Facial Shape Recovery By Feature Driven Stereo Analysis

Xiangyang Ju Andrew Naftel
Department Physics, Astronomy and Mathematics
University of Central Lancashire
Preston PR1 2HE U.K.
x.ju@uclan.ac.uk
a.j.naftel@uclan.ac.uk

Abstract

This paper describes a hybrid least squares stereo matching and feature detection approach for the accurate metric recovery of 3-D facial shape and shape change. We have incorporated an active shape model to enable *a priori* image segmentation, facilitate adaptive window stereo matching and to locate seed point correspondences on which full field disparity estimates can be reliably initialised. Results are presented which demonstrate that 3D facial reconstruction detail can be enhanced through thin plate spline disparity map interpolation. The proposed method is sufficiently sensitive to detect localised shape changes induced by simulated surgical correction of abnormal jaw relationships.

Keywords: Stereo, Active Shape Model, Facial Features, 3D Shape

1 Introduction

There has been increasing interest in the analysis of three-dimensional facial surface shape and shape change by oral and maxillofacial surgeons. Recent work [1] has shown that clinicians can benefit from a 3-D analysis of facial morphological changes which may occur as a result of growth processes, pathology or from soft tissue movements following corrective treatment. This requirement places exacting demands on the metric accuracy of 3-D reconstruction as well as the density of recovered depth maps, since the changes to be detected are usually small. Imaging techniques such as laser ranging [1], structured lighting [2] and stereo vision [3,4] have all been used for recovering facial depth information. However, it is our belief that the potentials of a stereo approach have yet to be fully exploited in this application since to date automatic facial feature detection has not been used directly to drive the correspondence search. In this paper we present a method that improves the performance of stereo matching in the metric recovery of facial shape by using active shape model detected facial features.

Two broad classes of techniques are used in stereo vision; area-based [5-9] and feature-based matching [10-14]. A feature-based approach provides a sparse disparity map by matching at texture-rich points only. By itself, it does not provide sufficient information for facial recovery. Area-based stereo matching chooses an intensity

BMVC 1999 doi:10.5244/C.13.53

window in one image and then searches for its homologous counterpart in the other image. It enables a correspondence search at any position of interest.

In this paper, we propose a stereo matching method, that combines an area-based approach with facial feature detection, to improve the performance of least squares stereo matching (LSSM) [5-8] by providing good initial disparity estimates and reducing the geometrical image differences arising from close range image capture. These differences can significantly degrade the performance of correlation–type algorithms in facial surface reconstruction [2]. Firstly, we locate the face features (eyebrows, eyelashes, nose, mouth and face boundary) using an active shape model (ASM)[21]. Secondly, we establish a few representative point correspondences in the neighbourhood of the detected features in order to estimate full-field disparity estimates between the stereo images using thin plate spline interpolation [23,24]. The success of LSSM is heavily dependant on the provision of good initial disparity estimates to guarantee convergence of the iterative search to the global minimum of the least squares criterion. We present some results to show that the use of ASM and disparity map interpolation improves the performance of ordinary LSSM and results in the production of dense 3D facial depth maps and reliable shape recovery. Finally, we extract the facial midline profile from stereo reconstructions of a patient before and after a simulated surgical correction to validate the intended application of the proposed method.

2 Least Squares Stereo Matching

Least squares stereo matching (LSSM) is used to determine stereo correspondence by minimising grey level intensity differences between homologous regions of the two images, whilst taking due account of geometric and radiometric image distortions. The optimal transformation is found between search and target window by a guided sequence of iterative steps which continually update the position of the search window.

Let (x_l, y_l) and (x_r, y_r) be the co-ordinates of representative pixels in the two homologous windows of left and right images to be matched, and let $I_l(x_l, y_l)$ and $I_r(x_r, y_r)$ denote their grey level intensities. Assuming both images are corrupted by additive noise $n_l(x_l, y_l)$ and $n_r(x_r, y_r)$, the generalised observation equation is

$$I_r(x_r, y_r) + n_r(x_r, y_r) = I_l(x_l, y_l) + n_l(x_l, y_l) \tag{1}$$

The geometrical differences in the images of the viewed scene can be modelled as an affine linear transformation f_x, f_y to account for scale, rotation and shearing effects

$$\begin{aligned} x_l &\equiv f_x(x_r, y_r) = a_1 x_r + a_2 y_r + a_3 \\ y_l &\equiv f_y(x_r, y_r) = a_4 x_r + a_5 y_r + a_6 \end{aligned} \tag{2}$$

A linear transformation function H_R is used to model radiometric effects such as brightness gain and contrast scale

$$I_r(x_r, y_r) \equiv H_R(I_l(x_l, y_l)) = b_1 + b_2 I_l(x_l, y_l) \tag{3}$$

In the presence of noise $I_r + n_r = H_R(I_l + n_l)$, and given that H_R is linear, the residual error $v(x,y)$ resulting from unmodelled noise is

$$v(x_r, y_r) = n_r - H_R(n_l) = H_R(I_l) - I_r \tag{4}$$

Expanding to a first order approximation

$$\begin{aligned} & v(x_r, y_r) + I_r = H_R(I_l(f_x(x_r, y_r), f_y(x_r, y_r))) \\ & \cong H_R^0(I_l) + \sum_{i=1}^{6} \left\{ \frac{\partial H_R}{\partial f_x} \frac{\partial f_x}{\partial a_i} da_i + \frac{\partial H_R}{\partial f_y} \frac{\partial f_y}{\partial a_i} da_i \right\} \\ & \quad + \sum_{j=1}^{2} \frac{\partial H_R}{\partial b_j} db_j \end{aligned} \tag{5}$$

Denoting the set of geometric ($a_{i=1..6}$) and radiometric ($b_{j=1..2}$) transformation parameters by $\boldsymbol{\beta}$, we choose initial values

$$\beta^0 = (1,0, d_x^0, 0,1, d_y^0, 0,1)^T \tag{6}$$

In LSSM, it is necessary to provide initial disparity estimates d_x^0 and d_y^0 for specified features between search and target windows. From equation (5), the difference in intensity between the candidate windows $\Delta I(x,y) = I_r - H_R^0(I_l)$ is thus given by

$$\begin{aligned} & \Delta I(x, y) + v(x, y) = \\ & \frac{\partial I_r}{\partial x} x_l da_1 + \frac{\partial I_r}{\partial x} y_l da_2 + \frac{\partial I_r}{\partial x} da_3 \\ & + \frac{\partial I_r}{\partial y} x_l da_4 + \frac{\partial I_r}{\partial y} y_l da_5 + \frac{\partial I_r}{\partial y} da_6 + db_1 + I_l db_2 \end{aligned} \tag{7}$$

where $\partial I_r/\partial x$ and $\partial I_r/\partial y$ are the derivatives of the intensity function with respect to the x, y directions and $v(x,y)$ is the residual error. The observation equation (7) relating the grey level difference $\Delta I(x,y)$ with the transformation parameter $\boldsymbol{\beta}$ can be solved so as to minimise $\|v(x,y)\|^2$ over all pixels in the candidate windows. In $v = \mathbf{A}\delta_{\boldsymbol{\beta}} - \Delta I$, $\delta_{\boldsymbol{\beta}}$ are the parameter corrections and $\mathbf{A}$ the coefficient matrix. Each corresponding pair of pixels in matched windows generates on observation equation (7), and for an over-determined system a solution of the parameter correction $\delta_{\boldsymbol{\beta}}$ is obtained as a weighted least squares estimate

$$\delta_\beta = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \Delta \mathbf{I} \tag{8}$$

with a weight matrix $\mathbf{W}$ for data snooping [6]. The updated parameters $\boldsymbol{\beta}$ are then calculated by iterative updating using a bilinear sampling procedure to calculate the

new search window. After each iteration, the transformed search array is resampled to obtain an interpolated set of pixel intensities around (x_r,y_r). Once convergence has been achieved, the centre of the search and target arrays are chosen as corresponding points. The initial disparity estimates for a pair of candidate match windows is estimated by selection of control points (x_{lc}, y_{lc}) and (x_{rc}, y_{rc}) in the two images and setting $d_{xc}^{0}=x_{rc}-x_{lc}$, and $d_{yc}^{0}=y_{rc}-y_{lc}$. We explain how these seed points are generated in the next section.

In order to improve the performance of LSSM, various refinements were incorporated including a multi-resolution algorithm [15-17] to enlarge its search range, outlier detection [6] was applied to suppress the effects of pixels with large intensity differences and bi-direction verification was used to eliminate obvious false matches.

3 Facial Feature Detection

Facial features in stereo images are *a priori* in correspondence because they relate to the same physical features on the actual face. Hence, automatic identification of facial feature points in stereo images will result in good initial disparity estimates (d_{xc}^{0}, d_{yc}^{0}) for these seed points. Facial feature recognition techniques are the subject of much current research [19-22]. Given that 'full frontal' facial images are necessarily used in our work, we have tested the Active Shape Model technique proposed in [21,22].

Active Shape Model (ASM) is a statistically based technique for building compact models of the shape and appearance of almost any flexible object and they can be used when searching an image for a new example of the object. The shape of a set of image features can be described by a labelled set of landmark points. The points can be found automatically once the feature outlines have been defined manually on a set of images (Figure 1). A Point Distribution Model (PDM) can then be generated from a training set of examples containing sufficient variation by performing a principal component analysis on the aligned members of the training set. The analysis results in a mean shape and a set of basis vectors representing the main modes of shape variation. The mathematical details are now well established and may be found in [21]. Additionally, the local grey-level profile statistics are analysed along directions normal to the outline at each labelled point. In principle, these models can be used to locate instances of features in new images. Essentially, the PDM is placed on the new image and then allowed to geometrically deform in an iterative fashion until it best matches the presented features.

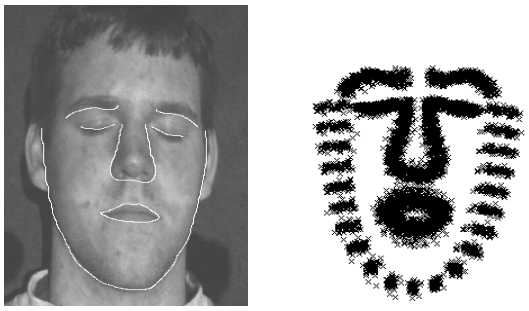


Figure 1. An example PDM and the aligned shapes in the training set

4 Prediction of Disparity Estimates

The detected features in the stereo images enabled us to define the seed points, such as the end points of eyebrows, eyelashes, 5 points on nose, corner points of mouth and middle points of lips, and 5 points on the facial border. However, these points were not yet in exact correspondence and cross-correlation was applied to establish more precise matches. Due to occlusion effects, points on the facial border were difficult to assign correspondences. Searching was directed towards the centre of the face until satisfactory correspondences were found. Final correspondence points are shown in figure 4.

Given an accurate, albeit sparse, set of point correspondences for facial features it is now possible to establish a thin plate spline interpolation function $\mathbf{f}(x,y)$ [23, 24] which maps each point (x_i^R, y_i^R) in the right image to its corresponding point (x_i^L, y_i^L) in the left image (Figure 4). The resulting mapping function $\mathbf{f}(x,y) = [f_x(x,y), f_y(x,y)]$ is vector-valued, one for the x-coordinate and one for the y-coordinate and has the general form

$$\mathbf{f}(x,y) = \mathbf{c}_1 + \mathbf{c}_x x + \mathbf{c}_y y + \sum_{i=1}^{N} \mathbf{w}_i r_i^2 \log r_i \qquad (15)$$

where $r_i^2 = (x - x_i^R)^2 + (y - y_i^R)^2$. Here, we determine a set of coefficients

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}_1 & \dots & \mathbf{w}_N & \mathbf{c}_1 & \mathbf{c}_x & \mathbf{c}_y \end{pmatrix}^T$$

such that

$$\mathbf{f}(x,y) \equiv [f_x(x_i^R, y_i^R), f_y(x_i^R, y_i^R)] = [x_i^L, y_i^L] \qquad (16)$$

The effect of using a thin plate spline is to produce a smooth disparity mapping function which interpolates between the sparse point correspondences. This is used to provide the initial disparity estimates d_x^0 and d_y^0 in equation (6).

5 Results and Discussion

5.1 Metric Reconstruction Accuracy

To establish the metric reconstruction accuracy of the stereo rig and validate the proposed matching algorithm, we tested its performance on images of calibration targets with known ground truth. We compared the disparity estimates obtained by LSSM with an accurate manually-assisted centroid detection procedure which for the purposes of the experiment were assumed 'error-free'. The differences in disparity estimates between the two approaches are displayed in Figure 2a. The rms errors in x and y directions are 0.111 and 0.064 pixels respectively. The reconstructed spatial coordinates of the target points were then compared using LSSM estimated correspondences with centroid detection. Stereo disparity estimates were then converted into 3D world coordinates through a modified form of Tsai's photogrammetric calibration technique [24]. The differences in 3D reconstructions are displayed in Figure 2b. The maximum error in depth (z-direction) is 1.824 mm and the rms errors are 0.116, 0.152 and 0.524mm in x, y and z direction respectively.

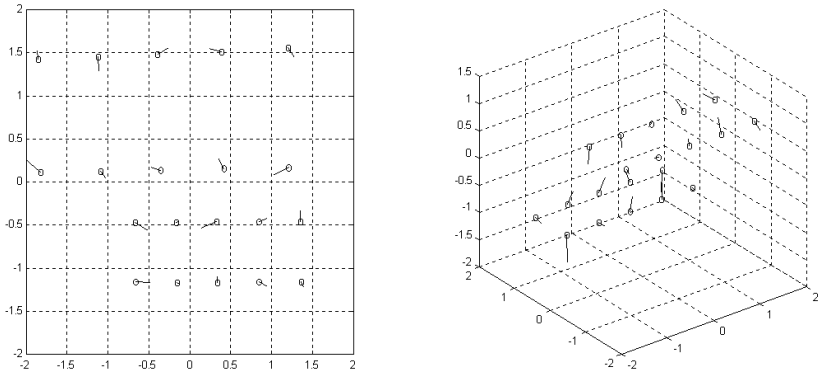
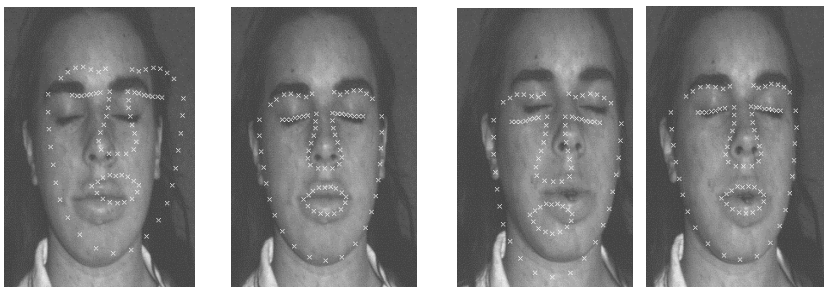


Figure 2. a) Disparity error vectors (magnification 200) for calibration targets (o)
b) 3D error vectors (magnification 100) at the calibration targets

5.2 Stereo Feature Detection using Active Shape Models

We collected frontal view facial stereo image pairs of 61 patients (aged 10 to 26 years; mean 16 years) attending a weekly orthodontic clinic at the Royal Preston Hospital. We randomly selected 20 patients to train the ASM. The shape examples were normalised and aligned by the Procrustes procedure; then the mean shape and the modes of the shape variation were analysed. At the same time, the grey profiles along the normal direction to the shape at all shape points were collected and their mean and covariance matrices were calculated. For a typical trial facial image pair, the ASM located the face features in about 50 iterations (Figure 3). To simulate the effect of 3D shape change induced by surgical correction to correct abnormal jaw relationships, stereo images of all patients were obtained before and after insertion of an orthodontic bite block



left stereopair before insertion left stereopair after insertion

Figure 3. Initial and ASM-detected features in the stereo images

Twenty-two seed points were selected from the detected feature set and adjusted by a cross-correlation matching to establish a set of sparse point correspondences (Figure 4). The homologous points were then used to establish a thin plate spline mapping for full field disparity interpolation.

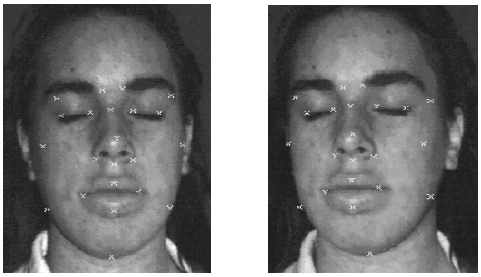


Figure 4. ASM-assisted point correspondences

LSSM can now be assisted by knowledge gained from the prior feature detection stage using the disparities interpolated from the sparse point correspondences. Due to greater texture variation, a smaller window size (9x9 pixels) is used in regions of the eyes, nose and mouth. A larger window size (27x27) is indicated for the smooth textured parts such as the forehead and cheeks.

5.3 Active Shape Model-Assisted Stereo

We have compared the performance of LSSM with and without ASM assistance. Using comparable search regions and window sizes in different face regions, we can observe the effect of the interpolated disparity estimates on the histogram of rms errors of matches (Figure 5). Without ASM assistance, the distribution of errors has a large tail (Figure 5a) indicating many unreliable matches (having larger rms errors). When LSSM is ASM assisted, there are smaller numbers of unreliable matches although the overall rms error is virtually unchanged (marginally decreased from 0.299 pixels to 0.279 pixels). However, the number of successful matches increased from 57.2 per cent to 72.4 per cent. A successful match has a consistent disparity in the bi-direction search and converges before 20 iterations. More successful matches lead to denser recovery of 3D facial depth map. This improvement is thought to justify the extra computational burden.

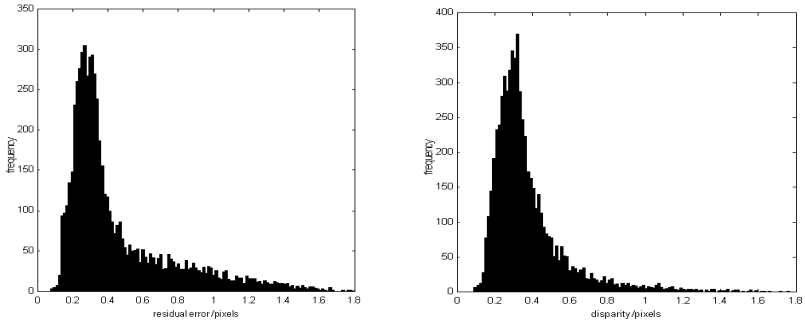
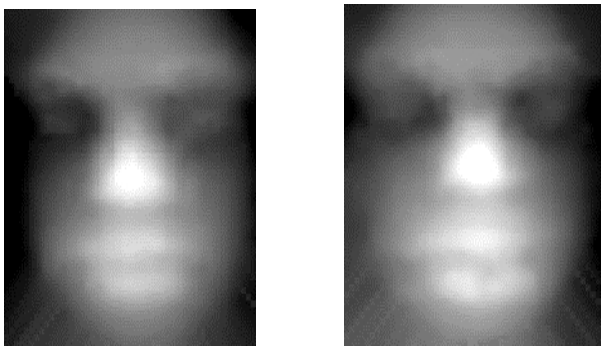


Figure 5. Comparison of the rms error distributions for point matches (a) without ASM assistance, (b) with ASM assistance

5.4 3D reconstruction and shape change detection

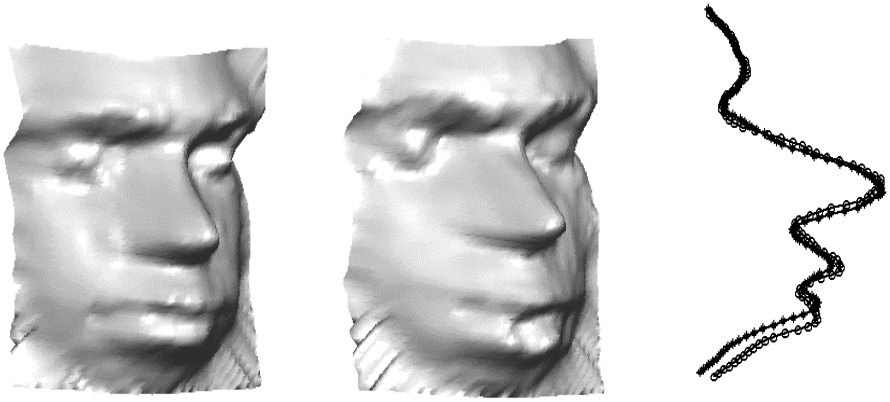
In Figures 6 and 7 we present 3D reconstruction results for a patient before and after

the simulated surgical correction. Facial midline profiles (Figure 7) have been extracted by a symmetry plane registration method [24] to highlight the shape change induced in the jaw. The midline profiles were first Procrustes registered using the unchanged upper portion. We have shown the proposed method is sufficiently sensitive to measure the shape change. The results obtained can then be further subjected to landmark driven morphometric analysis for statistical analysis of group differences [23].



a) b)

Figure 6. The disparity maps a) before, and b) after surgical correction



a) b) c)

Figure 7. 3D reconstruction a) before, and b) after surgical correction
c) Comparison of the 2D midline profiles before (*) and after (o) correction

6 Conclusion

In this paper we have proposed a hybrid stereo matching and feature detection approach for the accurate metric recovery of 3-D facial shape and shape change. This has been achieved using an active shape model to first detect the main features in a facial image and then to provide full field disparity estimates through a thin plate spline interpolation to initialise a least squares stereo algorithm. The advantages are a reduced search space for correspondence and an increase in the likelihood of successful matches being found. Results indicate that facial detail in the dense disparity map can

be enhanced through thin plate spline interpolation. In addition, the active shape model enabled *a priori* image segmentation to conduct adaptive window stereo matching and to locate seed point correspondences on which full field disparity estimates can be reliably initialised. The method is sufficiently sensitive to detect localised shape changes induced by simulated surgical correction of abnormal jaw relationships.

References

[1] McCance A M, Moss J P, Fright W R, Linney A D, James D R 1997 Three-dimensional Analysis Techniques - Part 2: Laser Scanning: A Quantitative Three-dimensional Soft-Tissue Analysis Using a Color-Coding System. *Cleft Palate-Craniofacial Journal*, **34**(1), 46-51.

[2] Gruen A and Baltsavias E 1988 Automatic 3-D Measurement of Human Faces with CCD-cameras. *SPIE Vol* **1030**. 106-116.

[3] Ayoub A F, Siebert P, Moos, K F et al. 1998 A Vision-based Three-Dimensional Capture System for Maxillofacial Assessment and Surgical Planning. *British Journal of Oral and Maxillofacial Surgery* **36**, 353-357.

[4] Bowskill J, Baldock C and Booth P W 1997 Measuring Facial Swelling Using Three-dimensional Imaging. *Medical Informatics* **22**(2), 155-164.

[5] Pertl A 1985 Digital image Correlation With An Analytical Plotter. *Photogrammetria.* **40**, 9-19.

[6] Rosenholm D 1987 Least Squares Stereo Matching Method: Some Experimental Results. *Photogrammetry Record* **12**(70), 493-512.

[7] O'Neill M and Denos M 1996 Automated System for Coarse-to-fine Pyramidal Area Correlation Stereo Matching. *Image and Vision Computing* **14**, 225-236.

[8] Gruen A W 1985 Adaptive Least Squares Correlation: A Powerful Image Matching Technique. *S Afr J of Photogrammetry, Remote Sensing and Cartography* **14**(3), 175-187.

[9] Kanade T and Okutomi M 1994 A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **16**(9), 920-932.

[10] Förstner W 1986 A feature based correspondence algorithm for image matching. *International Archives of Photogrammetry and Remote Sensing* **26**(3), 150-166.

[11] Lew M L, Huang T S and Wong K 1994 Learning and Feature Selection in Stereo Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(9)

869-881.

[12] Kim D H, Koo K B, Choi W Y and Park R H 1996 Stereo Matching Using Hierarchical Features for Robotic Applications. *Advanced Robotics* **10**(1), 1-14.

[13] Venkateswar V and Chellappa R 1995 Hierarchical Stereo and Motion Correspondence Using Feature Groupings. *International Journal of Computer Vision* **15**, 245-269.

[14] Hoff W 1989 Surface from Stereo: Integrating Feature Matching, Disparity Estimation, and Contour Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* **11**(2), 121-136.

[15] Tarel J P and Boujemaa N 1999 A Coarse to Fine 3D Registration Method Based on Robust Fussy Clustering. *Computer Vision And Image Universtanding.* **73**(1), 14-28.

[16] Edwards J L and Murase H 1998 Coarse-to-fine Adaptive Masks for Appearance Matching of Occluded Scences. *Matchine Vision And Applications.***10**(5-6), 232-242.

[17] Bonmassar G and Schwartz E L 1998 Improved Cross-Correlation For Temlate Matching on The Laplacian Pyramic. *Pattern Recognition Letters.* **19**(8), 765-770.

[18] Sobottka K and Pitas I 1997 A Full Automatic Approach to Facial Feature Detection and Tracking. *Lecture Notes in Computer Sciences,* **1206**, **77-84.**

[19] Brunelli R and Poggio T 1993 Face Recognition: Features versus Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(10) 1042-1052.

[20] Yow K C and Cipolla R 1997 Feature -based Human Face Detection. *Image and Vision Computing* **15**, 173-735.

[21] Cootes T F, Hill A, Taylor C J and Haslam J 1994 The Use of Active Shape Models for Locating Structures in Medical Images. *Image and Vision Computing* **12**(6), 355-366.

[22] Edward G J, Lanitis A, Taylor C J and Cootes T F 1998 Statistical Models of Face Images - Improving Specificity. *Image And Vision Computing.* **16**(3), 203-211.

[23] Bookstein F L 1997 Shape and the Information in Medical Images: A Decade of the Morphometric Synthesis. *Computer Vision and Image Understanding* **66**(2), 97-118.

[24] Mao Z and Naftel A J 1997 Improved area-based stereo matching using an image segmentation approach for 3-D analysis of facial soft tissue changes. *Medical Image Understanding and Analysis '97,* University of Oxford, pp.209-212.