14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

3D DETECTION AND MEASUREMENT OF FACIAL SWELLINGS

Anna Paviotti, Nicola Brusco, Guido M. Cortelazzo

Department of Information Engineering
via Gradenigo 6/B, 35131, Padova, Italy
phone: + (39) 0498277774, email: {paviotti,brusco,corte}@dei.unipd.it
web: lttm.dei.unipd.it

ABSTRACT

This work presents a procedure for the 3D detection and measurement of facial swellings, an issue of paramount importance in oral and maxillofacial surgery. The considered problems are the automatic acquisition of face masks of adequate precision, the registration of the 3D masks of the same patient taken at different times during the edema evolution, and the detection and estimate of the edema volumes.

1. INTRODUCTION

The accurate reproduction of post-surgical facial swellings is of paramount importance in oral and maxillofacial surgery. The identification and measurement of facial edemas can be used for the assessment of the effectiveness of a specific pharmacological treatment as well as for the simulation and planning of surgical interventions [1, 2, 3, 4, 5, 6]. There are several methods to obtain three-dimensional measurements of human body surfaces. Among them, 3D active reconstruction methods have recently received much attention [1, 2, 3, 4] due to their high precision, non invasivity and acquisition speed. However, fully automatic procedures for 3D modeling are not very common yet. Harrison et al. [1] used a hand-held laser scanner to acquire the faces of patients who had had their third molar extracted. After obtaining high-resolution 3D views interpolating the row data from the scanner by means of radial basis functions, they registered the partial views by manually aligning antropometric features with a MatLab Toolbox. Yip et al. [2] used a 3-D range camera for similar purposes. In their case, the 3D models of the entire face were automatically built by the acquisition system; on the contrary, the facial masks corresponding to different moments in time were registered by manually adjusting the two images until their sagittal and coronal outlines were aligned in the eyes and forehead region. Hasegawa et al. [3] cleverly skipped the problem of registration by building an acquisition system consisting of two lateral laser scanner and a central CCD camera. In this way, the lateral views could be easily aligned thanks to their pixel-to-pixel correspondences. However, their acquisition system is not very flexible. Kau et al [4] used two optical laser-scanning devices assembled as a stereo pair. Again, the left and right views were manually aligned with the help of a lens; similarly, the registration of facial masks taken at different times was performed by manually aligning five chosen points of the scans.

This work presents a procedure for the 3D detection and quantification of facial swellings and two clinical applications of the proposed method. The 3D model construction process is fully automatic, and so is the registration of different facial masks of the same patient. Therefore, the acquisition procedure can be performed by non specialized personnel and the registration result is independent from the operator's skills. Section 2 describes the acquisition system and the data acquisition protocol. The proposed method for the automatic construction and alignment of the 3D models is outlined in Section 3. The adopted algorithm for volume computation is explained in Section 4. Section 5 presents the results for the considered applications. Section 6 draws some concusions.

	Small Range	Medium Range
Field of view (XY plane)	10×75 mm^2	20×150 mm^2
Field of view (Z axis)	350-450 mm	600-750 mm
Objective focal length	25 mm + 1 mm ring	16 m
XY resolution	0.13 mm	0.26 mm
Z resolution	0.01 mm	0.015 mm
XY uncertainty ($\pm\sigma$)	$\pm$ 0.015 mm	$\pm$ 0.018 mm
Z uncertainty ($\pm\sigma$)	$\pm$ 0.030 mm	$\pm$ 0.050 mm

Table 1: Technical features of the 3-D range camera small and medium range setups.

2. ACQUISITION SYSTEM AND DATA CAPTURE

The acquisition system used in this study was an active system based on a CCD camera and a structured light projector. Depending on the desired acquisition range, the camera objective can be changed obtaining an uncertainty for the z-axis from 0.085 to 0.035 mm. In this study two different ranges were used, namely the medium range to build the 3D model of the entire face and to acquire relatively large regions such as the cheeks, and the small range to focus on specific face districts such as the mouth. The characteristics of the two range camera setups are summarized in Table 1.

A first application of our detection and measurement system was the quantification of facial edemas after a third molar extraction. The acquisitions were made at the Dental Clinic of University of Padova between January and July 2005. 23 patients were scanned before the intervention, soon afterwards, after 2 days and after 7 days. To acquire the 3D model of their face 5 scans were taken, namely a frontal view and two pairs of lateral views. In this way, regions affected and unaffected by edema were both included in every measurement session. As we well see in the following session, this is a crucial step for the alignment of facial masks taken at different times. Every scan took about 5 seconds; less than 5 minutes were necessary to process each patient.

The second experimental trial concerned the detection of the hyeluronic acid filler(Belotero) distribution in facial rejuvenation interventions to mouth, lip-nose or forehead region. Three patients were scanned in January 2006. Two measure-

ment sessions were performed, one before the filler injection and another soon after it. A third acquisition will be done after three months from the intervention in April. The regions affected by the operation were scanned using the small range setup. A first reason for this choice was of course the fact that the regions affected by the operation had a limited extension. A second but equally important motivation was that the volume of the injected filler was at most 500 mm^3, which had been estimated to be the volume resolution of the medium range setup. Moreover, after recalibrating the system for the medium range setup, an acquisition of the entire face (5 scans) was done for each patient. An overall amount of about 8 minutes was required by the whole process.

3. AUTOMATIC CONSTRUCTION AND ALIGNMENT OF THE 3D MODELS

3.1 Construction of a facial mask

In principle three views (a central one and a right and left one) may suffice to build the 3D model of a face. However, since submental and submandibolar region were of great interest in the presented applictions, we chose to take five views of the face, a central one and two pairs of lateral views (one slightly upwards-oriented, the other slightly downwards-oriented).

For a detailed description of all the necessary steps to build a 3D model see [7]. In this section we will report about an automatic procedure to perform pairwise registration, i.e. the alignment of two views by means of Textured Spin-Images (TSI) [8], a modification of the algorithm introduced by Johnson et al. in [9]. Let us consider a surface mesh. An oriented point is defined as the pair formed by the 3D vertex coordinates $\mathbf{p}$ and the surface normal $\mathbf{n}$ at the vertex. Consider a point $\mathbf{x}$ on the 3D surface and define in the following way its coordinates pair (α, β) with respect to the reference system $(\mathbf{p}, \mathbf{n})$ associated to an oriented point: radial coordinate α is the perpendicular distance to the line through the surface normal at the vertex point, elevation coordinate β is the signed perpendicular distance to the tangent plane defined by vertex normal and position. Moreover, let $I(\mathbf{x})$ indicate the luminance value associated to the (R, G, B) chromatic values of the surface texture at $\mathbf{x}$ and let us quantize $I(\mathbf{x})$ into L uniform gray level intervals $l_i, i = 1, 2, ...L$. A textured spin-map with respect to an oriented point is the set of triplets determined by all the points $\mathbf{x}$ of the surface and their intensity $I(\mathbf{x})$ in the following way:

$$\begin{aligned} TSI(\mathbf{x}) \quad &\rightarrow (\alpha, \beta, l_i) = \\ = \quad & (\sqrt{\|\mathbf{x}-\mathbf{p}\|^2 - (\mathbf{n}\cdot(\mathbf{x}-\mathbf{p}))^2}, \mathbf{n}\cdot(\mathbf{x}-\mathbf{p}), l_i) \end{aligned} \tag{1}$$

A TSI is obtained from the textured spin map by accumulating in each bin not just the number of (α, β) coordinates falling in it as in the case of standard spin-images, but also all the discrete gray level values l_i of the (α, β) coordinates falling in the bin. Hence a TSI is essentially the set of spin-images recording the (α, β) values associated to each l_i gray level interval, with an important difference with respect to the computation of standard spin-images: as the bin value for each $i = 1, 2, ...L$ is the sum of all the luminance values l_i falling in the bin, even a single-level TSI ($L = 1$) takes into account texture information. In principle, the larger is the number of intensity intervals L, the more texture information is recorded by the TSI. In practice it was shown that a sort of saturation takes place beyond L=4. With respect to Definition (1), standard spin-images can be considered as a special case of TSI with $L = 0$ (no luminance information).

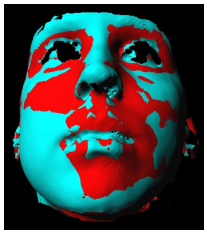


Figure 1: Superposition of two facial masks of the same patient.

TSI inherit some interesting properties from spin-images: first of all, they are invariant with respect to rigid rototranslations as they only depend on the instrinsic surface characteristics. Secondly, TSI of neighbour points are very similar because the features of the local surface area near the point have the greatest impact in TSI formation. The TSI of corresponding points of two partially overlapping 3D views will not be identical because of surface discretization effects and because the two patches share only a portion of their surface. However, if the overlap is substantial (no less than 30% of regions characterized by an adequate presence of geometrical features), corresponding points of the two 3D views located in the common region will have similar TSI. In this way the detection of the common region between two partially overlapping 3D views can be turned into the recognition of the most similar images of two sets of (textured spin) images, a problem for which a number of techniques are available.

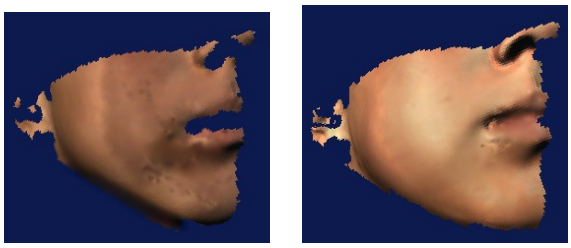
In order to detect corresponding points, a set of putative point correspondences is determined upon the similarity of the TSI. The correspondences are organized within geometrically consistent groups and for each group the rototranslations $(\mathbf{R}, \mathbf{t})$ moving the points of the first 3D view as close as possible to their corresponding points in the second 3D view is determined via Horn's algorithm. The rototranslation $(\mathbf{R}, \mathbf{t})$ giving the widest overlap between the two views is selected. The common region between the two views is taken to be this overlapping area.

The final estimate of the rotation and translation $(\mathbf{R}, \mathbf{t})$ between two 3D views is typically accomplished by the ICP algorithm [10] or by some of its many variants. We chose to adopt the method of [11], which in our case proved to be very robust. This algorithm has the important feature of not operating on a point-to-point correspondence logic as the ICP, since the Fourier transform makes a synthesis of all available spatial information.

3.2 Automatic alignment of facial 3D models taken at different times

The alignment of facial 3D models taken at different times essentially requires to match points among such facial models, exploiting the inclusion of face regions outside the edema such as nose, ears, eyes, etc.

Corresponding points outside the edema regions are robustly detected by textured spin-images. Fig. 1 shows the superposition of a patient's face before and after surgery. In



(a) (b)

Figure 2: A patient's cheek (a) before the intervention (b) after 2 days.

order to do this we defined a reference system associated with the frontal view (which was taken as the 3D model reference at acquisition time) with z-axis passing through the nose (actually through the 3D model point with largest z).

In the case of cheek edemas, the volume computation turned out to be more reliable if only the region affeced by the edema was comsidered. Therefore, as the edema region was located either at the left or at the right lower quadrant of the defined reference system, the parts of the 3D model lying in the other three quadrants were automatically eliminated. Fig. 2 shows portions of the cheek before and after surgery.

4. VOLUME COMPUTATION

The volume inside a closed surface can be readily computed as the sum of the volumes of the thetraedrons formed by adjacent triplets of points of the mesh [12]. Unfortunately, the superposition of cheek portions before and after surgery returns a surface which is only approximately closed. The error is due to the elastic nature of facial tissues. Indeed if one acquires a face without edema at two different times two slightly different surfaces will be obtained. An example of the distances between correspondent points is shown in Fig.5. The volume of their difference is of the order of $500mm^3$, which we consider the precision limit of our volume measurements.

In order to limit the face deformation effects, the edema region was reduced as much as possible by manually removing further nose and neck regions from the 3D local models obtained through the procedure described in subsection 3.2. Small holes due to stray light during acquisitions were filled before volume computation using standard interpolation techniques.

We implemented a very simple integration technique which takes one of the two surfaces as reference and uses small prisms to fill the gap between the two surfaces as shown in Fig. 3, in practice just a numerical implementation of Riemann' s integration. The volume of the surface difference was then evaluated as the signed sum of the volumes of each elementary prism. We considered the signed volumes in the sum and not their absolute values in order to avoid being penalized by the face elasticity effects.

More reliable results were obtained using classical deformable (or parametric) models, also known as snakes [13]. The basic idea of this method is to overlap the two models and place a sphere centered inside the region of interest, formed by the portion of the cheek without edema and the swallen one. The sphere will be hence deformed under the influence of three types of forces: an internal force, a balloon

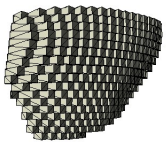
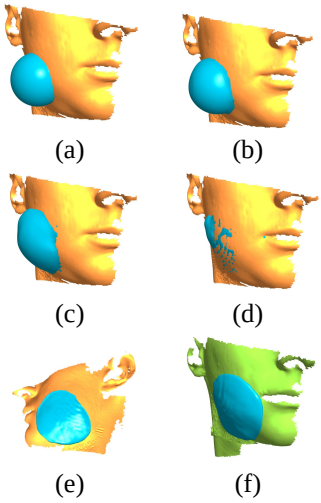


Figure 3: Volume computation with prisms.



(a) (b) (c) (d) (e) (f)

Figure 4: (a)-(d): Snake evolution over the face with edema; (e) Snake inside the face with edema; (f) Snake outside the face without edema.

force and an external one. The first will mantain the surface as smooth as possible, while the last will give it the desired shape, fitting the model to the surface containing the volume under measurement. The balloon force [14] is used to make the snake evolve until it covers all the edema volume (see Fig. 4). The evolution of the model can be formalized as follows:

$$s(0) = s_0 \quad (2)$$

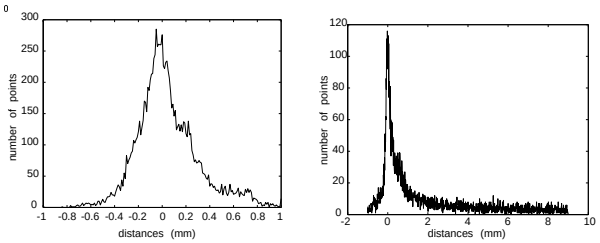
$$\frac{\partial s}{\partial t}(t)(P) = \alpha_{int} \cdot F_{int} + \alpha_{bal} \cdot n + \alpha_{ext} \cdot F_{ext} \quad (3)$$

where n is the outer normal of the surface and

$$F_{int}(P) = \nabla^2 s(P) - \nabla^4 s(P). \quad (4)$$

$\nabla^2 s$, $\nabla^4 s$ respectively represent the Laplacian and Bilaplacian of s in P. As for F_{ext}, instead of using the classical conservative force field we resorted to the Gradient Vector Flow (GVF) introduced in [15]. This vector field is a priori generated from the set of points of both models using a multiresolution-grid to avoid heavy memory requirements. The GVF can be computed through the method proposed in [16]. The position of the initial sphere can be determined by localizing the point P_{max} on the edema for which the distance from the cheek before surgery is maximum and then placing the center of s_0 in the median point of the segment joining P_{max} and its projection on the surface underneath.

We found precision and robustness of this method way superior to those of the above mentioned techniques and certainly adequate to our application.



(a) (b)

Figure 5: Histograms of distances (a) between the models of a patient's cheek without edema considered at two different times (b) between a patient's cheek with and without edema.

Patient	Vol. After intervention [mm^3]	Vol. After 2 days [mm^3]	Vol. After 7 days [mm^3]
D.A.	28684.5	31730.6	25785.5
G.A.	12321.6	39412.7	15751.6
C.A.	17616.0	21180.0	15260.5
C.E.	3113.7	17146.4	3782.8
G.E.	8143.06	21768.4	6752.9
C.E.	9793.7	20637.5	15890.9
R.E.	5387.1	11686.4	4420.3
G.G.	10664.7	18790.7	11469.8
C.G.	11491.9	23508.4	18714.2
V.I.	33023.0	35210.6	20162.0
M.M.	15332.0	46266.9	10568.5
F.M.	16346.9	21167.9	10565.1
M.N.	18797.7	24806.2	20699.3
F.P.	15762.6	49835.8	12654.9
G.S.	18619.8	46591.7	15600.9
G.S.	16679.6	34040.3	21885.4
F.V.	15977.8	25273.4	11476.8
C.D.	11800.3	33362.1	9451.5
B.M.	17859.9	49765.6	15560.0
P.A.	5909.3	33572.9	12675.9

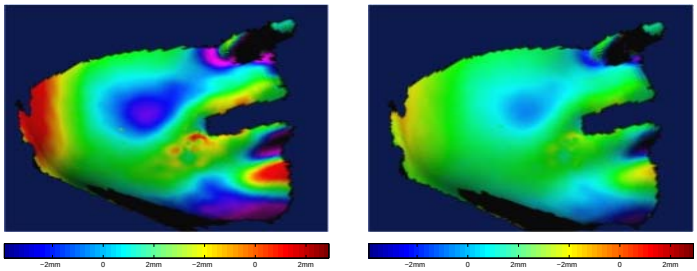
Table 2: Edema volumes at the three considered post-intervention times.

Fig. 5 shows a histogram of the distances between corresponding points on the two cheeks before and after surgery. The edema presence is clearly indicated by the histogram behaviour, rather different from that of Fig.5a). The maximum distance shown by the histogram is more significant than its mean and standard deviation in order to characterize the edema. Indeed, mean and standard deviation depend very much on the area selected for the volume computation, while maximum distances don't.

5. EXPERIMENTAL RESULTS

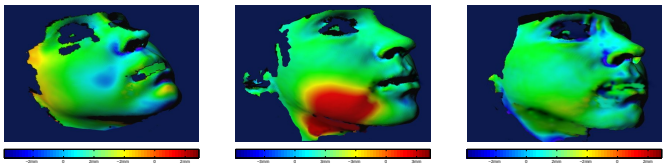
5.1 Quantification of facial edemas

Table 2 shows the volume computation results for the 23 patients of the Dental Clinic of the University of Padova. After calculating the volume mean values for each gender, it was clear how the edema volume is much higher for males, see Table 3. Namely, soon after the intervention males exhibited an edema volume 17% higher than females; 2 days afterwards the gap amounted to 16.2%; 7 days afterwards to 15%. The most interesting result of this study was that the presence of a residual swelling after 7 days was highlighted. The edema volume after a week from the intervention, often unnoticed by the patients themselvs, is about equal to that calculated soon after surgery for both genders.



(a) (b)

Figure 6: Edema map at two different scales.



(a) (b)

Figure 7: Edema map evolution: (a) after surgery (b) 2 days afterwards (c) 3 days afterwards.

Since edemas concern surfaces their presentation is naturally a 2D map. Map information can be further synthesized in few parameters such as maximum edema height, edema heights at specified craniometric locations, or edema volume as a convenient single parameter.

We realized that the appropriate minimum scale range to display the edema maps was $[-2mm, 2mm]$. Fig. 6 shows a patient's cheek before surgery at two different scales, namely $[-2mm, 2mm]$ and $[-1mm, 1mm]$. The map at scale $[-1mm, 1mm]$ shows artefacts related to face elasticity due to data acquisition and registration, giving rise to a minimum volume precision of $500mm^3$.

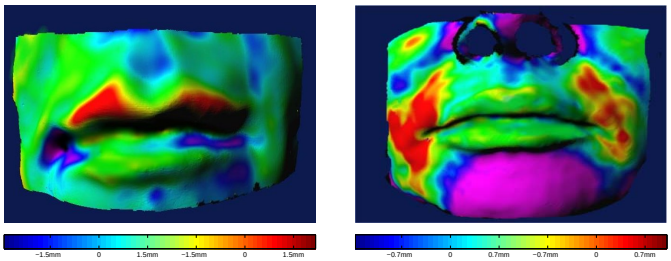
Fig. 7 shows some edema maps right after surgery and at the second and seventh day after the operation. It can be seen than in presence of considerable swelling the adequate display scale varies from $[-3mm, 3mm]$ to $[-5mm, 5mm]$.

5.2 Detection of hyaluronic acid filler distribution in face rejuvenation interventions

In this case, the major problem was that the small amount of injected filler seemed to hamper the detection of the hyaluronic acid filler distribution by the range camera. As can be seen in Fig.8, it turned out that the small range setup was perfectly adequate for the task. Fig.8(a) and 8(b) show the hyaluronic acid filler distribution for an intervention of lip augmentation and one to a patient's lip-nose region respectively. In this application, the appropriate map scale was smaller than for cheek edemas, varying from 0.7 to 1.5 mm.

6. CONCLUSIONS

This work reports on a procedure for the detection and the quantitative assessment of facial swellings. A first problem posed by this task concerned the acquisition of a 3D mask of the face and the registration of the 3D masks of the same patient taken at different times during the edema evolution. Moreover, a reliable method to estimate the volume of the



(a) (b)

Figure 8: Edema maps for (a) a lip augmentation intervention (b) an intervention to the lip-nose region.

Gender	Vol. After intervention [mm^3]	Vol. After 2 days [mm^3]	Vol. After 7 days[mm^3]
Male	15873.7	32827.8	14960,9
Female	13190.5	27188.1	12728.8

Table 3: Edema volume mean values for males and females.

difference between the obtained surfaces was searched for. The possibility of detecting the distribution of small volumes of injected fluid was then evaluated. The previous sections presented effective solutions for the above tasks.

We are currently defining a procedure for the metrological validation of our algorithm. The validation process will aim at quantifying both the error in the measurement of known volumes [1, 2] and the error arising from human tissue elasticity and patients' unperceived movements, i.e. the "pose variability"[1].

Further research will also address the definition of surface parametric models capable of capturing facial deformations in few parameters. Such parameters would be very useful for edema classification purposes.

A third area in need of investigation is the realization of *ad hoc* 3D face digitizers as accurate as the general purpose structured light system we used but less expensive, possibly more robust for data acquisition and handier.

7. ACKNOWLEDGEMENTS

We would like to acknowledge Alvise Gasparini, Antonio Gracco, Giampiero Cordioli, M.D., and Prof. M. Berengo of the Dental Clinic of the University of Padova for the organization of the measurement sessions of cheek edema volumes. We would like to thank Licia Armiraglio, M.D., Alessandra Camporese, M.D., Merz and Orlian for organizing the experimental trials on face rejuvenation interventions. We are also grateful to Gabriele Bardella, Enrico Zanella and Luca Ballan for their help with the acquisitions. This research was partially funded by "Progetto Ricerca di Ateneo CPDA024157-2002".

REFERENCES

[1] J. A. Harrison, M. A. Nixon, W. R. Fright, and L. Snape. "British Journal of Oral and Maxillofacial Surgery," *Use of hand-held scanning in the assessment of facial swelling: A preliminary study*, vol. 42, pp. 8-17, Feb. 2004.

[2] E. Yip, and M. Yoshino. "International Journal of Oral and Maxillofacial Surgery," *Volumetric evaluation of facial swelling utilizing a 3-D range camera*, vol. 33, pp. 179-182, 2004.

[3] K. Hasegawa, K. Hattori, and Y. Sato. "A High Speed Face Measurement System," in *Proc. Vision Interface*, Trois-Riviéres, Canada, May 19-21. 1999, pp.196–202.

[4] C. H. Kau, A. Cronin, P. Durning, A. I. Zhurov, A. Sandham, and S. Richmond . "Orthod Craniofacial Res," *A new method for the 3D measurement of postoperative swelling following orthognathic surgery*, vol. 9, pp. 31-37, 2006.

[5] R. M. Koch, S. H. M. Roth, M. H. Gross, A. P. Zimmermann, and H. F. Sailer. "A framework for facial surgery simulation," in *Proc. 18th Spring Conference on Computer Graphics*, Budmerice, Slovakia, April 24-27. 2002, pp. 33–42.

[6] J. Xia, D. Wang, N. Samman, R. W K. Yeung, and H. Tideman. "Int. J.. Oral Maxillofac. Surg.," *Computer-assisted threedimensional surgical planning and simulation: 3D color facial model generation*, vol. 29, pp. 2-10, 2000.

[7] H. Rushmeier, and F. Bernardini. "Computer Graphics Forum," *The 3D model acquisition pipeline*, vol. 21, pp. 149-172, 2002.

[8] N. Brusco, M. Andreetto, A. Giorgi, and Guido M. Cortelazzo. "3D-Registration by Textured Spin-Images in *Proc. 5th International Conference on 3-D Digital Imaging and Modeling (3DIM05)*, Ottawa, Ontario, Canada, June 13-16, 2005, pp 262-269.

[9] A. E. Johnson and M. Hebert. "IEEE Transactions on Pattern Analysis and Machine Intelligence," *Using Spin-Images for Efficient Multiple Model Recognition in Cluttered Scenes*, vol. 21, pp. 433-449, 1999.

[10] P. J. Besl, and N. D. McKay. "IEEE Transactions on Pattern Analysis and Machine Intelligence," *A method for registration of 3d shapes*, vol. 14, pp. 239-245, 1992.

[11] L. Lucchese, G. Doretto, and G. M. Cortelazzo. "IEEE Transactions on Pattern Analysis and Machine Intelligence," *A frequency domain technique for 3-d view registration*, vol. 1468-1484, 2002.

[12] C. Zhang and T. Chen. "Efficient feature estraction fo 2d/3d objects in mesh representation," in *ICIP*, Thessaloniki, Greece, 2001.

[13] M. Kass, A. Witkin, and D. Terzopoulos. "CVGIP: Image Undersand," *On active contour models and balloons*, vol. 53, pp. 211-218, 1991.

[14] L. D. Cohen. "International Journal of Computer Vision," *Snakes: Active contour models*, vol. 1, pp. 433-449, 1987.

[15] Xu, and Prince. "IEEE Transactions on Image Processing," *Snakes, shapes, and gradient vector flow*, vol. 7, pp. 359-369, 1998.

[16] C. H. Esteban, and F. Schmitt. "Computer Vision and Image Understanding," *Silhouette and stereo fusion for 3d object modeling*, vol. 96, pp. 367-392, 2004.