

Feature-based error processing for robust surface registration in computer assisted orthopedic surgery

Maxime Taquet #¹

*Communications and Remote Sensing Laboratory (TELE), Université catholique de Louvain
Place du Levant, 2, 1348 Louvain-La-Neuve, Belgium*

¹maxime.taquet@uclouvain.be

Abstract—Accurate registration algorithms are required in computer assisted orthopedic surgery to provide a reliable correspondence between the bone and its tridimensional model. In surface registration, the main source of errors lies in the points acquisition process. This paper presents an online point processing algorithm based on features correlated to the error. The features are organized in classification trees to detect and remove outliers. Different common classification trees are investigated. The efficiency of the approach to remove outliers is demonstrated through experiments on a pelvic bone.

I. INTRODUCTION

Computer assisted orthopedic surgery requires the registration between the patient's bone in the operating room and a preoperative tridimensional image of this bone [1]. This returns the transform between the coordinates in the reference frame of the computer model and those in the patient's frame. This transform is then used in real time to project the surgeon's tool position in the computer model. Any error made in the registration will therefore propagate to the location of the tool. The registration accuracy has thereby a strong impact on the overall accuracy of the surgery [2], [3].

A surface registration is typically used for orthopedic surgery [4]. The surgeon acquires a point cloud on the bone surface using a probe whose position is recorded by a tracking camera. A gradient-descent optimizer seeks the transform that minimizes a metric representing the distance between the projected digitized points and the bone surface in the computer model. Since the optimizer may be trapped in local minima of the metric, a preregistration step is used to provide a first approximation of the transform. This is typically performed by anatomical paired-point matching that uses three or four points acquired on the bone and their corresponding locations manually picked in the computer model.

Large inaccuracies can proceed from outliers present in the acquired point cloud. This problem has been identified by many researchers and methods to solve it were suggested. Zhang et al. [5] attempted to identify outliers in order to prevent misregistration of free-form surfaces. Palombara et al. [6] reported the consequences of outliers and proposed a method to remove them in surface registration for total knee arthroplasty. Their algorithm is based on a first registration and the use of a threshold value on the estimated matching error. Other researchers such as Rangarajan et al. [7] and Ma et al. [8] attempted to reduce the influence of outliers by means of a robust metric.

Three problems can be identified with these algorithms. First, they do not completely eliminate the influence of outliers on the final result and may thereby fail to be robust if the number of outliers is too high. Second, since the threshold filter processes points offline, the registration accuracy may suffer from a low cardinality of the point cloud after the filter. Third, the threshold method assumes a Gaussian distribution of the matching errors which is not the case when far outliers are present [9].

The novel approach presented in this paper suggests the use of a removal technique based on features statistically correlated to the matching error. This removes most outliers online, i.e. during the point acquisition itself, and yields a remaining error distribution closer to a Gaussian. It can be combined with an offline removal technique and a robust metric to lead to a significant improvement of the overall registration accuracy.

The algorithm developed to remove outliers and its integration is explained in section II. Section III presents the main results of the presented algorithm. Finally, section IV concludes and discusses the main contributions of the presented approach

and suggests further research directions that may improve its efficiency.

II. METHOD

The main idea of the method used to remove outliers online is to analyze features related to the acquired point i and correlated to its acquisition error $e_{\text{acq}}(i)$, i.e. its distance from the bone surface. The features can be organized in classification trees whose goal is to decide whether a point is regular or is an outlier.

The development of this method is performed in two steps. First, the features are selected to define a feature space in which outliers and regular points can be separated. Second, the features are organized in a single structure based on classification trees to detect outliers.

A. Feature space

Information can be extracted during the acquisition process to detect the presence of outliers. Basically, two categories of information can be pointed out: the preregistration information and the probe dynamic information.

Choices need to be made in order to derive analytic expressions for the features. These expressions should lead to a representative yet concise feature space. Throughout this section, the Kullback-Leibler divergence (KLD) is used to guide these choices [10]. The KLD measures the quality of a feature to give information about the acquisition error. If E is this acquisition error and Y is a feature, both defined as random variables with distributions $p(e)$ and $p(y)$ respectively, then the KLD measure is defined as:

$$I_{\text{KLD}}(E, Y) = \int_{\mathbb{R}} \int_{\mathbb{R}} p(e, y) \log_2 \frac{p(e, y)}{p(e)p(y)} de dy \quad (1)$$

The higher the KLD of a feature, the more correlated it is to the acquisition error.

Technically, an experimental sample of 2,700 points was acquired along with the value of many potential features Y_i and their acquisition error considered as realizations of E . This acquisition error can be computed by means of a prior accurate registration. The KLD $I_{\text{KLD}}(E, Y_i)$ was then computed to select the most relevant features.

1) *Pre-matching distance*: Since the preregistration is performed before the point set acquisition, the transform returned (\mathcal{T}_{pre}) can be used to compute the pre-matching distance $d_{\text{pre}}(i)$ for any point i . This quantity is a rough approximation of the acquisition error.

However, since \mathcal{T}_{pre} may not be very accurate, an error $e_{\text{pre}}(i)$ called the preregistration error may corrupt the approximation, i.e.

$$d_{\text{pre}}(i) = e_{\text{acq}}(i) + e_{\text{pre}}(i) \quad (2)$$

There is no accurate a priori knowledge about the distribution of this error over the volume since it strongly depends on the location of the paired points used in the preregistration as well as the accuracy with which they were acquired. To cancel the effect of the preregistration error, $d_{\text{pre}}(i)$ can be compared to the pre-matching distance of points within some neighborhood. The neighborhood must be small enough to legitimately consider locally constant preregistration errors. However, it must not be too small unless contiguous outliers may be interpreted as the result of the preregistration error and no more be detected.

To efficiently define the relative pre-matching distance, assumptions should be made about the distribution of outliers along the sampling track. It will be assumed that at most two contiguous outliers and at least three contiguous regular points can be encountered. These assumptions were observed to be legitimate at common sampling rates used in orthopedic surgeries (3 acquisitions/sec.) no matter the speed at which the surgeon moves his hand. Indeed, the time required for the probe to come back onto the bone after it slipped off was observed to be approximately constant. A digital high pass filter could be used instead and would probably be more robust with respect to the sampling rate. However, for a fix sampling rate, a specific analysis of the outliers occurrence is expected to yield better results.

All the patterns of five contiguous points respecting the above assumptions are the ones depicted in Figure 1 and their symmetrical with respect to the central point. Their complementary patterns, i.e. the patterns obtained by replacing outliers with regular points and reciprocally, if observed, must be the result of the preregistration error which cancels out the distance of outliers and exaggerates the acquisition error of regular points. Therefore, outliers can be detected even in case of large preregistration errors.

The above analysis brings about a natural definition of a continuous feature correlated to the acquisition error. The idea is to assume that, in any group of five contiguous points, the one with the lowest pre-matching distance lies on the bone. Its pre-matching distance is therefore equal to the preregistration error. This preregistration error can be subtracted to each one of the five pre-matching

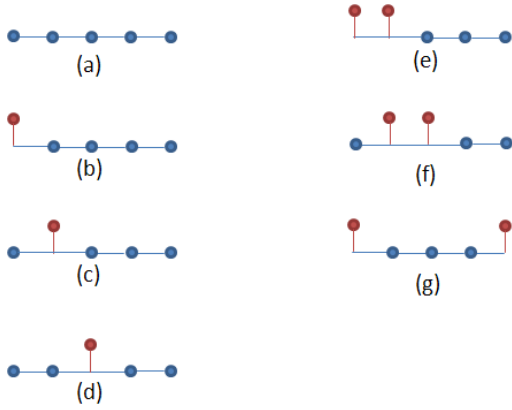


Figure 1. Different possible patterns, (a) with no outlier, (b)-(d) with one outlier, (e)-(g) with two outliers.

distances. If the pattern observed falls into one of the patterns described in Figure 1, then the distance of the point under analysis can be approximated by:

$$d_n(i) = d_{\text{pre}}(i) - \min\{d_{\text{pre}}(i+j) \mid |j| \leq 2\} \quad (3)$$

which defines a feature correlated to the acquisition distance. If the complementary of one of the patterns defined in Figure 1 is observed, then the point with the highest pre-matching distance is considered to lie on the bone and the feature correlated to the acquisition error is defined as:

$$\tilde{d}_n(i) = \max\{d_{\text{pre}}(i+j) \mid |j| \leq 2\} - d_{\text{pre}}(i). \quad (4)$$

2) *Probe dynamic*: When the probe slips off, its speed tends to increase. This speed is approximated by the central difference method using the probe position in three subsequent frames acquired by the tracking camera.

The speed must be normalized in order to take into account the discrepancies between surgeons and between areas of the bone. According to the assumptions about distance patterns for contiguous points, there are always at least three regular points in a group of five contiguous points. Therefore, the normalization term could be the mean of the speed corresponding to these three regular points. Since it is expected that these speeds will be lower than those corresponding to outliers, they can be defined as the three minimum speeds out of the five. The feature is then defined as

$$s_n(i) = s(i) - \text{mean}\{s(k), s(l), s(m)\} \quad (5)$$

where k , l and m are three different integers between $i-2$ and $i+2$ standing for the relative index of the points with minimal speed.

Other expressions of the speed such as the ratio of the speed and the mean of the minimum speeds or the use of a larger neighborhood with other assumptions on the distribution of outliers were also investigated but their computed KLD showed to be lower than the KLD of s_n by a factor of 1.5 at least. Besides, the probe acceleration was also in competition for the feature space. Its KLD was, however, observed to be three times lower than the KLD of s_n .

B. Classification trees

Classification trees can be represented as a set of nodes each defined by a feature and a threshold value. Data travel from node to node according to the result of the comparison of their feature value to the threshold. Their travel stops when they reach a leaf for which a decision is made about their class.

The detection of outliers can be interpreted as a statistical hypothesis test whose null hypothesis is to consider the point as a regular point. Two types of statistical errors can therefore occur: false positives (regular points classified as outliers) and false negatives (outliers classified as regular points).

The efficiency of classification structures can be assessed by the occurrence rates of both types of statistical errors. These rates can be computed through cross-validation. A 10-fold cross-validation is performed, i.e. the learning sample of 2,700 points is split into 10 subsamples of 270 points. The classification trees is grown using nine out of ten subsamples and subsequently processes the last subsample. This method prevents misinterpretations that would be induced if the data used to test the trees were the same as the data used to grow them.

Two main classification structures will be investigated: univariate classification and regression tree (CART) as first described by Breiman [11], [12] and extremely randomized trees (Extra-Trees) introduced by Geurts [13]. Two problems can be identified with these classifiers in their simplest shape. First, they do not deal with unbalanced data while the learning sample used contains much more regular points than outliers. Second, they do not allow weighing the misclassifications while in the case of outliers detection, one usually cares less about false positives than false negatives. Both these problems will be addressed by generalizing CART and Extra-trees.

1) *CART trees*: The learning step of CART trees selects the nodes by maximizing the gain in impurity defined as the difference between the impurity of the parent node and the impurities of

the resulting child nodes weighed by the number of records reaching them. Different definitions of the impurity can be provided [14]. However, the misclassification impurity allows generalizing the tree to deal with unbalanced classes. It is defined as

$$I(r) = 1 - \max_j P_j(r) \quad (6)$$

where j stands for the class and $P_j(r)$ is the proportion of learning data of class j at node r . Since probabilities sum up to one, $I(r)$ can be seen as the sum of all proportions of learning data that would be misclassified if the node was a leaf. Replacing this by a weighted sum leads to a straightforward generalization of the impurity that assigns different costs for the different misclassification errors. In the case of two classes, the weighted misclassification impurity can be written as

$$I_C(r) = C(t|1)P_1(r) + C(t|0)P_0(r) \quad (7)$$

where t is the decision made if the node r was a leaf, the ratio between the cost for false negatives, $C(0|1)$, and the cost for false positives, $C(1|0)$, is noted C . The value of C can be tuned by cross-validation to match the classification goal.

The tree needs to be pruned thereafter to prevent from overfitting. The leaves with the lowest impurity gain are pruned first. The number N of leaves to keep after pruning can also be tuned by cross-validation.

2) *Extra-Trees*: The problem raised with CART trees is that it results in a single tree whose node threshold values strongly depend on the learning sample [15]. One way to bypass this variance problem would be to use several trees and to perform a majority vote. Geurts introduced Extra-Trees for which randomization is introduced in the learning algorithm of the tree in order to generate a forest of several different trees [13].

The learning process of Extra-Trees consists in generating at each step of the iteration a set of random nodes and selecting the one with the highest impurity gain. If all computed gains have a negative value, another set of random nodes are generated.

Again, the method needs to be generalized in order to match the classification goal. Three generalizations are explored. First, the majority vote can be weighed in order to give more importance to pessimistic trees that classify the point in outliers. This method is called *votes generalization* and is parameterized by the weight W . Second, the

impurity gain used can be the weighed misclassification impurity as defined in (7) to define a *cost generalization* method parameterized by the cost C . Third, a *bootstrap aggregation* procedure can be used within the learning process [16]. In bootstrap aggregation, each classification tree is grown using only a subsample of the learning set made of all the outliers and P regular points per outlier.

III. RESULTS

The two main goals of the algorithm presented here are to remove most outliers and to yield an error distribution that is closer to a Gaussian. The different classification methods can be compared according to the number of false positives and false negatives that they introduce. This comparison as a result of cross-validation is depicted in Figure 2. The CART results are obtained using cost values C linearly increasing from 0 to 100. The bootstrap method is implemented with proportions P linearly increasing from 0.5 to 5. For the cost method, the parameter C is logarithmically increasing from 0.1 to 100,000. The weights of the weighed votes method are logarithmically increasing from 0.1 to 100.

The cost method yielded too many false negatives even with the highest cost value. No improvement were observed by further increasing the cost value. Therefore, a bootstrap with a proportion P of one was combined to it and the results of this combination are also depicted in Figure 2.

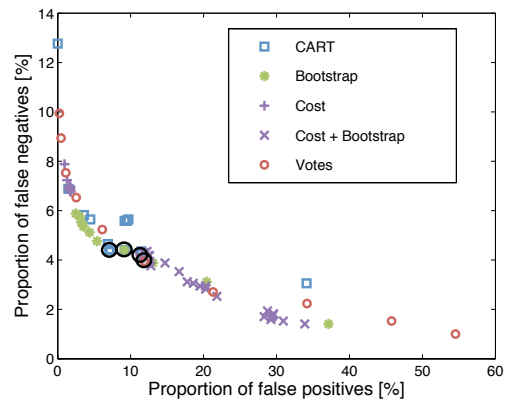


Figure 2. Statistical error rates of different classification methods. All methods have suitable results in the region of interest.

The region of interest of these results that combines efficient outliers removal and low enough false positive rates lies for false negative rates between three and five percents. In this region, four

different strategies have acceptable results allowing the removal of about two thirds of the outliers while removing between 5% and 15% of regular points. They are circled in black in Figure 2.

To choose between the four circled strategies, the standard deviation of the results needs to be taken into account. It assesses the reliability of the obtained results. In order to compute the standard deviations, each structure is grown ten times with random permutations of the data, so that the 10-fold cross-validation is performed with random splitting of the learning sample. The mean, μ , and the standard deviation, σ , of the removed points and false positives are summarized in table I.

Table I
PERFORMANCE OF THE GENERALIZATION METHODS AND THEIR STANDARD DEVIATIONS

Method	μ_{rem}	σ_{rem}	μ_{fp}	σ_{fp}
CART	57.7%	14.8%	10.1%	2.3%
Bootstrap	65.4%	0.7%	6.7%	0.3%
Cost + Bootstrap	66.8%	1.9%	7.3%	0.7%
Votes	69.0%	1.1%	13.4%	0.3%

This analysis of the variance clearly shows that CART should be avoided because of its large variance. None of the three others are outstanding. Depending on the desired sensitivity of the tree with respect to outliers and on the acceptable false negative rates, one could choose one or the other method. The following results of this section are obtained using the *votes generalization*.

To demonstrate the efficiency of the online removal method to yield an error distribution that is closer to a Gaussian, a quantile-quantile plot of this error versus a Gaussian distribution can be used [17]. These plots are depicted in Figure 3 and Figure 4 for the error distribution without and with online outliers removal respectively. It is clear that the removal of outliers brings the error distribution closer to a Gaussian.

The algorithm presented in this paper has been developed in the context of the development of a navigation system for computer assisted orthopedic surgery. The method was applied to the registration between a point cloud acquired around the cotyle of a pelvic bone and a computed model of the bone. Figure 5 compares the registration results with and without the online outliers removal. This method obviously removes most significant outliers located along the edges of the bone.

IV. CONCLUSION

This paper has introduced a novel point processing algorithm to remove outliers from an acquired

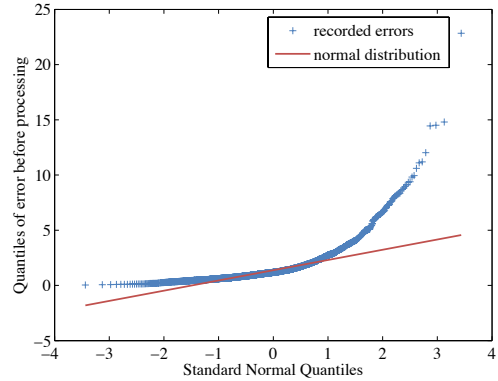


Figure 3. Q-Q plot of the error without online outliers removal. The red line represents the Gaussian distribution.

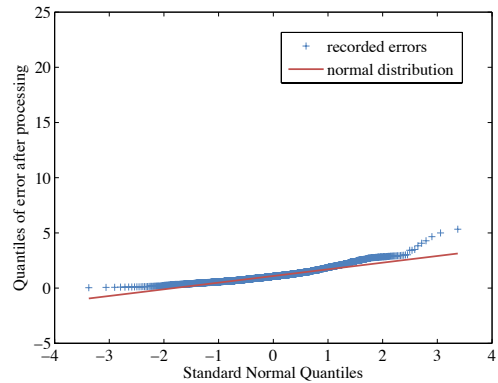


Figure 4. Q-Q plot of the error with online outliers removal. The red line represents the Gaussian distribution.

point cloud. While existing algorithms are all of-line methods, the algorithm presented here allows removing the outliers as they are acquired, based on features statistically correlated to the distance of the points from the bone. Two features have been selected: the relative pre-matching error and the relative speed of the probe.

A generalization of the extremely randomized trees has allowed combining these features in an efficient way. The experiment carried out showed that two third of the outliers can be removed by means of the generalized forest. The remaining outliers are expected to be removed by the threshold filter which is likely to be more efficient since the error distribution is now closer to a Gaussian.

Although cross-validation has been used to assess the actual performance of the algorithm, there may be a bias in this performance due to the fact that a single surgeon was asked to acquire the points of the learning sample. A more thorough

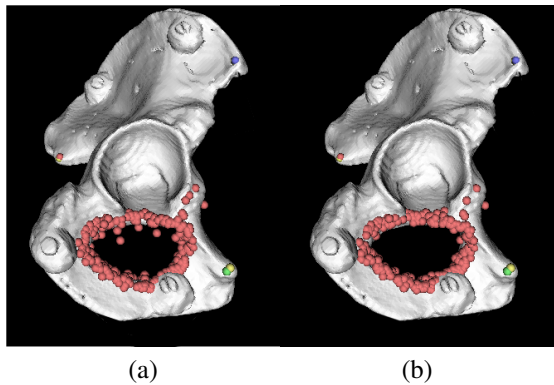


Figure 5. (a) Registration result of a point cloud containing outliers without using the removal method. Outliers are mostly present along the edges of the bone. (b) Registration result of a point cloud from which outliers have been removed online and replaced by additional regular points.

analysis with a larger learning sample combining the points acquired by different surgeons is required before the classification forest could indeed be used in a general way.

Finally, the small dimensionality of the feature space is required to reduce the computational cost of the learning process and to avoid overfitting the data due to the low number of elements of one class in the learning sample. However, the definition of the relative pre-matching error requires some strong assumptions about the preregistration error which, in practice, are rarely respected. This incorrect assumption is expected to induce most encountered misclassification errors. It was indeed observed that for some neighborhood, a linear model of the preregistration error would better represent the reality. This linearity could possibly be detected by multivariate classification trees. For these trees, the features used would be the bare pre-matching error of the five points constituting the neighborhood. Since the learning problem would be of higher dimensionality, the representativeness of the learning sample would be even more critical.

ACKNOWLEDGMENT

The author wishes to acknowledge Bruno Dehez and Benoît Macq for supervising the work presented in this paper.

REFERENCES

- [1] A. Hodgson, T. P. (eds.), and K. C. (eds.), *Image-Guided Interventions: Technology and Applications*. Springer US, 2008.
- [2] R. Phillips, "The accuracy of surgical navigation for orthopaedic surgery," *Current Orthopaedics*, vol. 21, pp. 180–192, 2007.
- [3] D. Simon, "Frameworks for evaluating accuracy in CAOS," August 1997.

- [4] J. Stiehl, W. H. Konermann, R. G. Haaker, J. B. (eds.), M. D. R. (eds.), and J. M. K. V. (eds.), *Total Knee Arthroplasty: A Guide to Get Better Performance*. Springer Berlin Heidelberg, 2005.
- [5] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Comput. Vision*, vol. 13, no. 2, pp. 119–152, 1994.
- [6] P. F. L. Palombara, M. Fadda, S. Martelli, L. Nofrini, and M. Marcacci, "A minimally-invasive 3-D data registration protocol for computer and robot-assisted total knee arthroplasty." London, UK: Springer, 1997, pp. 663–672.
- [7] A. Rangarajan, H. Chui, and F. L. Bookstein, "The softassign procrustes matching algorithm," in *Information Processing in Medical Imaging*, ser. Lecture Notes in Computer Science, J. S. Duncan and G. Gindi, Eds., vol. 1230. Springer, 1997, pp. 29–42.
- [8] B. Ma, R. E. Ellis, and D. J. Fleet, "Spotlights: A robust method for surface-based registration in orthopedic surgery," in *Medical Image Computing and Computer-Assisted Intervention*, ser. Lecture Notes in Computer Science, vol. 1679. Springer, 1999, pp. 936–944.
- [9] B. Ma and R. E. Ellis, "Robust registration for computer-integrated orthopedic surgery: Laboratory validation and clinical experience," *Medical Image Analysis*, vol. 7, no. 3, pp. 237 – 250, 2003.
- [10] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, pp. 79–86, 1951.
- [11] L. Breiman *et al.*, *Classification and Regression Trees*. Wadsworth, 1984.
- [12] Panel on Discrimination Analysis, Classification and Clustering, "Discriminant analysis and clustering," *Statistical Science*, vol. 4, no. 1, pp. 34–69, 1989.
- [13] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.
- [14] P. H. Kvam and B. Vidakovic, *Nonparametric Statistics with Applications to Science and Engineering (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2007.
- [15] T. G. Dietterich and E. B. Kong, "Machine learning bias, statistical bias, and statistical variance of decision tree algorithms," Department of Computer Science, Oregon State University, Technical Report, 1995.
- [16] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123–140, 1996.
- [17] H. Thode, *Testing for Normality; electronic version*. Abingdon: Dekker, 2002.