RECONSTRUCTION, REGISTRATION, AND MATCHING OF 3D FACIAL MODELS FROM STEREO-IMAGES

Yu-Chuan Chang and Ayman F. Habib

Department of Geomatics Engineering, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada
T2N 1N4 - (ycchang, habib)@geomatics.ucalgary.ca

WgS: WG V/6

KEY WORDS: Reconstruction, Registration, Photogrammetry, Matching, Biometrics, Computer Vision

ABSTRACT:

Interest in biometric systems has dramatically increased worldwide due to the rising public demand for more reliable and effective criminal identification and effective surveillance systems. Biometric facial recognition has been studied intensely over the past decade due to its potential applications in government, law enforcement and business. Previous studies in facial recognition have tended to focus on using two-dimensional (2D) images. The performance of current 2D image-based facial recognition systems, however, is clearly unsatisfactory due to their sensitivity to changes in facial expression, body position, lighting, sensor conditions and other disturbance factors. Further, some 2D systems are easily fooled by simply presenting a photo in front of the sensor. The move from 2D to three-dimensional (3D) recognition technology is expected to improve performance in real-life environments. This study developed a low-cost imaging system to capture overlapping imagery which was then used to construct a 3D facial model. The generated 3D model was then registered and matched with available 3D models in a central database for personal verification or identification purposes. The experimental results showed that low-cost digital cameras, after proper calibration, can construct accurate 3D facial models when combined with an active pattern projection system. Preliminary experiments also demonstrate the feasibility and robustness of the proposed automatic surface registration and matching procedure. This study discusses the performance, advantages and limitations of the proposed method.

1. INTRODUCTION

Biometric measurements are being studied for security applications as an alternative to Personal identification number (PIN) codes and cards. In cooperative environments, speech and face modalities are well accepted by individuals but have yet to demonstrate acceptable performance and reliability. A previous profile analysis (Beumier, 1995) demonstrated the adequacy of geometrical information for automated personal authentication. Such information is derived by analyzing rigid areas of the face, such as the forehead, nose and chin and are generally unaffected by makeup or lighting conditions. This explains the success of many profile works (Chellappa, 1995). A facial 3d description requires more geometrical information, especially where grey level features are lacking as in the chin, forehead and cheek regions. Such an analysis would benefit from actual 3D measures without scale or rotation influence. Depth information also helps to distinguish the face from background elements. These advantages make 3D geometrical approaches an important complement to grey level analysis.

Previous research in 3D facial recognition has yielded two main techniques: the stereo acquisition technique and laser-based acquisition technique. Most recent research have utilized laser scanning systems, which are typically more accurate but also more expensive and time consuming (Achermann, 1997). The acquisition of a single 3D head scan can take more than 30 seconds, which is a major limitation of laser-based systems. This relatively long scan time introduces errors due to the object's breathing and movement resulting in errors in 3D reconstruction of a face.

The stereo acquisition technique employs two or more cameras positioned and calibrated to simultaneously acquire images of the subject (Lao, 2000). The location for each point in 3D object space can be computed by using a photogrammetric procedure. This method has the lowest cost and highest ease of use. A simultaneous acquisition system using stereo-photogrammetric technologies does not have serious motion problems that a laser-based acquisition system has. However, photogrammetric systems are not widely used in commercial applications for facial model recognition due to their unsatisfactory automation of matching, which is a process for identifying conjugate landmarks of an object's surface in stereo images.

Landmark locations used for matching in previous work can be found either manually (Lao, 2000) or automatically (Yacoob, 1994). The correct localization of the landmarks is crucial to many algorithms, and judging the sensitivity of an algorithm to localization errors by its description alone is usually not possible. Nevertheless, automatic landmark localization remains an unsolved problem. Among the possible optical acquisition systems (Jarvis, 1993), structured light has emerged as the optimal solution for identifying landmarks in homogeneous areas.

This work presents an efficient and automatic algorithm for 3D model reconstruction from facial images. A photogrammetric model is proposed for surface reconstruction from stereo pairs using pattern projection. Its resolution, speed and adequate facial coverage make it attractive for numerous practical implementations. The next section briefly describes the photogrammetric principles for 3D facial measurements. The proposed system design for homogeneous surface model generation is then described and experimental results are presented. Finally, conclusions regarding the effectiveness and possible uses of the proposed system are summarized.

2. METHODOLOGY

Photogrammetric techniques define the shape, size and position of objects using images taken from different viewpoints. Photogrammetric reconstruction is based on the collinearity equation, which states that the image point, the perspective centre and the corresponding object space point are collinear. The internal orientation parameters (IOPs) of the implemented camera, which include the principal distance of the camera (c), the coordinates of the principal point (x_p, y_p) and distortion parameters, are accurately recovered through a camera calibration procedure. The exterior orientation parameters (EOPs) define the position (X_o, Y_o, Z_o) and the orientation (ω, φ, κ) of the reconstructed bundle relative to the object space coordinate system. The EOPs simulate the actual position and orientation of the camera at the moment of exposure. A photogrammetric system first identifies a pair of conjugate points in overlapping areas between two 2D images acquired by calibrated cameras. Reducing the search area for conjugate features achieves better results and reliability. Epipolar geometry is commonly used to constrain the search in matching. Conjugate light rays can be reconstructed after identifying conjugate points. The intersection of two conjugate light rays defines an object point in 3D space. According to the above concepts, the proposed procedures for 3D reconstruction modeling require the following fundamental procedures: Image acuqisition, epipolar transformation, matching and intersection. (see Figure 1).

2.1 Image Acquisition

First, the utilized cameras undergo a calibration and stability analysis procedures. The objective of the calibration process (Habib and Morgan, 2003) is to derive the cameras' internal characteristics including principal point coordinates, principal distance, lens distortions, etc. The stability analysis, on the other hand, aims at verifying that the estimated internal characteristics do not significantly change over time (Habib et al., 2006).

Provided with initial estimates of EOPs and a test field with 3D points which are measured in advance, bundle adjustment can perform an estimation which minimizes the re-projection error by adjusting the bundle of rays between each camera centre and the set of 3D points. The EOPs of the camera at the moment of exposure can be obtained through a bundle adjustment procedure.

Since the human face is a relatively homogeneous surface, few conjugate features can be identified. To overcome such a limitation, structured patterns can be projected onto the face during image acquisition to increase the density of identifiable points on the facial surface. The pattern projection technique was selected for several reasons. First, the pattern projection is especially useful for providing artificial landmarks in homogeneous areas by projecting a light pattern on the face. Second, this setup is relatively fast, inexpensive and enables acquisition of 3D information using easily available and low cost digital cameras. The additional cost is limited to a projector. Eleven 3 by 3 sub-blocks were used for the encoding pattern. The sub-blocks were randomly selected and arranged for this pattern (Figure 2). To minimize ambiguity, the sub-block should not be repeated within a certain radius.

After an imaging environment with pattern projection is setup, a subject with projected pattern can be imaged by using calibrated cameras with known IOPs and EOPs at different locations.

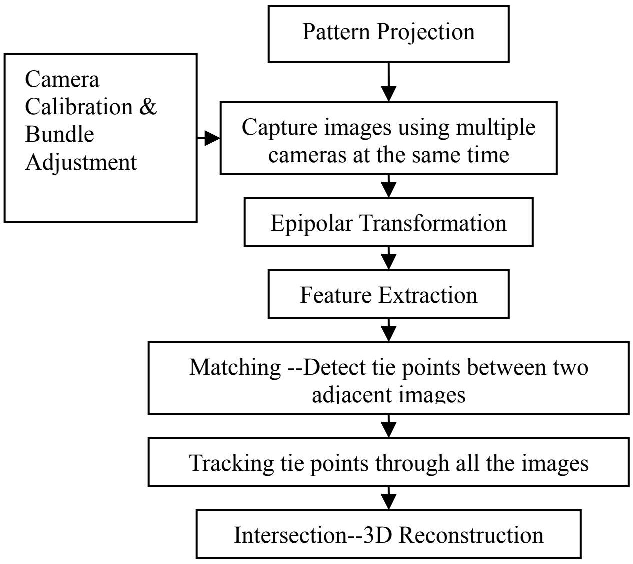

Pattern Projection
Camera Calibration & Bundle Adjustment
Capture images using multiple cameras at the same time
Epipolar Transformation
Feature Extraction
Matching --Detect tie points between two adjacent images
Tracking tie points through all the images
Intersection--3D Reconstruction

Figure 1. The flowchart for 3D reconstruction system design using pattern projection

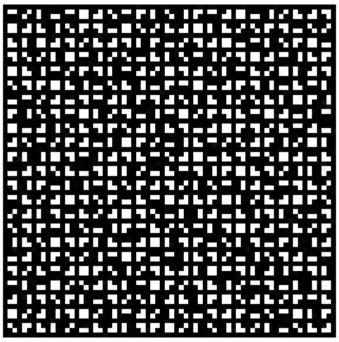


Figure 2. The designed pattern for pattern projection

2.2 Epipolar Transformation, Feature Extraction and Matching

The acquired images have to be pre-processed in order to perform better matching. The captured images are first resampled through Epipolar transformation. The main objective of epipolar transformation is to generate normalized images with corresponding points on the same rows. Epipolar trasnformation of frame images is a straightforward process. The resampling process involves projecting the original images onto a common plane in an orientation determined by the parameters of the original images. Original and normalized scenes share the same perspective centre. During normalization, the two optical axes should be parallel to each other and perpendicular to the baseline. For a normalized pair, we can search the conjugate points along the corresponding row.

During matching process, conjugate points can be more reliably detected by identifying features on the surface. In the proposed algorithm, feature extraction can be achieved by using the Harris corner detector (Harris, 1988), which is a popular interest point detector due to its reliable invariance to: rotation, scale, illumination variation and image noise (Schmid, 2000).

Searching conjugate points in stereo digital images can be automated using image matching procedures based on well defined techniques measuring the similarity of the local area surrounding a feature. The image matching technique was intended to precisely locate tie points during image-to-image registration and estimate the accuracy of the local image-to-image transformation. The choice was based on Normalized Cross Correlation (NCC), which returns the similarity of two windows of pixels (Equation 1). Two highly similar windows centered on corners indicate that the corners match. The matching result is the position in the test image with the highest correlation score.

$$S = \frac{\sum_{x,y}(I(x,y)-\bar{I})(J(x,y)-\bar{J})}{\sqrt{\sum_{x,y}[(I(x,y)-\bar{I})]^2 \sum_{x,y}[(J(x,y)-\bar{J})]^2}} \tag{1}$$

Where:

$\bar{I} = \frac{1}{N}\sum_{x,y} I(x,y)$ and $\bar{J} = \frac{1}{N}\sum_{x,,y} J(x,y)$

I(x, y) is the value of a pixel from window 1.
J(x, y) is the value of a pixel from window 2.
N is the total number of pixels in window 1 or window 2.

2.3 Reconstruction

A shorter baseline between stereo images is better for reducing occlusions and increasing the reliability of matching. However, because a smaller baseline produces larger errors along depth direction, a larger baseline with an intersection angle near 90 degrees is needed to improve accuracy. Ideally, the system would combine the advantages of short baseline and large baseline. Our hypothesis is as follows. First, intermediate images between two images with an appropriate baseline for best accuracy should be used for better matching. Second, tracking conjugate points through all images and reconstructing a point intersected by conjugate light rays from stereo images with an appropriate baseline can achieve an automatic matching procedure for generating accurate output of a reconstructed surface model.

By using stereo imaging, 3D object points can be derived by the intersection of conjugate light rays which are defined by the conjugate points, the IOP of the camera, and the EOP of the image. A set of randomly distributed points can be obtained by using the intersection process. A well-reconstructed 3D facial model established from these random points requires an interpolation method. Thus, Thin Plate Spline is used in this stage. Thin Plate Spline (TPS) is an interpolation method that finds a "minimally bended" smooth surface that passes through all given points. The thin plate spline is the two-dimensional analog of the cubic spline in one dimension and is the fundamental solution to the biharmonic equation. The TPS can represent the surface through a mathematical function where the facial surface can be resampled and modeled with regularly spaced points.

2.4 Registration

Cheng and Habib (2007) introduced an automated surface matching algorithm for registering 3D geographic datasets constructed relative to two reference frames. The Modified Iterated Hough Transform (MIHT) is coupled with the Iterative Closest Patch (ICPatch) algorithm to improve the convergence rate of the matching strategy as it relates to the nature of acquired surface models. This algorithm can be used to model surfaces with randomly distributed points when it is unknown how they correspond with each other.

Considering the characteristics of collected surface models, 3D points can be used to represent the first surface (S_1) while triangular patches can be used to define the second surface (S_2). 3D similarity transformation is used for describing the mathematical relationship or mapping function between the reference frames associated with the two surfaces. Seven parameters are involved in 3D similarity transformation, including three translations, one scale, and three rotational angles. A coplanarity constraint (Figure 3) is used here as the similarity measure in this algorithm. The enclosed volume of a transformed point and the corresponding patch should be zero if the point belongs to the same plane as the patch (Equation 2).

$$V = \begin{vmatrix} X_{q'} & Y_{q'} & Z_{q'} & 1 \\ X_{pa} & Y_{pa} & Z_{pa} & 1 \\ X_{pb} & Y_{pb} & Z_{pb} & 1 \\ X_{pc} & Y_{pc} & Z_{pc} & 1 \end{vmatrix} = 0 \tag{2}$$

Where $X_{q'}$, $Y_{q'}$, $Z_{q'}$ are the coordinates of the transformed point from S_1, and p_a, p_b, p_c denote the coordinates of the three vertices of the conjugate patch from S_2

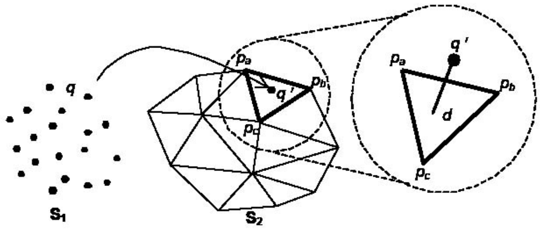


Figure 3. Similarity measure for relating conjugate primitives in two facial models.

The MIHT approach is a voting procedure that derives the most probable solutions of the transformation parameters needed for the best alignment of two surface models by considering all the possible matches between points in S_1 and patches in S_2. The MIHT also determines the correspondence between conjugate surface elements in the involved facial models. To improve the performance of the MIHT, the ICPatch is used to fine tune the estimated transformation parameters and corresponding elements in the involved facial models by using only matched point-patch pairs obtained from MIHT. This algorithm does not deform the surfaces and can perform registration and matching in a one-step procedure.

3. EXPERIMENTS

The proposed approach employs a low-cost imaging system to capture overlapping imagery, which are then used to derive a 3D facial model. The generated 3D model is then registered and matched with available 3D models in a central database for personal verification or identification purposes. A Canon EOS Digital camera (8 mega pixels; pixel size: 6.5 micrometers) was used in this study to capture images of two persons. The camera was accurately calibrated, checked for stability and mounted on a tripod. A test field with a set of 3D points which were previously measured was constructed to compute the position and orientation of the camera at the time of exposure. To reduce the motion caused by human operation, image capturing was remotely controlled. A pattern was projected onto the face during image acquisition by using a Sony projector to enable easier identification of a dense set of points. For each person, seven images were collected by using the same camera in different locations. Although only one camera was used in our experiments at the current stage, seven cameras should be used in the future to ensure simultaneous data acquisition. The three following datasets were acquired:

1. Images of the first person at time T_1. (David_1)
2. Images of the second person at time T_2. (Ivan_1)
3. Images of the first person at time T_3. (David_2)

The pattern projection produced encoded patterns on the face and provided specific landmarks on homogenous surfaces, thereby solving the problem of identifying conjugate features in the stereo-pair. To increase the reliability of the matching process, the internal and external characteristics of the utilized cameras and images, respectively, were incorporated to reduce the search space from 2D to 1D by using epipolar geometry for stereo-imagery. After the Harris operator was applied to automatically detect corners resulting from the pattern projection system, the NCC was then performed for identifying conjugate features between adjacent images (Figure 4).

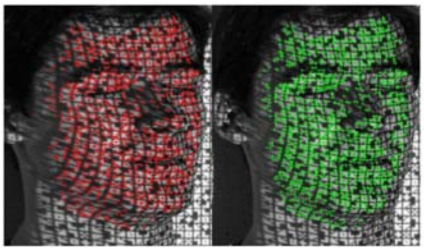


Figure 4. Extracted corners and detected tie points on images #1 and #2 in Ivan_1.

Following the matching process, the tie points were tracked through all images captured in different locations. Using tracked conjugate points between two images with an intersection angle close to 90 degrees for best accuracy, an intersection procedure was performed to derive the 3D coordinates of the corresponding object points. A TPS algorithm was then used to generate a 3D facial model with a dense set of points. Figure 5 shows the reconstruction results from conjugate light rays intersected by the stereo images with the largest baseline after tracking, while the surface shown in Figure 6 was reconstructed only using one pair of adjacent images. Compared with the noisy mesh produced without a tracking process in Figure 6, the accuracy of the output meshes in Figure 5 is much superior.

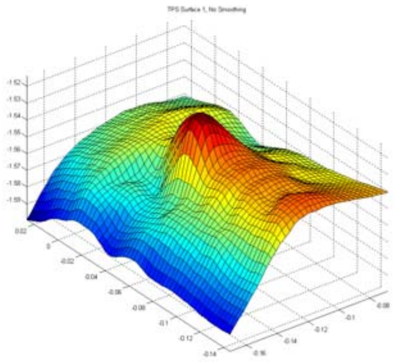


Figure 5. Surfaces reconstructed using stereo-pair with a wide baseline.

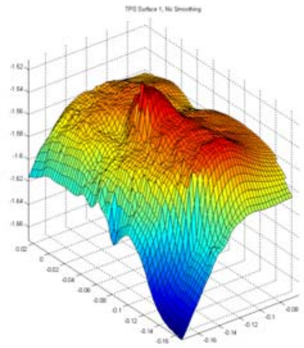


Figure 6. Surfaces reconstructed using stereo-pair with a short baseline.

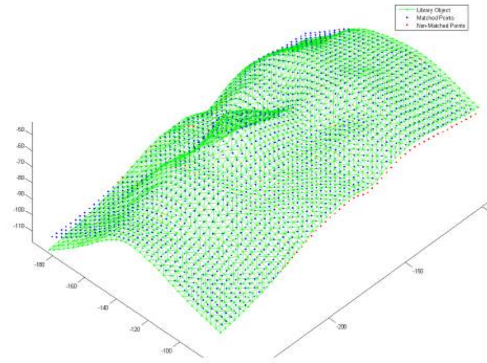


Figure 7. Co-registered facial models using David_1 and David_2 with 93.902% matched points. The green mesh represents the facial surface of David_1, and the points represent the facial surface of David_2 (blue: matches, red: non-matches).

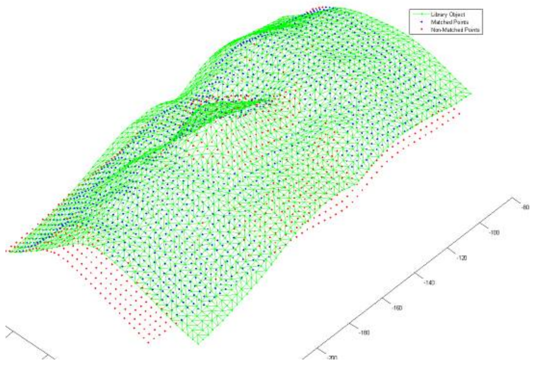
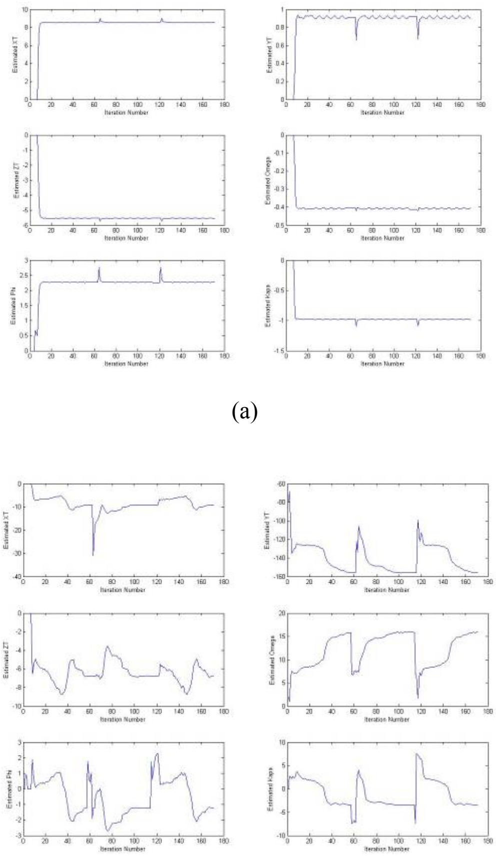


Figure 8. Co-registered facial models using David_1 and Ivan_1 with 79.458% of matched points. The green mesh represents the surface from David_1, and the points represent the surface from Ivan_1 (blue: matches, red: non-matches).



(a)

(b)

Figure 9. Iterative solutions for the transformation parameters in (a). ” David_2 vs. David_1” and (b).” Ivan_1 vs. David_1”.

For verification and identification purposes, acquired facial models were compared to available models. The comparison procedure required the co-registration of the facial models to a common reference frame and the matching of the registered models. The registration algorithm which combines the MIHT and ICPatch, was used to register and match the two facial models. Figure 7 and Figure 8 show the co-registration of the two facial models for a reference facial model David_1.

The scale factor of the transformation function was fixed in the experiments because the scale should not substantially differ between data acquisition epochs. In the experiment ” David_2 vs. David_1”, the iterative solution (Figure 9(a)) for the transformation parameters revealed smooth and quick convergence. The estimated RMS of the normal distance between matching surface elements following the registration procedure was 0.629 mm. A large percentage of the points (Fig. 7) were classified as matches (93.902%) with the non-matches mainly occurring around the edges of the facial models. The results showed a high quality of fit between two surfaces. In the experiment ” Ivan_1 vs. David_1”, the iterative solution (Figure 9(b)) for the transformation parameters did not exhibit a smooth and rapid convergence. The RMS of the normal distances between the matched elements was 1.715 mm. The procedure achieved 79.458% of matched points (Fig. 8). Compared with ” David_2 vs. David_1”, the results here showed a lower quality of fit between two surfaces.

4. CONCLUSION

This research presented a system for automated matching of facial models using a low-cost photogrammetric stereo system with pattern projection. The experimental results showed that, after calibration, low-cost digital cameras can reconstruct 3D facial models, and then facial model registration can be performed effectively. The proposed system has great potential for various applications such as surveillance, plastic surgery and personal verification. Because seven locations were sequentially imaged, movement of the subject may have produced errors. In the future, development of a system with multiple cameras, which can be controlled to capture images simultaneously, would improve both accuracy and matching without motion effect. Some limitations of the system are acknowledged. Certain surface details such as the hair and beard cannot be properly acquired by this system.

REFERENCES

Achermann, B., Jiang, X., and Bunke, H., 1997. Face recognition using range images, *Proc. Int. Conf. on Virtual Systems and MultiMedia*, pp.129-136.

Beumier, C., and Acheroy, M., 1995. Automatic Face Identification, *Applications of Digital Image Processing XVIII, SPIE, vol. 2564*, pp. 311-323.

Chellappa, R., Wilson, C.L., and Sirohey S., 1995. Human and Machine Recognition of Faces: A Survey, *Proceedings of the IEEE* 83 (5), pp. 705-741.

Cheng, R.W.T. and Habib, A., 2007. Stereo photogrammetry for generating and matching facial models. *Optical Engineering* 46(6), pp. 67203-67213

Jarvis, R., 1993. Range Sensing for Computer Vision, in: Three-Dimensional Object Recognition Systems, Jain, K. and Flynn, P.J., eds., Advances in Image Communication, vol 1, Elsevier Science Publisher, pp. 17-56.

Habib, A., and Morgan, M, 2003. Automatic calibration of low-cost digital cameras. *Optical Engineering*, 42(4), pp.948-955.

Habib, A., Pullivelli, A., Mitishita, E., Ghanma, M., and Kim E., 2006. Stability analysis of low-cost digital cameras for aerial mapping using different geo-referencing techniques. *Journal of Photogrammetric Record*, 21(113), pp29-43.

Harris C., and Stephens M.J., 1988. A combined corner and edge detector. *Alvey Vision Conference*, pp. 147–152.

Lao, S., Sumi, Y., Kawade, M., and Tomita, F., 2000. 3D template matching for pose invariant face recognition using 3D facial model built with iso-luminance line based stereo vision, *Proc. ICPR, vol.2*, pp.911-916.

Schmid, C., Mohr, R. and Bauckhage C., 2000. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2), pp. 151–172

Yacoob, Y., and Davis, L.S., 1994. Labeling of human face components from range data, *CVGIP: Image Understanding*, 60(2), pp.168-17