Noise versus Facial Expression on 3D Face Recognition

Chauã Queirolo, Maurício P. Segundo, Olga Bellon, Luciano Silva *
IMAGO Research Group - Universidade Federal do Paraná
P.O. Box 19092, 81531-990, Curitiba - PR, Brasil
Email: {ccqs03, mps04, olga, luciano}@inf.ufpr.br

Abstract

This paper presents a new method for 3D face recognition. The method combines a Simulated Annealing-based approach for image registration using the Surface Interpenetration Measure (SIM) to perform a precise matching between two face images. The recognition score is obtained by combining the SIM scores of four different face regions after their alignment. Experiments were conducted on two databases with a variety of facial expressions. The images from the databases were classified according to noise level and facial expression, allowing the analysis of each particular effect on 3D face recognition. The method allows a verification rate of 99.9%, at a False Acceptance Rate (FAR) of 0%, for the FRGC ver 2.0 database when only noiseless, neutral expression face images are used. Also, the results using face images with expressions and noise demonstrate that subjects still can be recognized with 87.5% of verification rate, at a FAR of 0%.

1. Introduction

Face recognition systems have applications in many areas, *e.g.*, law enforcement, human machine interfaces, access control. Recently, 3D face recognition has been subject of many researches mainly because both the limitations of 2D images and the advances in 3D imaging sensors [4].

One approach for 3D face recognition is using image registration to perform face matching [1, 5, 6, 7, 12], being the Iterative Closest Point (ICP) mostly applied to this purpose. In order to evaluate the quality of registration in face matching the Mean Squared Error (MSE), sometimes combined with other measures, is employed.

The ICP is guided by the MSE but it was proved [18] that this measure could allow imprecise local convergence for range image registration, even when improved ICP-based approaches are used [9, 16]. Also, in [18] the authors suggest that the MSE is a good measure for starting the image registration process, but the Surface Interpenetration Measure (SIM) could be more suitable to be used at "the end of the game" to assess the quality of the registration.

In [2, 3] the SIM was presented for 3D face matching. The experiments were conducted using different databases, containing mostly faces with neutral expression, and the experimental results showed that SIM allows better discrimination between faces as compared to other metrics. Also, it was proposed a Simulated Annealing-based approach (SA) guided by SIM for 3D face matching, instead of ICP. This approach produced precise alignments, achieving higher contrast between faces from different subjects [3].

This paper presents a novel, comprehensive method for 3D face recognition using the SIM. This approach uses new combinations of the SIM produced from the matching of four different 3D face segmentation as authentication score. More extensive experiments were performed using two databases, our private IMAGO [1] Database and the Face Recognition Grand Challenge Database ver 2.0 (FRGC)[2], both containing faces with several facial expressions. The experiments were based in state-of-art works presented in the literature [5, 6, 12] and the results confirm that the SIM is a very discriminatory measure for 3D face recognition.

2. 3D Face matching

The 3D face matching is performed by combining range image registration techniques and the Surface Interpenetration Measure (SIM). In [3] some initial experiments were conducted on the FRGC ver 1.0 3D face image database comparing two registration approaches: (1) an improved version of the ICP [9, 16], and (2) a SA-based approach [10]. The experimental results have shown that the SA produces more precise alignments and a higher recognition rate as opposed to ICP. Although SA is slightly slower than ICP, it does not depend on initial pre-alignment.

*The authors gratefully acknowledge CNPq and FINEP for the financial support.

[1] http://www.inf.ufpr.br/imago
[2] http://www.frvt.org/FRGC/

Supported by the former results, the SA-based approach was improved and applied for range image registration combined with the SIM, now on a more extensive database, including many different facial expressions. A brief explanation of the SIM and the SA approach for range image registration are presented as follows.

2.1. Surface Interpenetration Measure

The SIM was developed by analyzing visual results of two aligned surfaces crossing over each other repeatedly in the overlapping area. The interpenetration effect results from the nature of real range data, which presents slightly rough surfaces with small local distortions caused by limitations of the acquiring system.

By quantifying interpenetration, one can more precisely evaluate the registration results and provide a highly robust control. Registrations of two range images presenting good interpenetration have high SIM values, and erroneous alignments produce low SIM values and that small differences in MSE can yield significant differences in SIM. Furthermore, alignments with high SIM present a very low interpoint distance between the two surfaces. That is, the SIM is a far more sensitive indicator of alignment quality when comparing "reasonable" alignments.

For more details of this measure, the reader should refers to [17, 18]. Also, to be applied in 3D face matching some constrains were applied to the SIM as described in [3].

2.2. Simulated Annealing-based approach for range image registration

Simulated Annealing (SA) is a stochastic algorithm for local search in which, from an initial candidate solution, it generates iterative movements to a neighbor solution that represents a better solution to the problem as compared to the current one. The main difference between SA and other local search algorithms, *e.g.* Hill Climbing, is that SA can accept a worse solution than the current candidate in the iterative process. Then, SA does not remain "tied" to local minima and because of this it has better chances to reach its goal, which is a solution close enough to the global one.

To apply SA on registration of two range images, six parameters (three parameters each for rotation and translation relative to a 3D coordinate system) are needed to define the candidate solutions as a "transformation vector" that, when applied to one image, can align it with the other.

Our SA-based approach basically has three main stages: (1) an initial solution is obtained by aligning the centers of mass of the two face images; (2) a coarse alignment is performed using a SA-based searching procedure to minimizes a robust evaluation measure, based on the MSAC robust estimator [19] combined with the MSE of the corresponding points between the two face images; and (3) a precise alignment is obtained by a SA-based searching procedure with the SIM as the evaluation measure, where the goal is to maximize the interpenetrating points between the two faces.

In the SA-based searching procedure, small random values within $[-1, 1]$ are introduced to each element of the transformation vector in an attempt to reach better neighbor solutions. The stop criteria is attested if the best solution does not change within n iterations (*i.e.* the system is frozen). We observed that to obtain a good alignment it was not required to use all the valid points. A sampling rate s of valid points equally spaced is used. The values for n and s are going to be defined in section 3. Although we only use a sample of points, the final SIM value that evaluates the alignment is computed taking into account all the valid points from both surfaces.

The "temperature" of the SA is reduced very slowly and two iterations are performed for each allowed "temperature" until the final one is achieved, otherwise the system becomes frozen [13]. The initial "temperature" was defined as $t = 0.002$ and $t = 0.15$, for stages (2) and (3), respectively. The MSAC threshold was defined empirically as 3.0, using a small dataset of images from FRGC ver 1.0 database, and than validated on a larger dataset.

3. 3D Face authentication

To accomplish face authentication, first the face images are automatically segmented in four regions (see Fig. 1): (a) circular area around the nose, (b) elliptical area around nose, (c) upper head, including eyes, nose and forehead, and (d) entire face region. Regions (a) and (b) were used because the nose area suffers less influence from facial expression as compared with other face regions [6]. However, in [3] it was stated that, when using the SIM, the nose region alone is not enough discriminatory for face recognition because it represents only a small fraction of the face. Due to this fact, regions (c) and (d) were also used because they contain more information about the face.

(a) (b) (c) (d)

Figure 1. Segmented regions from a same face. (a) circular nose area, (b) elliptical nose area, (c) upper head and (d) entire face.

The segmentation process uses our own approach based on the depth of the range maps to segment the images, which is composed basically by two main stages: (1) locating homogeneous regions in the input image by using clustering combined with edge data, and (2) identifying candidate regions that belong to the face region by an ellipse detection method based on the Hough Transform. More details of this approach are described in [14].

After this step, corresponding regions from both faces are registered independently. The authentication score is achieved by combining the final SIM value obtained for each segmented region using the sum rule [11]. This approach is more discriminant than the SIM value computed from each region by itself.

When two faces of the same subject are compared the authentication score tends to have a high value. Otherwise, if the faces are from different subjects this score has lower value. For the experiments a discriminant threshold was defined to allow a False Acceptance Rate (FAR) of 0%, using the same strategy described in [3]. With this FAR we can avoid the possibility of identifying a non-authorized person in an authentication system.

The SA parameters, n and s mentioned in section 2.2, were empirically defined according to the region that is used for matching. For both nose regions we set $n = 500$ and $s = 1.5\%$ in stage (2), $n = 50$ and $s = 15\%$ in stage (3). For other regions, in stage (2) we set $n = 300$ and $s = 0.1\%$, and in stage (3) $n = 20$ and $s = 10\%$.

4. Databases

In our experiments we used both the IMAGO Database and the Face Recognition Grand Challenge (FRGC) Database[3]. IMAGO database has 150 images, from 30 subjects, acquired using a Minolta Vivid 910 scanner. Five images were taken from each person with four different facial expressions (3 neutral, 1 small happy and 1 very happy). Some images from this database are shown in Fig. 2.

The FRGC Database is the largest available database of range face images and works as benchmarking to evaluate algorithms for face recognition [15]. It is composed by 4,007 range images, from 466 subjects, mostly frontal and diversified facial expressions. Besides the great number of images, they have several particular features that may interfere in face matching, such as noise.

In order to make an accurate analysis from the proposed method and identify its weak points, all images were arranged according to their noise level, expression type and level. Since the provided facial classification from FRGC did not correspond to the ground truth, the new arrangement was based on the six basic emotions proposed by [8].

[3]The authors would like to thanks Dr. Jonathon Phillips and Dr. Patrick Flynn for allowing us to use the images.

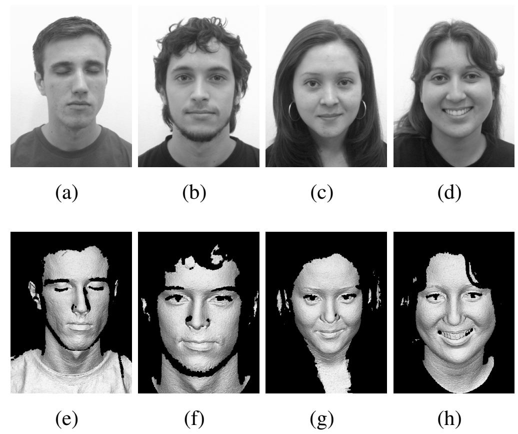


Figure 2. Example of images from IMAGO Database. (a) and (b) neutral expression, (c) small happy and (d) very happy.

The following classes were established: neutral, neutral with open mouth (O.M.), happy, surprise, sad, frown and disgust. Except for neutral expression, those groups were also separated according to its strength. Fig. 3 shows some images from this database, classified by expression.

All the range images were divided based on three main kinds of noise: (1) very noisy, includes images that are deformed, stretched, without nose, and wavy surface, (2) little noisy, includes faces with beard/moustache, or small holes around nose, or waves around mouth, (3) noiseless. Since the images from (1) are very damaged they were discarded for the experiments. Examples of these images are shown in Figs. 3(o) and 3(p).

Figs. 3(b) and 3(i) show examples of little noisy effect, beard and holes around nose, respectively. Despite beard and moustache are facial features, they were classified as little noisy because the scanner does not precisely acquire the information of these areas. Also, in this database it was identified two claimed different subjects that are actually the same. This case was discarded for the result analysis.

5. Experimental Results

The experiments were performed intending to reproduce a genuine face recognition system, where we have the matching between an acquired face image with a previous registered one in a database. Usually, for genuine systems, the stored image has good quality, *i.e.* noiseless and neutral expression. For this reason, all neutral images were combined against each other, for both databases. The images with non-neutral expression were matched only against all other neutral images, and for the FRGC database only against noiseless neutral images.

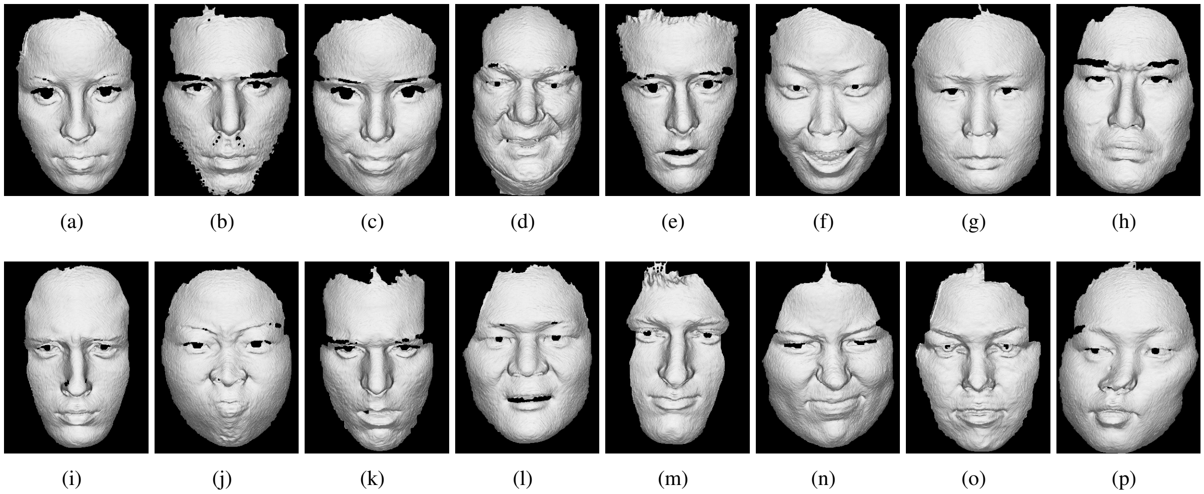


Figure 3. Images from FRGC Database. (a) and (b) Neutral Expression. Small Expression: (c) Happy, (e) Surprise, (g) Sad, (i) Frown, (k) Open Mouth, (m) Disgust. Larger Expression: (d) Happy, (f) Surprise, (h) Sad, (j) Frown, (l) Open Mouth, (n) Disgust, (o) and (p) Very noisy images.

For all images from both databases, the matching was performed using our SA-based approach for each one of the segmented regions and the similarity score was computed using the method described in section 3. Than, the combinations were split in several datasets according to its facial expression and noise level. For each dataset, a discriminant threshold was defined to allow a FAR of 0% [3].

One of the main worries about using SA to perform the 3D face matching in an authentication system is time constrains. Because of this, the experiments from IMAGO Database were executed in a controlled environment for time analysis, being totally computed 14,715 combinations. We used a computer with the following configuration: Linux operating system, AMD Sempron 2300+ processor, cache of 256KB and 512MB of memory.

The average time achieved for the nose regions was $0.55s$, and for the upper head and entire face regions the average times were $1.11s$ and $1.78s$, respectively. The maximum execution time achieved for combinations of same subjects was $3.78s$ using the entire face region.

The obtained results from IMAGO database are shown in Table 1. The S column shows the number of combinations from same subjects, TC is the total number of combinations, followed by the verification rate (VR) founded in that dataset. The last column is the number of combinations that were not authenticated. Those faces with low and high expression level are labeled as 1 and 2, respectively.

With the neutral expression dataset a verification rate of 100.0% was attained. However, this rate tends to decrease in the datasets with non-neutral expression images. It can be also observed that this verification rate is degraded when expression strength increases, comparing small happy with very happy, achieving a recognition rate of 96.6% and 83.3%, respectively.

Table 1. Results from IMAGO Database.

Datasets	S	TC	$VR\,(\%)$	E
Neutral	90	4,005	100.0	0
Happy 1	90	2,700	96.6	3
Happy 2	90	2,700	83.3	15

Table 2 presents the results from FRGC Database for faces with neutral expression. The labels from this table are the same ones described previously.

Table 2. Results from FRGC Database using faces with neutral expression.

Datasets	S	TC	$VR\,(\%)$	E
noiseless	2,489	434,778	99.9	1
all	7,286	2,379,471	99.2	53

The proposed method can indeed distinguish if two faces, neutral expression, belong to the same subject with a verification rate of 99.2%, at a FAR of 0%. From the noiseless dataset, only one combination was not recognized, because one of the faces was slightly non-frontal and the images do not have enough overlapping area in common. Besides, it can be observed in Table 2 that noise has significant effect to recognition of faces with neutral expressions.

Although noise affects greatly the results for neutral face images, its effect seems to be not so relevant in analyzing the results for faces containing some level of expression. A possible explanation is that an expression has greater influence to the alignment than noise regarding bad registration results. In this context, Table 3 shows the individual results among the different types of facial expression using all images of the datasets (noisy and noiseless).

Table 3. Results from FRGC Database using faces with non-neutral expression.

Datasets	S	TC	$VR\,(\%)$	E
Happy 1	707	205,088	91.6	59
Happy 2	643	245,514	58.1	269
Surprise 1	208	55,216	93.2	14
Surprise 2	590	169,592	86.4	80
Sadness 1	233	80,852	90.9	21
Sadness 2	140	56,202	94.2	8
O.M. 1	139	63,104	95.7	6
O.M. 2	598	209,032	90.8	55
Frown 1	71	31,552	83.0	12
Frown 2	200	51,272	80.5	39
Disgust 1	43	12,818	100	0
Disgust 2	96	58,174	30.2	67

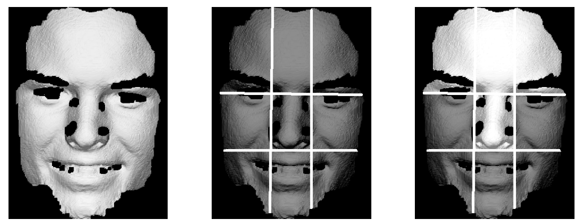
When analyzing the results for other facial expressions, higher expression levels becomes harder to identify as compared to low level expression, as expected. Although, it can be seen that the proposed approach still can discriminate many different expressions, with a high verification rate.

Faces with disgusting and very happy expression are the most challenging ones to be verified. Those kinds of expression tends to greatly deform the nose region, and since all the segmented areas used from the face includes the nose, a lower performance was achieved. Figs. 3(n) and 3(d) show an example from these expressions.

Another experiment was performed in a dataset with all the images to observe how does the proposed method behaves in a situation were we have faces with different expressions and presenting little noise. The combinations were generated as described previously, totalling 3,617,887 combinations (10,954 combinations from the same subject and 3,606,933 combinations from different ones). In this context a verification rate of 87.5% was achieved, at FAR of 0%. If this FAR is relaxed to 0.1% the verification rates is improved to 98.5%. In fact, the goal of the competition Face Recognition Grand Challenge [15] is a verification rate of 98% at FAR of 0.1%.

6. Enhanced SA approach

Because faces with happy expression are the most common among people during taking pictures and they are challenging for recognition, we propose an alternative approach for matching. By observing the matching of faces with happy expression we noticed that the most invariant regions are the forehead and the nose areas. Firstly, each happy face is divided in nine sectors based on four facial feature points (see Fig. 4(b)). During SA registration process, the matching tends to be stronger on those invariant areas. Fig. 4(c) shows the selected less invariant regions.



(a) (b) (c)

Figure 4. Sectors used for enhanced SA approach. (a) Original face image, (b) Sectors and (c) Brighter sectors used for matching.

In the original matching process the SIM is computed based on the number of interpenetrated points. For this enhanced method, a point that is interpenetrated and belongs to one of the invariant regions receives a high weight, otherwise, the applied weight is 1. The final SIM score is achieved by giving to all interpenetrated points weight 1.

Experiments were conducted for all the combinations from the same person that were not recognized from each dataset, and had happy facial expression. Another experiment that was performed used only the not recognized neutral faces from FRGC. Also, to verify if the matching score for combinations between different subjects are improved a set of these combinations were tested. On IMAGO database all of these combinations were used, and for FRGC database 4000 combinations were randomly generated.

The enhanced method was applied only to entire face region, and the final authentication score is still achieved as described in section 3. Results from this approach are shown in Table 4. First column indicates the database, followed by the datasets that were used. The last two columns show the number of combinations from same subjects that were not recognized using the original approach and the enhanced one, respectively.

Table 4. Results for Improved SA approach.

Base	Dataset	Original	Improvement
IMAGO	Happy 1	3	1
	Happy 2	15	10
FRGC	Neutral	53	42
	Happy 1	59	54
	Happy 2	269	224

Observing results from Table 4, one can see that using the second approach for face region matching the verification rate can be improved in more than 15% for faces with high expression level. On the neutral expression dataset, from those images that were not recognized by the original approach, this is improved in 20% by using the enhanced version of SA. Also, the results from different subjects combinations showed that the enhanced SA approach did not significantly improved their authentication score.

7. Final Remarks

In this paper, we presented a new method for 3D face recognition. The method combines a Simulated Annealing-based approach for range image registration with the Surface Interpenetration Measure. To evaluate this approach we used two different databases, IMAGO database and FRGC ver 2.0 database, containing a great number of face images, with several facial expressions and noise levels. In order to provide an accurate analysis regarding the behavior of our method, we classified the FRGC database according the facial expression and noise level.

By using this novel approach one can distinguish if two face images with neutral expression belong to the same subject or not with a verification rate of 99%, at a FAR of 0%. The results for faces with expressions demonstrate that subjects still can be recognized with 87.5% of verification rate, at a FAR of 0%. Changing this FAR to 0.1%, this verification rate is improved to 98.5%. This rate reaches the goal of the competition Face Recognition Grand Challenge [15], which is a verification rate of 98% with a FAR of 0.1%.

We observed from the results that faces with disgust or happy expression are one of the most challenging to be recognized. For this reason, an enhanced SA approach was proposed to perform the matching of faces with happy expression. When comparing a face with neutral expression against one with very happy expression, the enhanced approach has performance over 15% when compared with the original one. Also, this approach can improve the matching of neutral expression faces in 20%. As future work, we plan to identify the presence of facial expression to improve the 3D face recognition.

References

[1] B. B. Amor, M. Ardebilian, and L. Chen. New experiments on icp-based 3d face recognition and authentication. In *ICPR*, volume 3, pages 1195–1199, 2006.

[2] O. Bellon, L. Silva, and C. Queirolo. 3D face matching using the surface interpenetration measure. In *Lecture Notes in Computer Science*, volume 3617, pages 1051–1058. Springer-Verlag, 2005.

[3] O. Bellon, L. Silva, C. Queirolo, S. Drovetto, and M. Pamplona. 3D face image registration for face matching guided by the surface interpenetration measure. In *IEEE ICIP*, pages 2661–2664, 2006.

[4] K. W. Bowyer, K. I. Chang, and P. J. Flynn. A survey of approaches to three-dimensional face recognition. In *ICPR*, pages 358–361, 2004.

[5] K. I. Chang, K. W. Bowyer, and P. J. Flynn. Adaptive rigid multi-region selection for handling expression variation in 3D face recognition. In *Proc. IEEE CVPR - Workshops*, volume 3, page 157, 2005.

[6] K. I. Chang, K. W. Bowyer, and P. J. Flynn. Multiple nose region matching for 3d face recognition under varying facial expression. *IEEE PAMI*, 28(10):1695–1700, 2006.

[7] J. Cook, V. Chandran, S. Sridharan, and C. Fookes. Face recognition from 3d data using iterative closest point algorithm and gaussian mixture models. In *3DPVT*, pages 502–509, 2004.

[8] P. Ekman. *Emotion in the Human Face*. Cambridge University Press, 1982.

[9] N. Gelfand, L. Ikemoto, S. Rusinkiewicz, and M. Levoy. Geometrically stable sampling for the ICP algorithm. In *Proc. 3DIM*, pages 260–267, 2003.

[10] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[11] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE PAMI*, 20(3):226–239, 1998.

[12] X. Lu, A. K. Jain, and D. Colbry. Matching 2.5D face scans to 3D models. *IEEE PAMI*, 28(1):31–43, 2006.

[13] M. Lundy and A. Mees. Convergence of an annealing algorithm. *Mathematical Programming: Series A and B*, 34(1):111–124, 1986.

[14] M. Pamplona, C. Queirolo, O. Bellon, and L. Silva. Face and facial features detection for 3d face recognition under facial expression variations. In *to appear in ICIAP*, 2007.

[15] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. IEEE CVPR*, pages 947–954, 2005.

[16] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. 3DIM*, pages 145–152, 2001.

[17] L. Silva, O. Bellon, and K. Boyer. *Robust Range Image Registration Using Genetic Algorithms and the Surface Interpenetration Measure*, volume 60 of *Machine Perception and Artificial Intelligence*. World Scientific Publishing, 2005.

[18] L. Silva, O. Bellon, and K. Boyer. Robust range image registration using the surface interpenetration measure and enhanced genetic algorithms. *IEEE PAMI*, 27:762–776, 2005.

[19] P. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *CVIU*, 78:138–156, 2000.