

SRICP: An Algorithm for Matching Semi-Rigid Three-Dimensional Surfaces

Ajmal Mian, Mohammed Bennamoun and Robyn Owens

School of Computer Science and Software Engineering

The University of Western Australia

35 Stirling Highway, Crawley, WA 6009, Australia

Email: ajmal@csse.uwa.edu.au

Abstract

The Iterative Closest Point (ICP) algorithm can accurately register and match 3D rigid surfaces. However, when it comes to non-rigid or semi-rigid surfaces such as the human face, the performance of ICP drops significantly. In this paper, we extend the ICP algorithm and propose a Semi-Rigid ICP algorithm which can match and register semi-rigid surfaces. We compare the performance of SRICP and ICP algorithms in a challenging scenario whereby 3D faces under exaggerated facial expressions are matched to 3D faces under neutral expression for recognition. Our results show that the proposed SRICP algorithm performs significantly better than the original ICP algorithm.

Keywords: Non-rigid registration and matching, semi-rigid surface, 3D face recognition, ICP.

1 Introduction

Surface registration and matching are fundamental problems in computer vision. Surface registration has applications in 3D modeling [8] whereas surface matching has applications in object recognition [11]. This paper mainly focuses on the latter problem i.e. surface matching with particular emphasis on 3D face recognition. A 3D face is essentially a three-dimensional surface represented by a 3D vector of its x, y, z coordinates. 3D face recognition is believed to have the potential to achieve higher accuracy compared to its 2D counterpart mainly because 3D surface matching is robust to changes in illumination, makeup and pose [3]. However, 3D face recognition is more sensitive to varying facial expressions compared to 2D face recognition. For literature review of existing 2D and 3D face recognition algorithms, the interested reader is referred the surveys of Zhao et al. [16] and Bowyer et al. [3].

The Iterative Closest Point (ICP) [2] is a classic algorithm used for the registration and matching of 3D rigid surfaces. It has been extensively used for 3D face recognition [5][7][9][10]. However, strictly speaking, face is a non-rigid object and varying facial expressions can significantly change the 3D surface of the face. This is why the performance of ICP significantly deteriorates under varying facial expressions. For example in [7], the 3D face recognition rate for neutral faces is 98% whereas it drops to 68% for smiling faces. The recogni-

tion rate drops significantly even though there are no exaggerated facial expressions. In our earlier work [10], we demonstrated that it is possible to achieve high 3D face recognition accuracy by using only partial regions of the face which are comparatively less sensitive to expressions. However, such regions not only vary between individuals but also vary between different facial expressions. To determine these precise regions for an individual requires many training images under all possible facial expressions which are usually not available in practical situations. Furthermore, this approach will also require the pre-classification of the expression type (e.g. smile, frown, anger, disgust) before performing recognition.

On the pretext that under different facial expressions, some regions of the 3D facial surface deform significantly lesser compared to others, we treat the face as a semi-rigid object in this paper. We extend the ICP algorithm and present SRICP (Semi-Rigid Iterative Closest Point) which can match and register semi-rigid 3D surfaces such as the human face. Like the ICP algorithm, SRICP is generic and can be applied to any 3D or even nD non-rigid datasets. Briefly, SRICP dynamically determines the points of the probe face which are less likely to have been affected by facial expressions and matches them to the corresponding points of the gallery face. These points are different for each match i.e. for each probe versus gallery face. Moreover, we also calculate a weighted distance error between the two faces by giving confidence weights

to different points (according to their sensitivity to facial expressions) of the gallery faces. We compare the recognition performance of the two algorithms using the FRGC (Face Recognition Grand Challenge) dataset [14] and demonstrate that SRICP outperforms the ICP algorithm.

This paper is organized as follows. Section 2 gives a brief description of the ICP algorithm. Section 3 gives details of the proposed SRICP algorithm and its differences from the ICP algorithm. Experimental setup and results are given in Section 4 and Section 5 respectively. Conclusions are given in Section 6.

2 Iterative Closest Point Algorithm

The ICP algorithm assumes that the two surfaces (to be matched or registered) are already coarsely registered or an initial set of correspondences have been identified between them either manually or automatically through a feature matching algorithm [8]. The distance between these corresponding points is then minimized by applying a rigid transformation to one of the surfaces (see Section 3 for details). Next, the ICP algorithm iteratively establishes correspondences between the closest points of the two surfaces and minimizes the distance between them by applying a rigid transformation to one of the surfaces. The iterations stop when the distance error reaches a saturation value and cannot be further decreased. The end effect of the algorithm is that the two surfaces are registered and the final distance error value is used as a similarity metric between the two surfaces. The more accurately the two surfaces resemble each other, the lower is the error. Note that this error value is also dependent upon how accurately the “closest points” of the two surfaces represent the correspondences between the two surfaces.

A number of modifications have been proposed to improve the registration performance of the ICP algorithm. These modifications are mainly targeted at improving the correspondence establishment which determines the final accuracy of the algorithm. Setting thresholds on the allowed distance between the closest points and the angular difference between their normals have improved the registration performance of the ICP algorithm. Establishing correspondences along the sensor viewing direction has been found to improve face recognition performance [10]. Since ICP is a computationally expensive algorithm, many efficient variants have also been proposed [15]. ICP has also been extended to non-rigid intensity based registration of 3D volumes [6].

3 Semi-Rigid ICP

SRICP mainly differs from the ICP algorithm in determining the eligible closest point correspondences and calculation of the distance error. In ICP, the closest pairs of points that are within a certain distance threshold are considered corresponding points. However, in SRICP, the closest pairs of points whose *weighted* distance is less than a threshold are considered corresponding points. Moreover, these weights also count towards the calculation of the final error in SRICP. The SRICP algorithm is described in detail below.

We use our automatic pose correction algorithm [10] for providing a initial registration of the faces for onward refinement by SRICP. All faces are pre-processed to remove spikes and noise and fill holes. Next each face is normalized with respect to pose and sampled on a uniform square grid [10]. The resultant faces are facing front with origin (coordinate [0 0 0]) at their nose tip.

Let $\mathbf{P} = [x_i, y_i, z_i]^T$ (where $i = 1 \dots n_P$) and $\mathbf{G} = [x_j, y_j, z_j]^T$ (where $j = 1 \dots n_G$) be the point cloud of a probe and a gallery face respectively. \mathbf{P} and \mathbf{G} are matrices of size $3 \times n_P$ and $3 \times n_G$ respectively. Let \mathbf{k} be a vector of size n_G whose elements represent the confidence in the respective point of the gallery face \mathbf{G} . \mathbf{k} is calculated as follows.

$$\hat{\mathbf{G}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_j \\ y_j \end{bmatrix} \quad (1)$$

$$\mathbf{r} = \sqrt{\sum_{j=x,y} \hat{\mathbf{G}}.^2} \quad (2)$$

$$\mathbf{k} = \frac{r}{\max(\mathbf{r})\pi} \arccos(\hat{\mathbf{G}}_{x./\mathbf{r}}) \quad (3)$$

Eqn. 1 maps the gallery face to the xy-plane and rotates it by 90° so that the x-axis passes between the eyes. In Eqn. 2, \mathbf{r} is a vector of the distances of each point of $\hat{\mathbf{G}}$ from the origin i.e. nose tip. In Eqn. 2 and Eqn. 3, $.^2$ and $./$ stand for point wise square and divide respectively. $\hat{\mathbf{G}}.^2$ is equal to a matrix whose each element is equal to the square of the corresponding element in $\hat{\mathbf{G}}$. Similarly, $\hat{\mathbf{G}}_{x./\mathbf{r}}$ is equal to a vector whose elements are equal to ratio of the corresponding elements of the vectors $\hat{\mathbf{G}}_x$ and \mathbf{r} ($\hat{\mathbf{G}}_x$ is a vector of the x coordinates of the pointcloud $\hat{\mathbf{G}}$). Note that \mathbf{k} has a negative polarity i.e. lower values of \mathbf{k} mean higher confidence. The confidence values decrease with increasing distance from the nose tip and with increasing absolute angle from the x-axis as shown in Fig. 1. From Fig. 1, we can see that the upper part of the face has higher confidence whereas the lower part of the face (the mouth) has the lowest

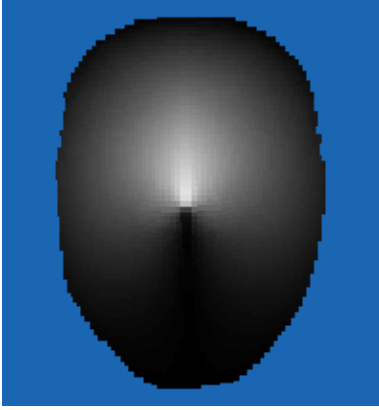


Figure 1: Confidence values calculated for a gallery face. The brighter regions represent higher confidence. Note that the upper part of the face and nearest to the nose has higher confidence.

confidence. This is to minimize the effects of an open mouth facial expression.

Let F be a function that finds the nearest point in \mathbf{P} to every point in \mathbf{G} .

$$(\mathbf{c}, \mathbf{d}) = F(\mathbf{G}, \mathbf{P}) \quad (4)$$

$$\mathbf{d}_k = \mathbf{d}(\mathbf{c})\mathbf{k} \quad (5)$$

In Eqn. 4, \mathbf{c} and \mathbf{d} are vectors of size n_P each such that c_i and d_i respectively contain the index number and distance of the nearest point of \mathbf{G} to the i th point of \mathbf{P} . \mathbf{d}_k is the confidence weighted distance calculated by multiplying the distance of a corresponding gallery point by its confidence (Eqn. 5). The correspondences are sorted according to the increasing value of \mathbf{d}_k and the last 10% are removed. Points of \mathbf{P} that fall in this category are also removed. Next, the 3D distance error e given by Eqn. 6 is minimized.

$$e = \frac{1}{N} \sum_{i=1}^N \|\mathbf{R}\mathbf{g}_i + \mathbf{t} - \mathbf{p}_i\| \quad (6)$$

In Eqn. 6, \mathbf{p}_i and \mathbf{g}_i are the corresponding points of the probe and gallery and $i = 1 \dots N$ (where N is number of remaining points of the probe). \mathbf{R} is a rotation matrix and \mathbf{t} is a translation vector that minimizes the distance between the corresponding points of the gallery and probe. Their values can be calculated using the classic SVD (Singular Value Decomposition) method [1]. Note, that this method can easily be generalized to any number of dimensions and is presented below for completeness. The mean of \mathbf{p}_i and \mathbf{g}_i is given by

$$\mu_p = \frac{1}{N} \sum_{i=1}^N \mathbf{p}_i \quad \text{and} \quad (7)$$

$$\mu_g = \frac{1}{N} \sum_{i=1}^N \mathbf{g}_i \quad \text{respectively.} \quad (8)$$

The cross correlation matrix \mathbf{K} between \mathbf{p}_i and \mathbf{g}_i is given by

$$\mathbf{K} = \frac{1}{N} \sum_{i=1}^N (\mathbf{g}_i - \mu_g)(\mathbf{p}_i - \mu_p)^\top \quad (9)$$

Performing a Singular Value Decomposition of \mathbf{K}

$$\mathbf{U}\mathbf{A}\mathbf{V}^\top = \mathbf{K} \quad (10)$$

gives us two orthogonal matrices \mathbf{U} , \mathbf{V} and a diagonal matrix \mathbf{A} . The rotation matrix \mathbf{R} can be calculated from the orthogonal matrices as

$$\mathbf{R} = \mathbf{V}\mathbf{U}^\top, \quad (11)$$

whereas the translation vector \mathbf{t} can be calculated as

$$\mathbf{t} = \mu_p - \mathbf{R}\mu_g \quad (12)$$

\mathbf{R} is a polar projection of \mathbf{K} . If $\det(\mathbf{R}) = -1$, this implies a reflection of the face in which case \mathbf{R} is calculated using Eqn. 13.

$$\mathbf{R} = \mathbf{V} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(\mathbf{U}\mathbf{V}^\top) \end{bmatrix} \mathbf{U}^\top \quad (13)$$

The above steps (from Eqn. 4 onwards) are iteratively repeated until the number of remaining points in \mathbf{P} are half the starting value i.e. $N \leq 0.5n_p$. The final value of error e_f between the two faces is calculated as

$$e_f = \sum_{i=1}^N (\|\mathbf{R}\mathbf{g}_i + \mathbf{t} - \mathbf{p}_i\| + \|\mathbf{R}\mathbf{g}_i + \mathbf{t} - \mathbf{p}_i\|(1 - k_i) + \frac{k_i}{2}) \quad (14)$$

All three terms in Eqn. 14 are first normalized on a scale of 0 to 1 for each recognition trial i.e. matching a single probe to all the gallery faces. This removes the bias between the terms. Moreover, the last term (k) is given half the weight of the other two terms.

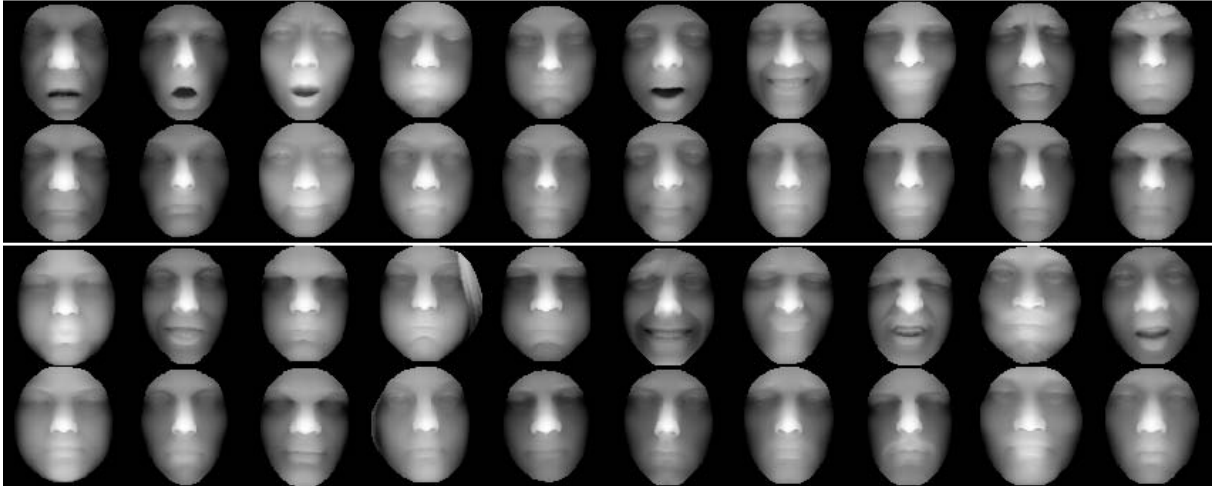


Figure 2: Sample 3D faces rendered as range images. First and third row contains the probe faces and second and last row contains their respective gallery faces.

4 3D Face Data and Experimental Setup

We used the FRGC v2.0 [14] data for our experiments which contains 3D faces along with their texture maps acquired with the Minolta Vivid 910 scanner [12]. However, we only used the 3D data for matching in this paper. There are 466 identities in the FRGC validation set. Out of these, only 374 individuals have a 3D face under neutral as well as non-neutral expression. For each of these 374 individuals, we selected one 3D face under neutral expression to form the gallery and one 3D face under exaggerated expression to form the probe. Selection was performed manually to ensure only those faces are selected which are significantly deformed due to facial expression e.g. blown cheeks and open mouth. Fig. 2 shows some example probes and their corresponding gallery faces. Moreover, where the choice of probe was to be made between a face that was covered with hair and a face that was not covered with hair, the former was selected to make the recognition extremely challenging.

Note that the aim of this paper is to compare the performance of the proposed SRICP algorithm to ICP under non-rigid (or semi-rigid) deformations using the same dataset. This is why we have only selected 3D probe faces under extreme facial expressions. We did not include the faces with neutral or minor expressions so that the results reflect matching performance on semi-rigid surfaces only. It is not the aim of this paper to achieve high recognition accuracy on the FRGC v2.0 data. Therefore, once the faces are preprocessed as described in [10], their alternate rows and columns are removed (i.e. the resolution is reduced by a factor of 4) in order to gain efficiency.

5 Results

We matched each probe to all the faces in the gallery once using the ICP algorithm and a second time using the SRICP algorithm. Fig. 3 shows the rank identification results. For each identification trial, the gallery faces are ranked according to their error scores e_f (Eqn. 14). A rank x recognition rate means the rate at which the correct identity is among the top x ranked identities. A rank one identification rate is the number of probes that were correctly identified (to its correct identity in the gallery) divided by the total number tested probes.

Fig. 4 shows our verification results. Each time a probe is matched with its correct identity in the gallery, the value of e_f is treated as a genuine score. However, when a probe is matched with a different identity in the gallery, the value of e_f is treated as an impostor score. The ROC curves are plotted as follows. The threshold for accepting a probe as a genuine client is varied and at every threshold, the verification rate is calculated as the number of genuine probes that fall below the threshold divided by the total number of genuine probes. Similarly, at every threshold, the False Acceptance Rate (FAR) is calculated as the number of impostors that fall below the threshold divided by the total number of impostors. At 0.001 FAR, SRICP achieves 62.57% verification rate whereas the ICP algorithm achieves 47.06% verification rate. These results show that SRICP outperforms the ICP algorithm.

The aim of these experiments was to perform an unbiased comparison of the SRICP and ICP algorithms on a challenging dataset. Note that the performance of both algorithms is quite low since

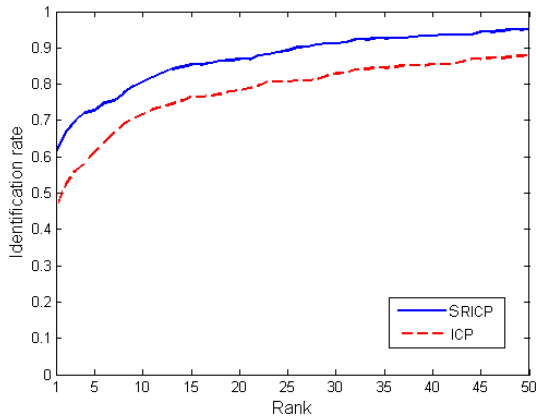


Figure 3: Identification results. SRICP achieves 61.23% rank one identification rate whereas the ICP algorithm achieves 45.99% rank one identification rate.

the data on which these experiments were performed was highly challenging due to exaggerated facial expressions. Moreover, the resolution of the faces was reduced (by a factor of 4) to gain computational efficiency. Therefore, these results can not be compared to others which used the entire FRGC v2.0 database (e.g. [10] and [13]) since the database also contains easy to recognize faces (i.e. with neutral and minor facial expressions). Moreover, these results can not be compared to those of Bronstein et al. [4] since their approach apparently does not deal with open mouth expressions.

6 Conclusion

We presented an algorithm for the registration and matching of semi-rigid 3D surfaces and demonstrated its performance on the challenging case of 3D face recognition under exaggerated facial expressions. Even though SRICP uses only half of the probe points for matching, it performs significantly better than the ICP algorithm which uses all the probe points from matching. SRICP is a generic algorithm and can be used for matching any 3D semi-rigid data. Moreover, like the ICP algorithm, it can also be extended to the nD case.

7 Acknowledgments

We would like to thank the organizers of FRGC [14] for providing the face data. This research is funded by an Australian Research Council discovery grant number DP0664228.

References

[1] K. Arun, T. Huang and S. Blostein, “Least-squares Fitting of Two 3-D Point Sets”, *IEEE TPAMI*, vol. 9(5), pp. 698–700, 1987.

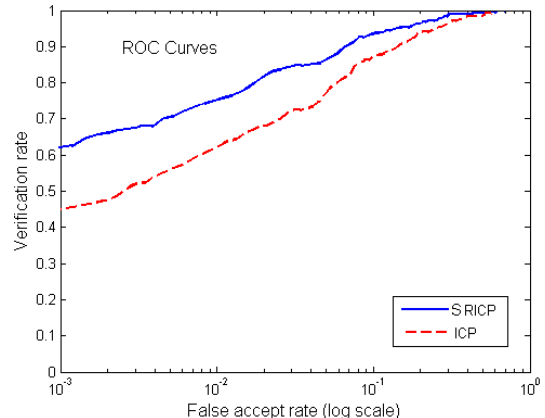


Figure 4: Verification results. At 0.001 FAR, SRICP achieves 62.57% verification rate whereas the ICP algorithm achieves 47.06% verification rate.

[2] P. J. Besl and N. D. McKay, “Reconstruction of Real-world Objects via Simultaneous Registration and Robust Combination of Multiple Range Images,” *IEEE TPAMI*, vol. 14(2), pp. 239–256, 1992.

[3] K. W. Bowyer, K. Chang and P. Flynn, “A Survey Of Approaches and Challenges in 3D and Multi-modal 3D + 2D Face Recognition,” *CVIU*, vol. 101, pp. 1–15, 2006.

[4] A. M. Bronstein, M. M. Bronstein and R. Kimmel, “Three-Dimensional Face Recognition,” *IJCV*, vol. 64(1), pp. 5–30, 2005.

[5] K. I. Chang, K. W. Bowyer and P. J. Flynn, “Multiple Nose Region Matching for 3D Face Recognition under Varying Expression”, *IEEE TPAMI*, vol. 28(10), pp. 1695–1670, 2006.

[6] J. Feldmar, B. Malandain, J. Geclerk, N. Ayache, “Extension of the ICP Algorithm to Non-Rigid Intensity-Based Registration”, *Workshop on Mathematical Methods in Biomedical Image Analysis*, 1996.

[7] X. Lu, A. K. Jain and D. Colbry, “Matching 2.5D Scans to 3D Models,” *IEEE TPAMI*, Vol. 28(1), pp. 31-43, 2006.

[8] A. S. Mian, M. Bennamoun and R. A. Owens, “A Novel Representation and Feature Matching Algorithm for Automatic Pairwise Registration of Range Images”, *IJCV*, vol. 66, pp. 19–40, 2006.

[9] A. S. Mian, M. Bennamoun and R. A. Owens, “2D and 3D Multimodal Hybrid Face Recognition”, *ECCV*, part III, pp. 344–355, 2006.

- [10] A. S. Mian, M. Bennamoun and R. A. Owens, "Automatic 3D Face Detection, Normalization and Recognition", *3DPVT*, 2006.
- [11] A. S. Mian, M. Bennamoun and R. A. Owens, "3D Model-based Object Recognition and Segmentation in Cluttered Scenes", *IEEE TPAMI*, vol. 28(10), pp. 1584–1601, 2006.
- [12] Minolta 3D Digitizers, "Non-contact 3D Laser Scanner", <http://www.minolta3d.com>, 2006.
- [13] G. Passalis, I. Kakadiaris, T. Theoharis, G. Tederici and N. Murtaza, "Evaluation of 3D Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach", *IEEE Workshop on FRGC Experiments*, 2005.
- [14] P. J. Phillips, P. J. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, "Overview of the Face Recognition Grand Challenge", *IEEE CVPR*, 2005.
- [15] S. Rusinkiewicz and M. Levoy, "Efficient Variants of the ICP Algorithm," *3DIM*, pp. 145–152, 2001.
- [16] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey", *ACM Computing Survey*, pp. 399-458, 2003.