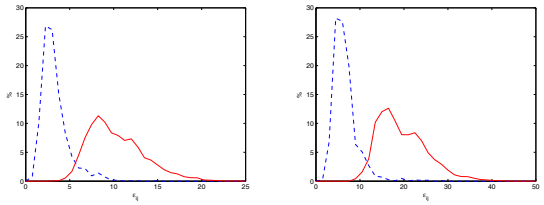

Figure 7. Histogram of data reconstruction error (error of Problem 3. Left- One image per probe set. Right- Two images per probe set. X-axis represent the error ϵ_{ij}. Blue dashed curve represents percentage of correct cases: Probe and gallery sets are compatible.Red curve corresponds to incompatible sets

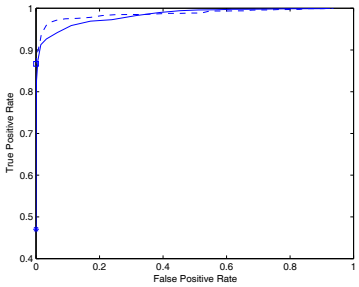

Figure 8. ROC curve: Solid line - 1 image in each probe set Dashed line - 2 images in each probe set

4.1. Conclusions and future work

In this paper, we shown a new face recognition method to deal with strong pose, which by nature generate occlusion of the face. By fusing information from several images it handles any type of pose. Also, since it relies on orthographic models it works with images from uncalibrated cameras. Because the reconstruction algorithm models reality more faithfully (imposing the full camera model), it can use degenerate data like frontal images where the viewed points lye on planar surfaces (degenerate 3D data!).

Even though it requires a scaling up of the data set for more accurate evaluation, we believe that this innovative concept is quite promising and definitely robust to strong pose.

The methodology presented here assumes that faces exhibit a neutral and similar expression over the whole data set. However, there are several factorization algorithms that can deal with nonrigid objets (strong articulations and small deformations) [8, 20]. Intrinsically the modeling is similar in the sense that observations (image points) are generated by bilinear models. The missing data is clearly extendable. However the parametrization and the meaning of shape and motion is not straightforward. Here the motion matrix includes other terms than rigid transformations. In rigid situations the error is clearly defined and easy to measure, which is not the case in the non-rigid case. Experiments will be conducted to evaluate the possibility of extending the methodology to these more broader cases.

References

[1] B. Amberg, A. Blake, A. Fitzgibbon, S. Romdhani, and T. Vetter. Reconstructing high quality face-surfaces using model based stereo. *ICCV 2007*, pages 1–8, 14-21 Oct. 2007.

[2] O. Arandjelovic and A. Zisserman. Automatic face recognition for film character retrieval in feature-length films. In *CVPR '05 - Volume 1*, pages 860–867, 2005.

[3] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 19(7):711–720, 1997.

[4] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In A. Rockwood, editor, *Siggraph 1999, Computer Graphics Proceedings*, pages 187–194, Los Angeles, 1999. Addison Wesley Longman.

[5] K. W. Bowyer, K. I. Chang, and P. J. Flynn. 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.

[6] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Expression-invariant representations of faces. *IEEE Transactions on Image Processing*, 16(1):188–197, 2007.

[7] A. Buchanan and A. Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *CVPR 2005*.

[8] A. D. Bue and L. de Agapito. Non-rigid stereo factorization. *International Journal of Computer Vision*, 66(2):193–207, 2006.

[9] C. D. Castillo and D. W. Jacobs. Using stereo matching for 2-d face recognition across pose. *CVPR '07*, pages 1–8, 17-22 June 2007.

[10] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[11] S. Du and R. Ward. Face recognition under pose variations. *Journal of Franklin Institute*, 343:596–613, 2006.

[12] A. Georghiades, D. Kriegman, and P. Belhumeur. From few to many: Illumination cones for face recognition under variable lighting and pose. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[13] R. Guerreiro and P. Aguiar. Estimation of rank deficient matrices from partial observations: Two-step iterative algorithms. *Proc. Conf. Energy Minimization Methods in Computer Vision and Pattern Recognition.*, 2003.

[14] P. Huisman, R. van Munster, S. Moro-Ellenberger, R. Veldhuid, and A. Bazen. Making 2d face recognition morerobust using aams for pose compensation. In *Proc. of the 7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 108–113, 2006.

[15] Y. Lee, H. Song, U. Yang, H. Shin, and K. Sohn. Local feature based 3d face recognition. In *AVBPA 2005, NY, USA*, pages 909–918, 2005.

[16] M. Marques and J. Costeira. Optimal shape from motion estimation with missing and degenerate data. In *WACV-MOTION '08: Proc. of the IEEE Workshop on Motion and Video Computing*, Copper Mountain, CO, USA, 2008.

[17] G. Medioni and R. Waupotitsch. Face modeling and recognition in 3-d. In *AMFG '03: Proc. of the IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, page 232, 2003.

[18] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp. A survey of 3d face recognition methods. In *AVBPA 2005, NY, USA*, pages 891–899, 2005.

[19] C. Tomasi and T. Kanade. Shape and motion from image stream under orthography: a factorization method. *Int. Journal of Computer Vision*, 9(2):137–154, 1992.

[20] L. Torresani, A. Hertzmann, and C. Bregler. Learning nonrigid 3d shape from 2d motion. In *NIPS*, 2003.

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

[22] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):733–742, 1997.

[23] S. Wang, M. Jin, and X. D. Gu. Conformal geometry and its applications on 3d shape matching, recognition, and stitching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(7):1209–1220, 2007.

[24] P. Yan and K. W. Bowyer. Ear biometrics using 2d and 3d images. *CVPR 2005*, 3:121–121, 20-26 June 2005.