A New Method for Automatic 3D Face Registration

Venkat R. Ayyagari*, Faysal Boughorbel, Andreas Koschan, and Mongi A. Abidi
The University of Tennessee, Knoxville, TN 37996, USA.
**Email: vayyagar@utk.edu*

Abstract

In view of today's security concerns, 3D face reconstruction and recognition has gained a significant position in computer vision research. Depth information of a 3D face can be used to conquer the problems of illumination and pose variation associated with face recognition. Registration is an integral process of any reconstruction process and hence we focus on the problem of automatic registration of 3D face point sets through a criterion based on Gaussian fields. The method defines a simple energy function, which is always differentiable and convex in a large neighborhood of the alignment parameters; allowing for the use of powerful standard optimization techniques. The lack of necessity of close initialization overcomes the limitations of Iterative Closest Point algorithm. Moreover, the use of Fast Gauss Transform reduces the computational complexity of the registration algorithm.

1. Introduction

The need for a robust and effective biometric system for security application has been highlighted by security agencies all over the world. The human face seems to be one of the effective biometric features even in the uncooperative environment. Although many security systems based on 2D analysis of face are prevalent in the market, most of them suffer from the inherent problem of illumination and pose [1] [2]. This is one of the main motivating factors for research in 3D face reconstruction and recognition for security purposes. The field of 3D face reconstruction has been rapidly growing during the recent past as range finders became more accurate, affordable, and commercially available. In fact its applications are not restricted just to recognition but spread over wide areas ranging from medical purposes to computer animation, from video surveillance to lip reading systems, from video teleconferencing to virtual reality [3].

Automatic reconstruction of 3D face models typically involves three stages: a data acquisition stage wherein the samples of the face are collected from different views using sensors, a data registration stage which aligns the different 3D views into a common coordinate system, and an integration stage which simplifies the aligned views into parametric models. Generally, some parts of the face will be unobservable from any given position, either due to occlusion or limitations in the sensor's field of view. When seen from a slightly different viewpoint, the missing data in unobserved regions is readily apparent. However, these different views will be in their local coordinate system and some transformations have to be employed to align them in a common coordinate system. It is in this capacity that registration becomes an integral part of the reconstruction process.

In this paper, we address the automatic registration problem at the point level without any external correspondence. The main contribution of this work is the design of a point set criterion which is differentiable and convex in the large neighborhood of the aligned position, overcoming the shortcomings of standard registration techniques, and in particular ICP. The ICP algorithm which is a locally convergent scheme requires parameter initialization close to the aligned position. Without a-priori approximate estimate of the transformation, the ICP often ends in a local minimum instead of the global minimum which represents the best transformation. This is not surprising since the ICP is searching a non-convex, multi - dimensional space using a gradient descent algorithm.

Our energy function is convex in the neighborhood of the solution and always differentiable allowing for the use of a wide range of well proven optimization techniques. We make use of a straightforward sum of Gaussian distances that is defined for point sets with associated attributes; local moments in our case. More importantly, the computational complexity of this criterion is reduced using the numerical technique known as Fast Gauss Transform. The obtained results affirm that our criterion can be used for accurate registration of 3D face datasets while at the same time extending the region of convergence, thus avoiding the need for close initialization. In the following sections we first present the literature overview, then describe the Gaussian criterion and the local attributes used, followed by an overview of the FGT evaluation method. In the results section we present an analysis of our approach regarding (a) the effect of parameter sigma on the registration accuracy, (b) the robustness of the proposed algorithm to different levels of noise, and (c) the influence of the data resolution on the results. Furthermore, we compare the performance of our algorithm to the performance of the standard ICP algorithm. Finally, we end with a conclusion.

2. Literature Overview

3D face reconstruction techniques can be broadly classified into active and passive methods based on their imaging modalities [6]. Active reconstruction techniques such as laser scan and structured light use external source of illumination for reconstruction whereas passive techniques such as stereo vision, morphing, structure from motion etc do not depend on external source of illumination. Most of the above mentioned methods invariably make use of registration technique in the process of building a complete face model.

The majority of the registration algorithms attempt to solve the classic problem of absolute orientation: finding a set of transformation matrices that will align all the data sets into a world coordinate system [7]. In the literature, a common distinction is found between fine and coarse registration methods [4], which are often used in a two stage fashion: a coarse registration followed by fine registration using the ICP and its variants.

The original ICP algorithm developed by Besl and MacKay [5] aligns the two point sets by minimizing the sum of squared distances between them. However, this approach converges monotonically to a local minimum and hence a good estimate of the initial transformation between point sets is required. Modifications to the original ICP algorithm have been made to improve the rate of convergence and register partially overlapping datasets. Chen and Medioni [8] used an iterative refinement of initial coarse registration between views to perform registration and utilized orientation information. They devised a new least square problem where the energy function being minimized is the sum of the distances from points on one view surface to the tangent plane of another views surface. Zhang [9] proposed a method based on heuristics to remove inconsistent matches by limiting the maximum distance between closed points allowing registration of partially overlapping data. While the basic ICP algorithm was used in the context of registration of clouds of points, Turk and Levoy [10] devised a modified registration metric that dealt with polygon meshes. They used uniform spatial subdivision to partition the set of mesh vertices to achieve efficient local search.

In order to improve the robustness of ICP, Masuda and Yokoya [11] used a Least Mean Square (LMS) error measure that is robust to partial overlap. Some other methods involved in the same effort at robustness were the Minimum Variance Estimate (MVE) of the registration error proposed by Dorai *et al.* [12], Least Median Squares (LMedS) proposed by Trucco *et al.* [13]. Also for reducing the computational complexity some other variants were introduced such as the use of k-D trees to partition datasets [9] and the use of spatial subdivision to partition mesh vertices[10] .

Stoddart *et al.* [14] studied the relationship between surface shape complexity and registration accuracy. Early work by Arun *et al.* [15] on estimating 3D rigid body transformations presented a solution using the singular value decomposition (SVD). The method requires a connected set of correspondences and accurately registers the 3D data. Faugeras and Hebert [16] employed the quaternion method to solve the registration problem directly.

Eggert *et al.* [17] proposed a method in which data from each view is passed through Gaussian and Median filters, and point position and surface normal orientation are used to establish correspondence between points. Chen *et al.* [18] proposed a random sample consensus (RANSAC) scheme that is used to check all possible data-alignments of two data sets. The authors claim that their scheme works with featureless data and requires no initial pose estimate.

The non differentiability of the ICP cost function imposes the use of specialized heuristics for optimization. Addressing the registration in the context of Gradient based optimization has attracted some interest recently. In his work Fitzgibbon [19] showed that a Levenberg-Marquardt approach to point set registration problem offers several advantages over current ICP methods. The proposed method uses Chamfer distance transforms to compute derivatives and Huber kernel to widen the basins of convergence of existing techniques. We try to overcome the limitations of the ICP algorithm by introducing a straightforward differentiable cost function, directly and explicitly expressed in terms of point coordinates and registration parameters.

3. Our Approach

The main idea employed in our 3D Registration method is to make use of the Gaussian fields to measure both the spatial proximity and the visual similarity of the two datasets in the point form.

3.1 Gaussian fields and Energy Function

We introduce our criterion on two point-sets $M = \{(P_i, S(P_i))\}$ and $D = \{(Q_j, S(Q_j))\}$, with their associated attribute vectors. As we consider our datasets in point form we utilize 3D moments as our attributes. However, those vectors can include curvature for smooth surfaces and curves, invariant descriptors, as well as color attributes when available. The Gaussian measure is given by:

$$F(P_i, Q_j) = \exp\left(-\frac{d^2(P_i, Q_j)}{\sigma^2} - \frac{(S(P_i) - S(Q_j))^T \Sigma^{-1} (S(P_i) - S(Q_j))}{C_a^2}\right) \tag{1}$$

with $d(P_i, Q_j)$ being the Euclidean distance between the points and C_a being the attribute confidence parameter. In the context of particle physics the expression (1) can be seen as a force field whose sources are located at one point and are decaying with distance in Euclidean and attribute space. We can now define an energy function that measures the registration of M and D as:

$$E(Tr) = \sum_{\substack{i=1\ldots N_M \\ j=1\ldots N_D}} \exp\left(-\frac{d^2(P_i, Tr(Q_j))}{\sigma^2} - \frac{(S(P_i) - S(Tr(Q_j)))^T \Sigma^{-1} (S(P_i) - S(Tr(Q_j)))}{C_a^2}\right) \tag{2}$$

where Tr is the transformation that registers the two point-sets. The Force Range parameter (σ) controls the region of convergence, while the parameter Σ normalizes the differences in the attributes, and the parameter C_a compensates the effect of noise on the features used in Gaussian criterion. If we choose the decay parameters very small, the energy function E will just 'count' the number of points that overlap at a given pose. This is due to exponential being very small except for $P_i = (RQ_j + t)$ and $S(P_i) = S(Q_j)$. In particular, if M is a subset of D we will have at the registered position:

$$(R^*, t^*) : \lim_{\substack{\sigma \to 0 \\ \Sigma \to 0}} E(R^*, t^*) = N_M \tag{3}$$

Thus, for this case we meet a rigorous definition of registration as maximization of both overlap and local shape similarity between the datasets.

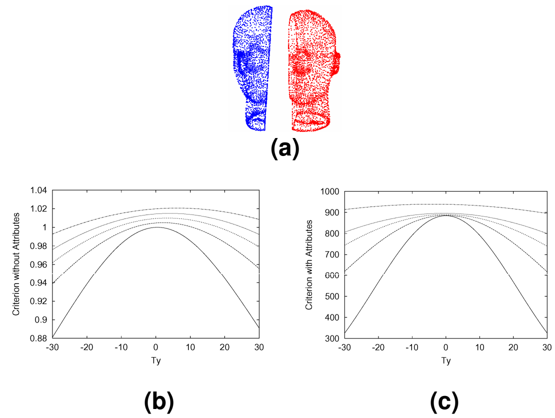


Figure 1. Profiles of the Gaussian energy function for a displacement around the registered position of the dataset shown in (a). In (b) the profiles are plotted in the case without attributes for σ = 30,50,70,90,150 (from narrowest to widest). Plots with moment invariants as attributes for the same values of σ are shown in (c) (For (b) magnitudes were rescaled for comparison).

The Gaussian energy function is convex and is differentiable in a large region around the registered position, allowing us to make use of the standard optimization techniques such as Quasi-Newton method. As mentioned earlier the parameter σ controls the convex safe region of convergence. Higher its value larger will be the region of convergence, but this comes at an expense of reduced localization accuracy. However, the region of convergence can be extended considerably with limited reduction in localization accuracy if the datasets have sufficient shape complexity and many independent local descriptors are used. This tradeoff can be illustrated with the behavior of the matching criteria with and without attributes as illustrated in Fig. 1. The profile of the criterion with increasing values of sigma was plotted for relative displacement of the two point sets of Fig. 1(a). It is noticed that for the non-attributed case (Fig. 1(b)) as σ increases the width of the Gaussian bell increases too, but the maximum starts to drift away from the correct position. However, when we use the Gaussian criterion with moment invariants, as attributes associated with the points, the maximum is stable for the same values of σ (Fig. 1(c)). Instead of just continuously incorporating additional information from the point sets, we employ a strategy of tuning the parameter σ to increase the ROC without losing localization accuracy. A rough alignment is performed initially using a large sigma and then its value is decreased for future refinement steps.

3.2 The Fast Gauss Transform

The registration criterion which is a mixture of N_D Gaussians evaluated at N_M points then summed together has a high computational cost of $O(N_M \times N_D)$, which is very high for large datasets. This problem which is also encountered in other computer vision applications can be solved by a new numerical technique called as Fast Gauss Transform (FGT). The method introduced by Greegard and Strain [20] is derived from a new class of fast evaluation algorithms known as "fast multipole" methods and can reduce the computational complexity of the Gaussian mixture evaluation to $O(N_M + N_D)$. The basic idea is to exploit the fact that all calculations are required only up to a certain accuracy. In this framework the sources and targets of potential fields were clustered using suitable data structures, and the sums were replaced by smaller summations that are equivalent to a given level of precision.

The FGT method is used to evaluate sums of the form $S(t_i) = \sum_{j=1}^{N} f_j \exp(-(\frac{s_j - t_i}{\sigma})^2)$, $i = 1,\dots,M$ where $\{s_j\}_{j=1,\dots,N}$ are the centers of the Gaussians known as sources and $\{t_i\}_{i=1,\dots,M}$ the targets. The following shifting identity and expansion in terms of Hermite series are used:

$$\exp(\frac{-(t-s)^2}{\sigma^2}) = \exp(\frac{-(t-s_0-(s-s_0))^2}{\sigma^2}) \tag{4}$$
$$= \exp(\frac{-(t-s_0)^2}{\sigma^2})\sum_{n=0}^{\infty}\frac{1}{n!}(\frac{s-s_0}{\sigma})^n H_n(\frac{t-s_0}{\sigma})$$

where H_n are the Hermite polynomials. Given that these series converge rapidly, and that only few terms are needed for a given precision, this expression can be used to replace several sources by s_0 with a linear cost at the desired precision, these clustered sources can then be evaluated at the targets. For a large number of targets the Taylor series (5) can similarly be used to group targets together at a cluster center t_0, further reducing the number of computations.

$$\exp(-(\frac{t-s}{\sigma})^2) = \exp(\frac{-(t-t_0-(s-t_0))^2}{\sigma^2}) \tag{5}$$
$$\approx \sum_{n=0}^{p}\frac{1}{n!}h_n(\frac{s-t_0}{\sigma})(\frac{t-t_0}{\sigma})^n$$

where the Hermite functions $h_n(t)$ are defined by $h_n(t) = e^{-t^2}H_n(t)$. The method was shown to converge asymptotically to a linear behavior as the number of sources and targets increases.

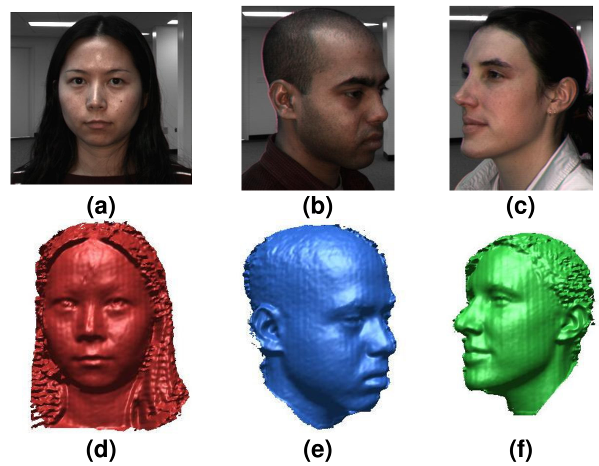


Figure 2. Sample models in different poses from our IRIS 3D face database. The two dimensional intensity images (a) (b) (c) along with their associated 3D models (d) (e) (f)

4. Experimental Analysis

In our experiments we have used two different data sets: the synthetic dataset (Fig.1.a) and the IRIS 3D face dataset (Fig.3) from our database. The 3D face dataset was scanned using the Genex 3DFaceCam which operates on the principle of structured light. Some of the models from our database are depicted in Fig. 2.

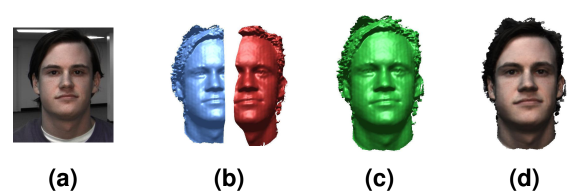


Figure 3. 3D face model from our database. The 2D image (a); two different views of the face (b); registered model (c); and the complete 3D model with texture (d). In our experiments the texture is discarded and point sets are used instead.

4.1 Effect of varying the parameter Sigma

The parameter σ controls the region of convergence which should be large for better practical applications. However, increasing the value of σ without any constraints causes the decrease in the localization accuracy. It is with this motivating factor we analyze the effect of varying sigma on the registration accuracy using the synthetic and 3D face dataset from our database. The results of this experiment are shown in Fig. 4.

It is interesting to find that both the models exhibit similar trends in the sense that the registration error increases linearly as a function of σ. However rate of increase slows down for larger values of σ and tends towards an asymptotic limit. This can be the explained by the fact that as σ exceeds the average distance between the points in the datasets the exponential can be approximated by its first order development:

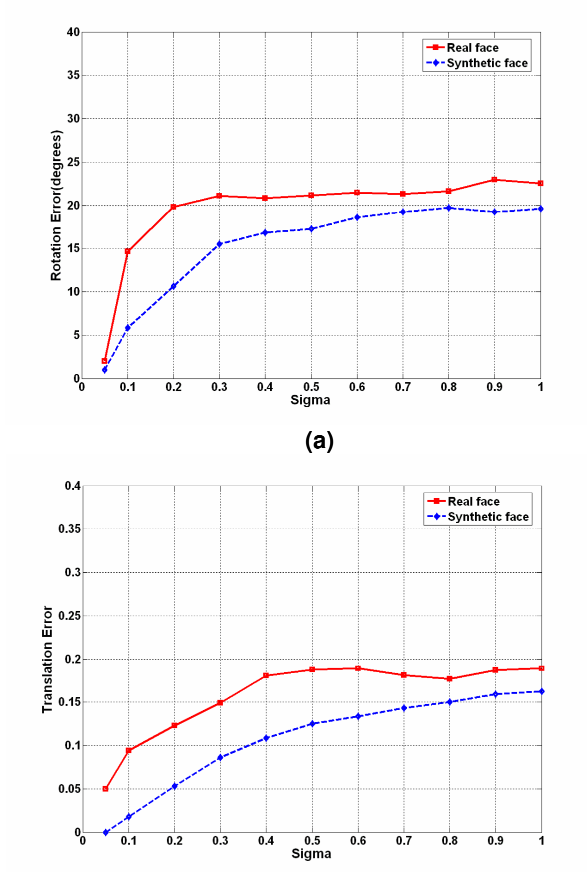
$$\exp\left(-\frac{d^2(Tr(P_i), Q_j)}{\sigma^2}\right) \approx 1 - \frac{d^2(Tr(P_i), Q_j)}{\sigma^2} \qquad (6)$$

The optimization problem now reduces to minimizing the sum of average distances from one point set to other dataset and doesn't depend anymore on σ. Another interesting observation is that the registration error generally remains less that 20% the length of the model for translation and less than 25° for rotation. Based on this behavior we can develop an algorithm that starts with initial rough alignment with a large σ and then end up with a refinement step where σ is sharply decreased leading to a very low registration error.

4.2 Resolution Analysis

The main criterion of a good registration method is the level of accuracy and the computational complexity involved. There are many optimization techniques which would reduce the computational complexity burden. Although Fast Gauss transform was utilized to reduce the computational complexity of the criterion, the sub- sampling of the datasets would lead to a lower level of complexity. However, the number of points in the datasets should be sufficient to maintain the accuracy level. Hence this turns out to be an optimization between the computational complexity and level of accuracy. It was this factor which drove us to experiment on the minimum number of points in space required for an effective 3D Registration.

The dataset utilized was taken from our IRIS 3D face database. We start with the relatively low number of 6000 points for each view then sample by two to obtain the next pairs until we reach 750 points. To study the influence of reduction in resolution we sub-sampled our datasets in three different ways: uniform sampling wherein the points are sampled at equal intervals; curvature based sampling wherein points in high curvature regions are less deleted than low curvature region in order to maintain the accuracy of the curvature line and random sampling wherein the points are randomly sampled throughout the dataset.



(a)

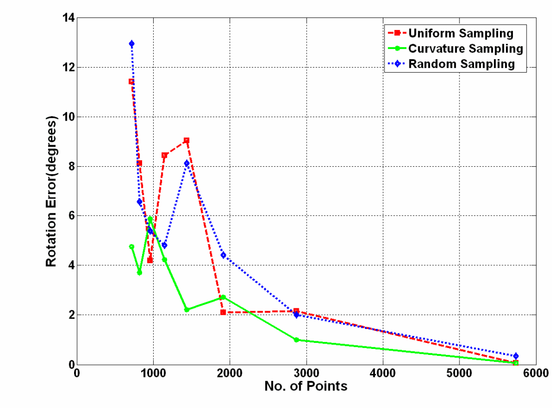
(b)

Figure 4. Plots showing the rotation (a) and translation error (b) of the real face data and the synthetic face as a function of parameter σ. The parameter sigma and translation error are in terms of fraction of the length of the face model

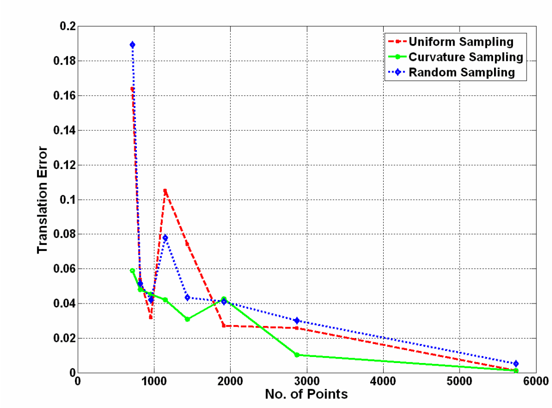
Although at higher levels of sampling (lower number of points) the curvature sampling provides a slight insignificant edge over others, no particular method can be considered superior to others. The reason that no particular sampling method can be attributed as perfect is due to the following reasons:

- Uniform sampling has better spatial distribution of points but this may lead to coarser description of objects.
- Curvature sampling has better visual description but may sometimes lead to complications due to clustering of points in certain areas.
- Random sampling may create complications due to uneven distribution of points.

Another observation from Fig. 5 is that the criterion does not break down even at higher levels of sampling and remains intact even for a few points around 800, thus reducing the computational burden by multi resolution strategy that initializes at coarser levels.

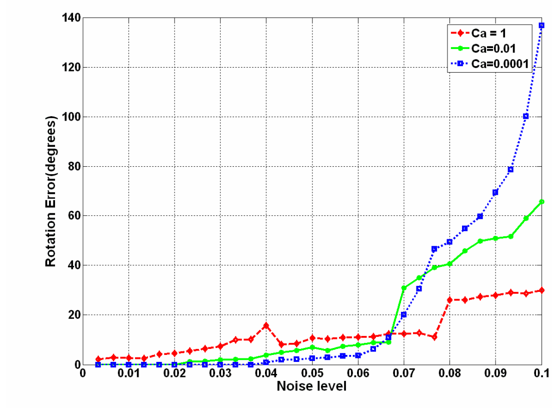


(a)

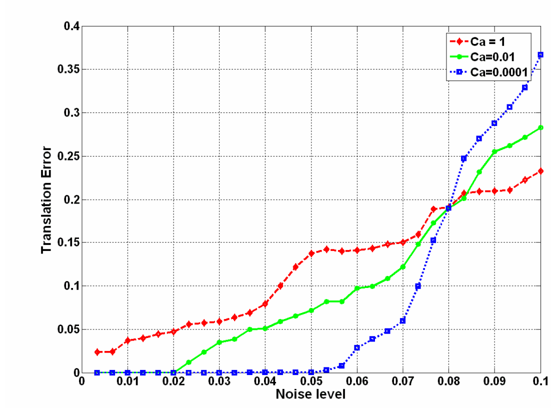


(b)

Figure 5. Effect of sampling on the registration accuracy. Rotation error (a) and translation error (b) as a function of number of points for three different sampling methods



(a)



(b)

Figure 6. Registration error versus uniform noise: rotation error (a) in degrees, translation error (b) as a fraction of the length of the face model. We show plots for three values of the confidence parameter.

4.3 Noise Analysis

Noise has a significant effect on the 3D registration process especially in the Gaussian criterion framework because it influences both the position of the point-sets as well as the descriptors computed from them. In practical applications noise is more dominant in the radial direction with respect to camera's coordinate frame. However, we focus our experimental analysis on uniform noise to study the worst case scenario. As mentioned in earlier sections the parameter C_a is added to our criterion to compensate the effect of descriptors which become practically useless at higher levels of noise. This is achieved by forfeiting a part of discriminatory power that the descriptors add at higher levels of noise. For practical applications the confidence level factor is typically chosen to be around 10^{-3} for datasets with low noise levels and around unit value for higher noise values. For the purpose of noise analysis we add uniform noise of amplitude going up to 10% of the length of the head to both the models.

The effect of uniform noise on the drift in the maximum of the criterion can be studied from the plots shown in Fig. 6. The first conclusion made from plots is that our algorithm is robust for levels of uniform noise upto ±7% which is very high by any practical standards. The effect of C_a in moderating the effect of registration accuracy at higher levels of noise can also be clearly seen.

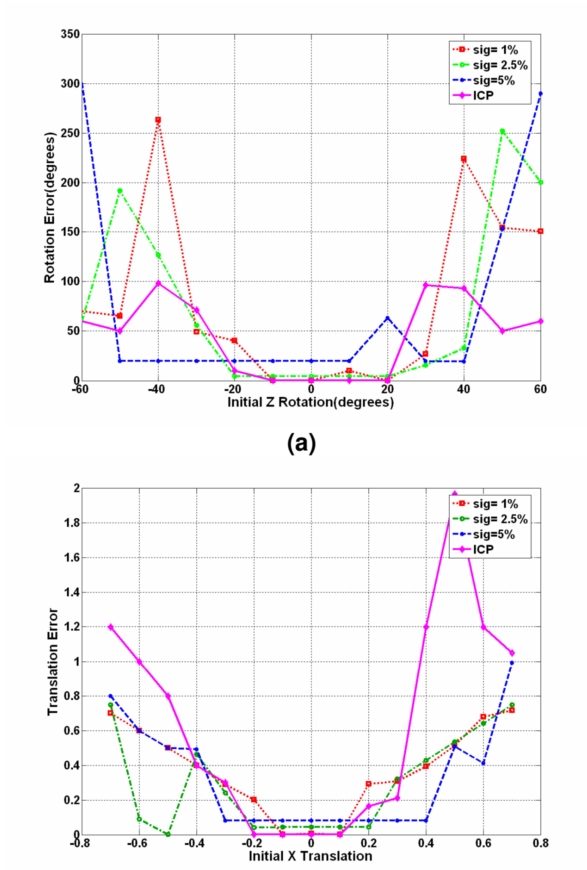
4.4 Comparison with ICP

In order to study the effect of σ on the region of convergence and to prove its advantages over the ICP algorithm, we analyzed the basins of convergence of the algorithm for the Head dataset. A relationship between the initial value of transformation parameters provided to the algorithm and the residual error at the end of the process with different values of σ can be seen from Fig. 7.

These plots confirm the tradeoff between a large basin of convergence for a large value of σ associated with a large residual error as well, and a smaller basin of convergence for a small value of σ that comes with a better registration accuracy. It can also be found that the width of the basins will grow fast first but then does not increase much after a certain value of the force range parameter which can be deduced from earlier sections. Also when these basins are compared with that of ICP, it is found that they are wider even for small values of σ. This can be attributed to the fact that ICP is a locally convergent scheme and needs close initialization. However, the ICP has a small residual error except when compared with algorithm tuned for close Gaussian fields. Thus a balance between residual error and the region of convergence can be obtained by a suitable adaptive optimization scheme.

5. Conclusions

In this paper we demonstrate a new automatic registration method based on Gaussian Fields applied to 3D face reconstruction. The method overcomes the close initialization limitation of ICP and avoids the two stage registration process employed by the other algorithms. Moreover, the method allows us to start from arbitrary initial position and converge to the registered position. A simple energy function is utilized and by the application of Fast Gauss Transform, the computational complexity is reduced to linear level. The experiments performed on real noisy 3D Head datasets demonstrate the effectiveness of our method. Furthermore, the simple energy criterion can be applied for the task of 3D face recognition which is currently under investigation.



(a)

(b)

Figure 7. Comparison of our method's basin of convergence to that of ICP; rotation error (a) and translation error (b) for three different values of sigma and the ICP.

Acknowledgements

This work is supported by the University Research Program in Robotics under grant DOE-DE-FG02-86NE37968, by the DOD/RDECOM/NAC/ARC Program, R01-1344-18, and by the Office of Naval Research under grant # N000143010022.

6. References

[1] W. Zhao and R. Chellappa, "3D Model Enhanced Face Recognition," Proc. of Int'l Conf. on Image Processing , vol. 3, 2000, pp.50-53.

[2] K.W. Bowyer, K. I. Chang and P. J. Flynn, "A Survey of Approaches to Three-Dimensional Face Recognition," Int'l Conf. on Pattern Recognition, England, vol. 1, August 2004, pp. 358-361.

[3] N. D'Apuzzo, "Modeling Human Faces with Multi-Image Photogrammetry," Three-Dimensional Image Capture and Applications V, Proc. of SPIE, San Jose, California, vol. 4661, 2002.

[4] R. Campbell and P. Flynn, "A Survey of Free-form Object Representation and Recognition Techniques," Computer Vision and Image Understanding, vol. 81, No. 2, 2001, pp. 166-210.

[5] 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, 1992, pp.239-256.

[6] A. Bronstein, M. Bronstein and R. Kimmel, "Expression - invariant 3D face recognition," Proc. Audio & Video-based Biometric Person Authentication (AVBPA), Lecture Notes in Comp. Science 2688, Springer, 2003, pp. 62-69.

[7] B. K. P. Horn, "Closed-form Solution of Absolute Orientation using Unit Quaternions," Journal of the Optical Society of America A, vol. 4, No. 4, April 1987, pp. 629-642.

[8] Y. Chen and G. Medioni, “Object modeling by registration of multiple range images,” Image Vision Computing, vol.10, 1992, pp.145–155.

[9] Z. Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces,” Int’l Journal of Computer Vision, vol. 13, No. 2, 1994, pp 119-152.

[10] G. Turk and M. Levoy, "Zippered Polygon Meshes from Range Images,” in Proc. of Siggraph 94, Orlando, Fl, 1994, pp. 311-318.

[11] T. Masuda and N Yokoya, "A Robust Method for Registration and Segmentation of Multiple Range Images,” Computer Vision and Image Understanding, vol. 61, No. 3, 1995, pp. 295-307.

[12] C. Dorai, J. Weng and A. K. Jain, “Optimal Registration of Object Views using Range Data,” IEEE Trans. on Pattern Analysis Machine Intelligence, vol. 19, No. 10, 1997, pp. 1131-1138.

[13] E. Trucco, A. Fusiello and V. Roberto, “Robust motion and correspondence of noisy 3-D point sets with missing data,” Pattern Recognition Letters, vol. 20, 1999, pp. 889-898.

[14] A. J. Stoddart and A. Hilton, "Registration of multiple point sets,” Proceedings 13th Int’l Conference on Pattern Recognition, Vienna, Austria, 1996, pp. 40-44.

[15] K. S. Arun, T. S. Huang and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," IEEE Trans. on Pattern Analysis and Machine Intelligence, number 5, Sept. 1987, pp. 698-700.

[16] O. D. Faugeras and M. Hebert, “The Representation, Recognition, and Positioning of 3D Shapes From Range Data,” Techniques for 3D Machine Perception, INRIA North-Holland, 1986, pp 13-51

[17] D.W. Eggert, A.W. Fitzgibbon and R. B. Fisher, “Simultaneous registration of multiple range views for use in reverse engineering of CAD models,” Computer Vision and Image Understanding, vol. 69, 1998, pp 253–272.

[18] C. Chen, Y. Hung and J. Cheng, “RANSAC-based DARCES: A New Approach for Fast Automatic Registration of Partially Overlapping Range Images,” IEEE Trans. on Pattern Analysis and Machine Vision, vol. 21, number 11, Nov. 1999.

[19] A. W. Fitzgibbon, “Robust registration of 2D and 3D Point Sets,” Image and Vision Computing, vol. 21, 2003, pp. 1145-1153.

[20] L. Greengard and J. Strain, “The Fast Gauss Transform,” SIAM J. Scientific Computing, vol. 12, 1991, pp. 79-94.