

# FACIAL SCAN CHANGE DETECTION

Prathap Nair\*, Lifong Zou<sup>†</sup>, Andrea Cavallaro\*

<sup>†</sup> Barts and The London, Queen Mary's School of Medicine and Dentistry

\*Multimedia and Vision Lab

Queen Mary University of London, Mile End Road, London E3 5NS (UK)

andrea.cavallaro@elec.qmul.ac.uk

**Keywords:** facial scans, low-level features, semantic features, 3D object database.

## Abstract

We present a method for quantifying and localising changes in two facial scans of the same person taken at two different time instants. The method is based on rigid registration and semantic feature extraction, followed by discrepancy computation. The proposed method combines the Landmark Transform (LT) method, which is applied on *semantic feature points*, and the Iterative Closest Point (ICP) algorithm, which is performed on *semantic regions*. Finally, the discrepancy between the two scans is computed using the Symmetric Hausdorff distance. Experimental results with both synthetic and real data show the effectiveness of the proposed method which has also been validated by an experienced clinical scientist. Moreover, the method is being used as support in clinical studies on a 3D object database with more than 1000 facial scans.

## 1 Introduction

Advances in data acquisition systems are facilitating the creation of 3D object databases. Facial scans are an important example of 3D objects and are increasingly being used for biometrics and medical imaging applications. The automated analysis and comparison of the scans is desirable for tasks such as object retrieval, access control and evaluation of facial surgery interventions, just to mention a few. Several medical conditions can be determined from the face. Hence there is an interest in accurately relating information in different images for diagnosis and treatment of patients. Although the distinctive symptoms associated with certain disorders are not always obvious to a layperson, an experienced clinical scientist can distinguish them by studying the morphology of the face.

With the advances of laser scanning equipment, capturing 3D surface scans has become a convenient and non-invasive process of capturing the structure of a patient's face at a high resolution. Scans captured at different time instances need to be compared in order to determine changes that might have occurred in the patient's face in between the time of the two scans. The current solution to this problem is to use an interactive technique that employs a semi-automatic registration process. This process requires the operator to specify the reference points or regions to be used for the comparison of the scans. The performance of the system is

highly dependent on the expertise of the operator in choosing the right regions as reference. To remove this dependency on the operator, our goal is to design an automatic system that is capable of automatically detecting relevant points and regions for the analysis of the scans. A commonly used technique to register face laser scans is the Iterative Closest Point (ICP) algorithm [1,2,5,6]. In particular, in [2] a method is proposed that uses an initial coarse registration using salient face points followed by the ICP method on the whole dataset. This method is expected to perform well in an application such as face recognition, where drastic changes between scans under study are not present. Unlike biometrics, such an assumption is not valid in medical applications as various types and magnitudes of changes are expected on two different scans of the same subject. In this paper we present a method that overcomes this limitation.

The paper is organised as follows. Section 2 introduces the data acquisition system for the generation of the facial scan database. In Section 3, low-level and high-level features that are extracted for registration are described. The change detection procedure is explained in Section 4. Section 5 reports the experimental results and the subsequent analysis and validation. Finally, in Section 6 we draw the conclusions.

## 2 Generation of the facial scan database

A common way of representing a facial scan is in the form of a mesh. A scan represented in this form is also referred to as a 2.5D range image. The facial scans that constitute the database under test have been obtained with an optical surface scanning of the face [8]. This system of data collection is non-hazardous and non-invasive. It collects 60,000 points over the facial surface in 10 seconds: approximately 250 facial profiles are recorded on each subject and the 300 data points within each profile are then imaged by faceting and Gouraud shading [9] of the polygons to give a realistic picture of facial form (Figure 1). The acquisition system is depicted in Figure 2.

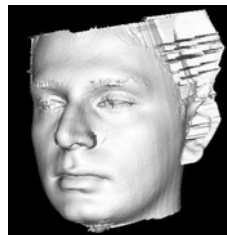


Figure 1. Visualisation of an optical surface scan using Gouraud shading

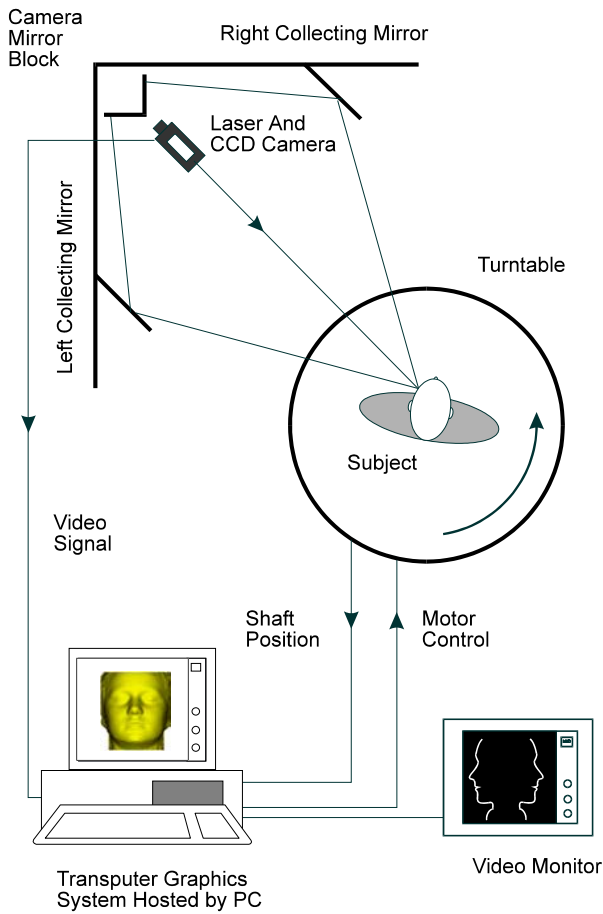


Figure 2. The data acquisition set-up used to generate the facial scan database

### 3 Feature extraction

Registration and change detection are based on low-level as well as high-level features that are extracted from the facial scans, as described below.

#### 3.1 Low-level feature map

In order to characterise the curvature property of each point on a face, two features maps are computed, namely the shape index and the curvedness index [3]. These features maps are derived based on the principal curvature values at the vertices of the surface mesh. The maximum and minimum of the normal curvature at a given point on a surface,  $\kappa_1$  and  $\kappa_2$ , are called the *principal curvatures*. Based on the principal curvatures, the shape index,  $\sigma$ , at a point  $P$ , is defined as

$$\sigma(P) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left( \frac{\kappa_1(P) + \kappa_2(P)}{\kappa_1(P) - \kappa_2(P)} \right), \quad (1)$$

with  $\kappa_1 \geq \kappa_2$  and  $\sigma$  in  $[0, 1]$ . A planar surface has an indeterminate shape index at all its points, since the principal

curvatures  $\kappa_1 = \kappa_2 = 0$ . The feature map generated by  $\sigma$  can describe subtle shape variations from concave to saddle to convex, and provides a continuous scale between salient shapes. However,  $\sigma$  does not give an indication of the scale of curvature present in the shapes. For this reason, an additional feature is introduced,  $\gamma$ , the curvedness of a surface. The curvedness of a surface at a point  $P$ , is defined as

$$\gamma(P) = \frac{\sqrt{(\kappa_1^2(P) + \kappa_2^2(P))}}{2}. \quad (2)$$

The feature maps generated by  $\sigma$  and  $\gamma$  are computed after an iterative Laplacian smoothing. The choice of the number of iterations,  $\alpha$ , is discussed in Section 3.2. A comparison between the feature maps generated with a smoothed and non-smoothed surface is shown in Figure 3.

Based on  $\sigma$  and  $\gamma$ , higher level features based on a priori knowledge of the face shape are introduced as described in the next section.

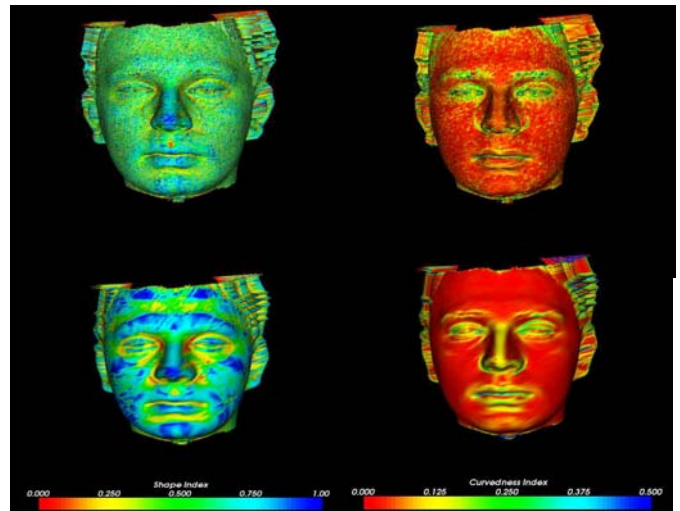


Figure 3. Comparison between the feature maps generated with (top) a non-smoothed surface and (bottom) a smoothed surface. Left: shape index. Right: curvedness index.

#### 3.2 High-level features

The curvedness index and the shape index can be used to characterise the different semantic feature points, or landmarks, in a face. Using  $\sigma$ , for instance, regions such as the nose (which resemble a spherical cap) can be distinguished from areas like the inner eye region (which resembles a spherical cup). The valley formed at the junction of the lips can also be easily identified with the use of this feature. Using  $\gamma$ , regions such as the nose tip, chin, eyebrow (which all resemble a spherical cap) can be distinguished from each other. In addition to  $\sigma$  and  $\gamma$ , depth information,  $z$ , is used to determine the semantic features (Figure 4).

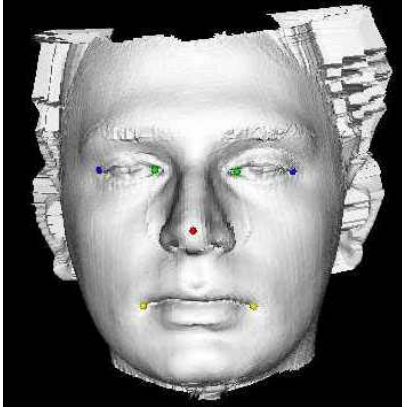


Figure 4. Semantic feature points on a facial scan

The first point to be detected is the *nose tip*. The algorithm assumes that the major axis of the scanned image is vertical, i.e. that the eyes are above and the mouth below the nose. A search area is defined around the middle of the head to find a point with  $\sigma > 0.8$  (spherical cap),  $\gamma > 0.1$  and highest z-coordinate value. Based on the nose point, the *inner eye points* are detected. A search area is defined above the nose tip to find two points on either side of the nose tip with  $\sigma < 0.2$ ,  $\gamma > 0.1$  and least z-coordinate value.

The *outer eye points* are the next semantic feature points to find. Since this search is more difficult, in addition to  $\sigma$ ,  $\gamma$  and  $z$ , a 2-D line is defined based on the position of the inner canthi that passes through the inner canthi and the distances of points from this line were used as additional feature. The search is defined to detect points with  $\sigma < 0.4$ ,  $\gamma > 0.1$  and  $z$  in the vicinity of the depth values of the detected inner eye points. Points satisfying the above criteria are further filtered using the value of the distance of the points from the defined line that passed through the inner canthi. The search narrows down until a single point is found to the right of the nose tip and a single point to the left of the nose tip with minimum distance value and which satisfied the above mentioned parameters. This approach yielded a robust method for detection of the outer canthi in all the test scans.

The detection of the *extremes of the lip* uses a method similar to that for inner canthi detection, with the exception that the search area is below the nose tip. The search looks for points that lie below the nose tip with  $\sigma < 0.3$ ,  $\gamma > 0.2$ . One single point to the right of the nose tip that satisfies the above criteria and which lies to the rightmost extreme is singled out as the right side lip corner. Similarly a point to the leftmost extreme is singled out as the left side lip corner.

The parameters to detect the semantic feature points described above are sufficiently robust across scans and are summarised in Table 1. The optimal value of the thresholds for  $\sigma$  and smoothing parameter  $\alpha$  was defined experimentally. Testing was carried out on 20 facial scans of patients which included both males and females and also included patients from different ethnic origins. The threshold for  $\gamma$  is set to 0.1 since this value is stable and consistent in the whole set of

landmark	$\sigma$	$\gamma$	$z$	other
nose tip	$\sigma > 0.8$	$\gamma > 0.1$	max	(a)
inner canthi	$\sigma < 0.2$	$\gamma > 0.1$	min	(b)
outer canthi	$\sigma < 0.4$	$\gamma > 0.1$	~min	(c)
lip extremes	$\sigma < 0.3$	$\gamma > 0.2$	~max	(d)

**Legend**

- (a) Search restricted to middle of face
- (b) Search restricted to middle of face and above nose tip
- (c) Minimum distance from line passing through nose tip and midpoint of inner canthi
- (d) Search restricted to middle of face and below nose tip

Table 1. The set of parameters and rules used to define the semantic features points on a facial scan

facial scans. The number of hits in accurate detection of the points was used as a detection score (Figure 5, top). A score of 20 indicates that the semantic points associated to a curve are detected accurately in all scans.

The number of smoothing iterations,  $\alpha$ , has been set to 350 based on the results generated on testing the detection accuracy by varying the number of iterations (Figure 5, bottom). A detection score measuring the number of points accurately detected was used for this evaluation. A score of 7 indicates that all the points were detected accurately, i.e. two inner canthi, two outer canthi, the nose tip and two lip extremes. Figure 5 (bottom) shows that in the 3 samples all face points were detected successfully on smoothing the mesh in 300-350 iterations. This value was seen to be consistent in almost all the remaining test scans as well. A smoothing of  $\alpha = 350$  iterations is therefore chosen to be most suited.

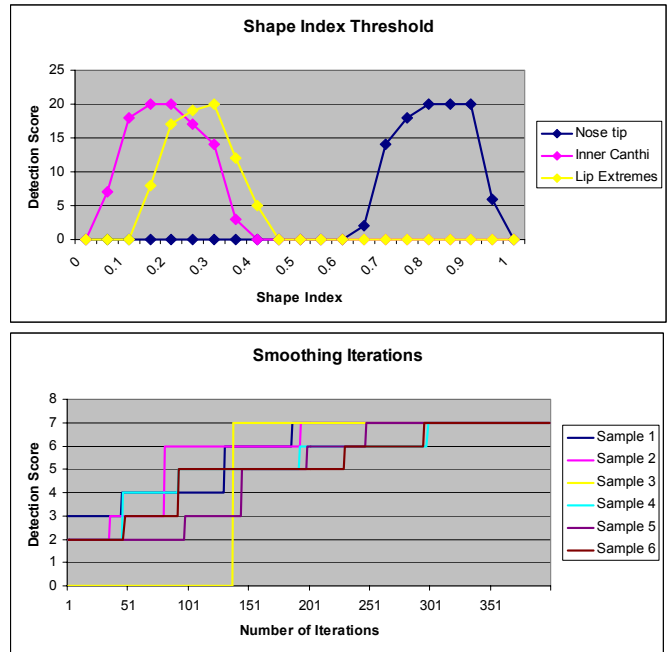


Figure 5. Optimal values for  $\alpha$  and for the threshold on  $\sigma$ . (top) Determination of the optimal threshold on  $\sigma$ ; (bottom) determination of the optimal number of smoothing iterations

## 4 Registration and change detection

Changes between two facial scans are detected by rigid registration followed by distance computation. Scan *registration* facilitates the subsequent change detection because it reduces the number of iterations required in the transformation and in turn helps automatic high-level feature extraction.

### 4.1 Registration

Two registration methods are compared, namely the Iterative Closest Point method and the Landmark Transform method. Based on these two methods, a new method is proposed which exploits the a priori knowledge of the application and overcome the limitations of the first two.

The *Iterative Closest Point (ICP)* method [7] registers the scans based on the closest point associations from one scan to the other. The steps involved in ICP are:

1. Choose one scan as reference.
2. Pair each point of the *reference scan* to the closest point of the *test scan*.
3. Compute the displacement that minimizes the Mean Square Error (MSE) between the paired points.
4. Apply the displacement to the test scan and update the MSE.
5. If the MSE is larger than a predefined threshold and the number of iterations is smaller than a maximum number of iterations, then go to 2.

The iterations have been proven to converge in terms of the MSE. Even if the ICP algorithm assumes that the two scans are roughly pre-aligned, satisfactory results were observed even with an initial relative rotation of up to 30°. While specifying the points to be used in the above algorithm, care was taken to avoid choosing points from the outliers. This was obtained by defining a region in the middle of the coordinate extents of the dataset from where the points would be selected. This region corresponds to the front part of the scan, i.e. the anterior or ventral aspect of the head from the forehead to the chin, including the eyes, nose, mouth, cheek and chin, but excluding the auricles. In almost all scans, the auricles include outliers and are therefore excluded.

The *Landmark Transform (LT)* method computes the best fit mapping in a least squares sense of the set of landmarks (semantic feature points) in the test scan onto the same landmarks in reference scan. A combination of the high-level features described in Section 3.2 is chosen: (i) the inner canthus, (ii) the tip of the nose, (iii) the outer canthus, and (iv) the extremes of the lips. These feature points (and in particular the inner canthus and nose tip) are relatively stable on the facial scan and do not change in terms of relative positions from one scan to another of the same person. The option of choosing the other semantic feature points such as the outer canthus or the extremes of the lips are provided for use in cases where the above mentioned points are known to be unstable due to known deformations in that area of the face for the patient.

In order to achieve an optimal automatic registration, a method incorporating both LT and the ICP is defined.

region	$\sigma$	$\gamma$	other
inner canthi	$\sigma < 0.2$	N/A	(a)
nasal bridge	$0.2 < \sigma < 0.7$	N/A	(b)
eyebrow	$\sigma > 0.4$	$\gamma > 0.05$	(c)
cheekbone	$\sigma > 0.75$	N/A	(d)
<b>Legend</b>			
(a) Around the detected inner canthi points			
(b) Between the detected inner canthi points			
(c) Above the detected inner canthi points			
(d) Below the detected outer canthi points			

Table 2 Parameters defining the segmentation rules for semantic region detection

This proposed method is referred to as *Combo Method (CM)*. The initial registration is performed using LT on the semantic feature points. Then, the ICP algorithm is applied only on *semantic regions* of the face. Semantic regions are ‘stable’ facial regions, which include the region around the inner canthi, the nasal bridge between the inner canthi and the eyebrow region. Although the cheekbone area is also considered to be quite stable, it has not been included in the chosen regions since its segmentation is not very accurate.

The semantic regions are detected using the same two feature maps used in the detection of the semantic points. The thresholds used in the segmentation are summarised in Table 2. Figure 6 shows the regions used by ICP after the initial registration with LT on the semantic points. The inner canthi region is indicated in navy blue, the nasal bridge region is indicated in green, the eyebrow region in yellow and the cheekbone area in sky blue.

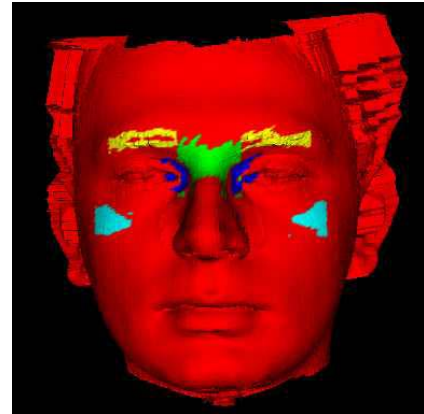


Figure 6. Visualisation of the semantic regions on a facial scan.

### 4.2 Change detection

Once the reference scan and the test scan are registered, the next step is to evaluate the distance between the two meshes so that regions of changes can be identified. To this end, the Symmetric Hausdorff distance is used [4]. Let  $\Sigma$  and  $\Sigma'$  be the



two facial scans and  $\delta(P, \Sigma')$  be the distance between a point  $P$  in  $\Sigma$  and a surface  $\Sigma'$ . If we define

$$\delta(P, \Sigma') = \min_{P' \in \Sigma'} (\|P - P'\|^2), \quad (3)$$

then the Hausdorff distance,  $\delta(\Sigma, \Sigma')$ , is given by

$$\delta(\Sigma, \Sigma') = \max_{P \in \Sigma} \delta(P, \Sigma') \quad (4)$$

and the Symmetric Hausdorff distance,  $\delta_s$ , is then given by

$$\delta_s(\Sigma, \Sigma') = \max[\delta(\Sigma, \Sigma'), \delta(\Sigma', \Sigma)]. \quad (5)$$

The Symmetric Hausdorff metric provides a more accurate estimate of the distance than a one-sided distance computation that would yield highly biased results.

## 5 Experimental results

In this section, the results obtained with the proposed facial scan change detection method are compared with those obtained with ICP and LT. A database with 22 facial scans of real patient data (which are not shown here to respect patient confidentiality) and 6 scans of non-patient data (with natural and intentional deformations) was used and the results have been validated by an experienced clinical scientist.

Sample results are shown in Figure 7 and Figure 8. The changes in two facial scans of the same person taken at different times (top) are computed using ICP, LT and CM. In Figure 7 is clearly visible that the region of change is located around the mouth region, while the rest of the face shows zero error. This is consistent with the morphological knowledge of the person under study. The mouth region consists of soft tissue which has changed over time, while the upper regions of the face consist of hard tissue (bone) which is known to be invariant in this person.

The best matches are achieved using CM. LT has yielded the worst result. This is due to the fact that the face points detected in both scans are not at exactly the same position. Due the change in pose of the patient, the points detected was slightly displaced with respect to each other and the use of only these points in matching the scans leads to an inaccurate matching of the meshes. Unlike LT, a much more accurate registration is achieved using CM since only initial coarse registration is performed using these points followed by fine matching using all points within the stable regions of the face.

In Figure 8, the person changes pose with an intentional deformation of the left cheek so that the entire mouth region gets displaced. The best registration is achieved by CM, while the worst is achieved by LT. The fact that the person has his eyes open in one scan and closed in the other adds further complexity in detecting exactly the same points in both scan for the inner and outer canthi, thus misleading LT.

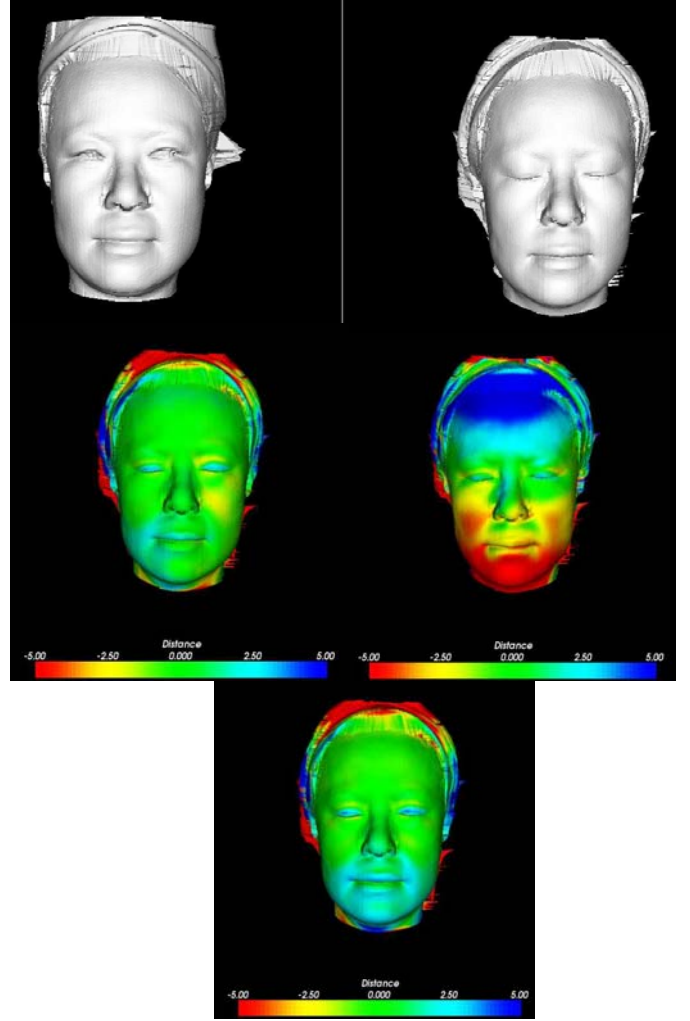


Figure 7. Evaluation of facial scan change detection techniques. (top) Two facial scans taken at two different time instants and change detection using ICP (middle-left), LT (middle-right), and CM (bottom). The discrepancy between the two scans is colour-coded according to the scale shown in the pictures

Based on the testing on the whole database, the proposed CM algorithm did not perform satisfactorily in 2 cases. In both cases, due to a jaw abnormality or to the ethnicity of the patient, the upper and/or lower jaw protrudes such that the lips protrude further than the tip of the nose. In such a case, the tip of the nose is not detected accurately since regions on the lip satisfy the criteria defined for the nose tip leading to the error in detection. This problem could be solved by validating the semantic feature points using the a priori knowledge of their relative positions (model). If the detected semantic feature points are not satisfying the model, then the process could iterate with updated parameters until a satisfactory configuration is found.

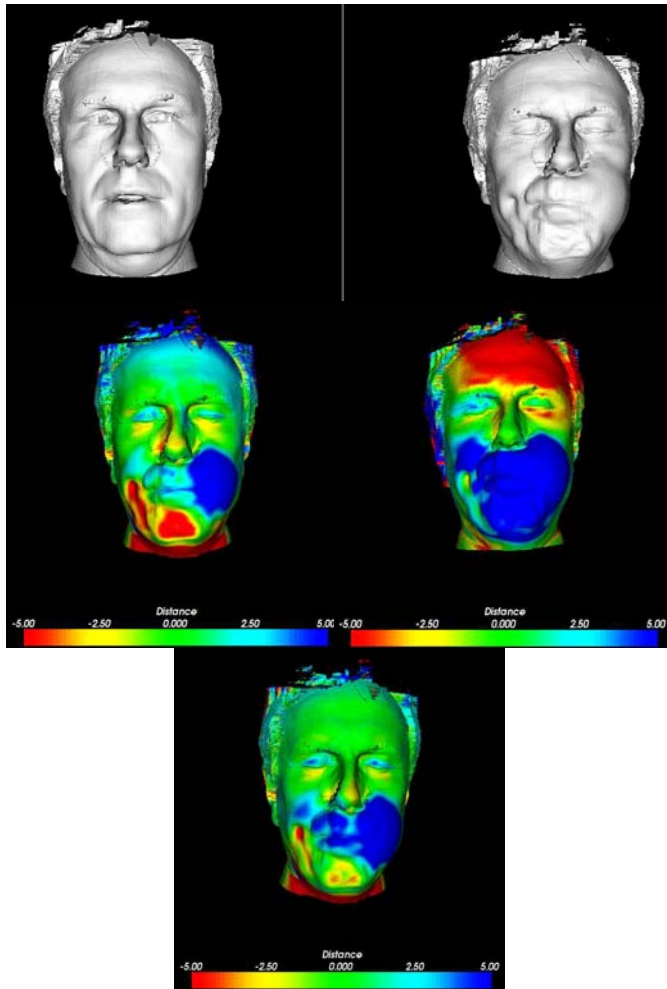


Figure 8. Evaluation of facial scan change detection techniques. (*top*) Two facial scans taken at two different time instants and change detection using ICP (*middle-left*), LT (*middle-right*), and CM (*bottom*). The discrepancy between the two scans is colour-coded according to the scale shown in the pictures

## 6 Conclusions

We have presented an accurate method for rigid registration of face laser scans that uses semantic features automatically extracted based on a priori knowledge and low-level feature maps. The proposed method has been used for the localisation and quantification of changes in facial scans and combines the Iterative Closest Point method with the Landmark Transform method.

The method is incorporated into a complete end-to-end system that has been validated by an experienced clinical scientist on a database of 28 facial scans. The system will be used as support in clinical studies on a database with more than 1000 facial scans.

## References

- [1] G.G. Gordon, "Face Recognition Based on Depth and Curvature Features," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, Champaign, Illinois, pp.108-110, June 1992.
- [2] X. Lu, D. Colbry, and A.K. Jain, "Matching 2.5D Scans for Face Recognition". Int. Conference on Biometric Authentication, LNCS 3072, pp. 30-36, Hong Kong, July 2004.
- [3] C. Dorai and A.K. Jain, "COSMOS - A Representation Scheme for 3D Free-Form Objects," IEEE Trans. on PAMI, vol. 19, no. 10, pp. 1115-1130, 1997.
- [4] N. Aspert, D. Santa-Cruz, T. Ebrahimi, "MESH: Measuring Errors between Surfaces using the Hausdorff Distance". In Proc. of the IEEE Int. Conference in Multimedia and Expo 2002 (ICME), vol. 1, pp. 705-708, Lausanne, Switzerland, 2002.
- [5] D. Nairn, "3D Surface Feature Detection". 4th year project report, Artificial Intelligence and Computer Science, University of Edinburgh School of Informatics, pp. 10-25, May 2004.
- [6] T. J. Hutton, "Dense Surface Models of the Human Face". PhD Thesis, Biomedical Informatics Unit, Eastman Dental Institute, University College London. September, pp. 10-20, 2004.
- [7] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 14, no.2, pp.239-256, 1992.
- [8] S.C. Sung, R.C.K Ngim and S T Lee, "Evaluation of the Laser Scanner as a Surface Measuring Tool and Its Accuracy Compared with Direct Facial Anthropometric Measurements", British Journal of Plastic Surgery, 48, 551-8, 1995.
- [9] H. Gouraud, "Computer Display of Curved Surfaces" Department of Computer Science, University of Utah, UTEC-CSc-71-113, 1971.