3D Reconstruction of Body Parts Using RGB-D Sensors: Challenges from a Biomedical Perspective

Pedro COSTA [1, 2], Hooshiar ZOLFAGHARNASAB [1,2], João P. MONTEIRO [2], Jaime S. CARDOSO [1,2], Hélder P. OLIVEIRA [2]
[1] Faculdade de Engenharia, Universidade do Porto, Porto, Portugal;
[2] INESC TEC, Porto, Portugal

http://dx.doi.org/10.15221/14.378

Abstract

The patient 3D model reconstruction plays an important role in applications such as surgery planning or computer-aided prosthesis design systems. Common methods use either expensive devices or require expert personnel which are not available in every clinic. Thus to make patient-specific modelling more versatile, it is required to develop efficient methods together with feasible devices. Body parts such as head and torso present valid challenges with different degrees of complexity, especially because of the absence of relevant and abundant features.
Considering Microsoft Kinect, it is a low-cost and widely available sensor, which has been successfully applied in medical applications. Since single depth-map acquired by Kinect is often incomplete and noisy, different approaches have been proposed to perform the reconstruction by merging multiple depth-maps, by registering single view point clouds generated form each point cloud. As human body is a non-rigid model, most of previous reconstruction methods using Kinect fail to perform accurate reconstruction since they do not address non-rigid surfaces.
In this paper we present the challenges of using low-cost RGB-D sensors to reconstruct human body. Additionally, we analysed coarse registration stage to understand its impact on the quality of reconstruction on both rigid and non-rigid data. Also comparative research has been performed to study different coarse registration methods such as Spin Image (SI), Curvedness, and Principal Component Analysis (PCA). Studies showed that the quality of reconstruction is directly related to robustness of reconstruction method to the rotational and translation noise. Regarding analytical comparisons, results indicate the positive impression of applying coarse registration on both rigid and non-rigid data. Moreover, evaluations show PCA presents better results among other considered methods. Finally it is shown that down-sampled models present less error.

Keywords: 3D models, calibration, multiple views, rigid and non-rigid registration

1. Introduction

In recent years, the emergence of three-dimensional (3D) scanning technologies has influenced a variety of areas of knowledge, including those related with biomedical research and the development of applications for healthcare. Forasmuch as the 3D representations of patients are to be used within a medical environment, it is required to use devices and methods with enough accuracy. In addition, to be used in a broad manner, issues like cost, transportability and usability requirements need to be taken into account. In that sense the introduction of consumer depth cameras like Microsoft Kinect, which capture real-time per-pixel depth information along with RGB images, has supported an extensive amount of related work [1]. Planning of surgery is an example which has been enhanced by the use of patients' 3D representations. For instance, in [2], Oliveira et al. utilized 3D data of patients to extract complementary features required for breast shape aesthetical objective evaluation, with the objective of evaluating the outcome of such medical procedure.

As a mid-resolution and low-cost device, the Microsoft Kinect has been proved to be accurate enough to be used as medical acquisition device in particular contexts [2, 3, 4, 5], however it is not compromise free. Beside limitations related to inevitable tolerances in its manufacturing process, its own characteristics (e.g. field of view, resolution, acquisition rate and working range) and different sources of noise affect any acquired single-view data, thus making it impossible to retrieve the 3D shape using only a single depth-map. In that manner it is often desirable to merge different views into a richer and more complete 3D representation of body parts. This process is called multi-view registration and can be done in several ways as depicted in figure 1.

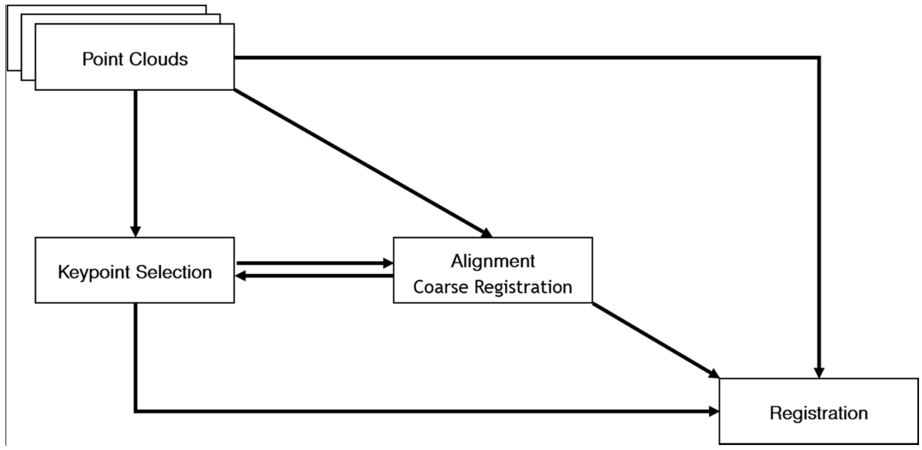

Fig. 1. The different possible approaches for multi-view registration.

A 3D representation of a body part can be obtained by: (a) merging the views directly, without any pre-processing; (b) detecting robust keypoints which provide a representative description of the point clouds and aligning the views bases on the transformation that aligns the keypoints; (c) aligning the views first and then performing the registration. The alignment stage is called coarse registration, and aims at computing an initial estimation of the rigid motion between the views, either based on keypoint detection or not. It is followed by a fine registration stage, which looks for the most accurate solution possible, being the Iterative Closest Point (ICP) [6] the gold standard methodology for fine registration. This is the standard registration approach, since the human body is nearly featureless and the preliminary alignment stage diminishes the computational complexity and the processing time. At each stage of the pipeline, inlier selection methodologies can be applied, such as the Random Sample Consensus (RANSAC). They look for the most representative subset of points, eliminating those which can compromise the results of the registration.

The point cloud registration process has been studied in [7] comprehensively. In it, Salvi *et al.* review the different registration approaches. We inferred about the capabilities of such methodologies towards the reconstruction of body parts and to meet the requirements of usability and complexity. Concerning keypoint extraction methodologies, Spin Images [8] and Curvedness [9] are two commonly used descriptors, respectively related with distances between points and planes and with bending energy of surfaces. On the other hand, for coarse registration (alignment stage), Principal Component Analysis (PCA) [7, 10], is expected to perform well.

The objective of this paper is to evaluate the usability of the aforementioned registration methodologies for the reconstruction of body parts. In the other hand, we intend to discuss challenges related with 3D reconstruction within medical context, concerning the acquisition and computing constraints with respect to quality of reconstructed objects. In this regard, this paper is organized in three parts. In section 2 related works on point cloud generation from RGB-D data and 3D reconstruction methods are discussed followed by a brief summary of prominent challenges. In section 3 we present a study of the impacts of using different coarse registration methodologies, together with the effects of number of points on the quality of reconstruction for both rigid and non-rigid datasets. Comparisons and further discussions on the studied coarse methods are presented in section 4. Finally in Section 5 the main conclusion of the present work are present.

2. Related work

Over the past few years, different research groups worked on the reconstruction of body parts using the Microsoft Kinect and were able to identify and overcome some of the existing challenges. Such previously identified challenges are related with calibration parameters, noise, acquisition and reconstruction software.

The Kinect uses structured light and machine learning tools to output RGB-D data that can be used to generate coloured point clouds given calibration parameters [11]. Although Microsoft Kinect provides manufacturing calibration data stored in the device's memory, alternative methods for Kinect calibration have been proposed ([12, 13]). Han *et al.* [14] recognized the valuable contribution of such works on the Kinect calibration since the precision of the point cloud generation increased.

Nonetheless, newer versions of the Software Developers Kit (SDK) already address such calibration related issues. The usage of factory supplied parameters avoids additional calibration procedures, thus improving usability and reducing the need for further intervention. However, due to possible structural changes in the device body, either by manufacturing tolerances and/or small structural changes that affect the Kinect components over time, there is still the necessity of re-calibration procedures.

Besides the accuracy of the calibration parameters, different sources of noise affect the Kinect data. Mallick *et al.* [14] recently presented a review paper systematizing Kinect noise-related work. In that regard, the authors highlight the fact that Kinect's depth sensor relies on the comparison of a known pattern against each new frame retrieved from its infra-red (IR) camera. Each frame is supposed to contain a specular pattern projected by Kinect itself and deformed by real world scene. Since it is an active sensor, not only multiple devices can interfere with the measurements of each other, but also other sources of IR light, like halogen operating room lights [16] or even direct sunlight, can degrade the reconstructed depth images. Also, transparent, specular, and reflective surfaces must be avoided since they greatly affect the projected dot pattern and are, therefore, prone to introduce noise in the data [14].

Different reconstruction studies that merge information from multiple-views can be found. In many situations, the patient is asked to rotate in front of the Kinect, which creates some occlusions and difficulties on the scanning (for instance, the underarms). Solutions that modify predefined silhouettes of the body while scanning, such in [17], have been proposed. Alternatively the usage of multiple Kinect devices [18] has been tested; however, some interference between devices has been observed [14], as well as an increase in the computational cost and a decrease on the portability. Nonetheless, it is possible to define a set of constraints, similarly to Tong *et al.* [18] in their three-Kinect scanning system. Concerning the usage of a single Kinect, the gold standard reconstruction software is the Kinect Fusion [19], developed by the homonymous Microsoft research group. It merges iteratively the incoming depth-maps into a single reconstructed model. However, the reconstruction is only accurate for rigid objects, presenting a high computational cost.

Other methods, although presenting accurate reconstruction results, but refer manual intervention as a limitation within the medical context. For instance, the methodology proposed by Li *et al.* in 2013 [20] can be mentioned. The authors claimed to perform the registration without the presence of operator, since user can select eight input views; nonetheless, for medical applications, neither patients nor clinic operators are expert enough to select appropriate views.

The human body itself is a non-rigid model, since patient breathes or moves unexpectedly during the acquisition, needing compensation parameters or deformable registration methods. In [21], the authors studied such problem and proposed a method which can compensate the effect of breath on the chest, while patient is being scanned.

2.1. Wrapping up challenges

The previously identified challenges related to the application of reconstruction methods within the biomedical context are listed in table 1Table. It is possible to understand that the majority of the previously identified challenges are related with either the used models or the acquisition set-up, demanding the identification of the reconstruction software-related challenges.

Table1. Comparing methods and identified challenges within medical context

	Heavy computation cost	Reconstruct only rigid objects	Restrictive acquisition set-up	Additional intervention necessary	Description
Tong et al. [18]			X		Requires 3 Kinects to be installed in predetermined positions;
Izadi et al. [19]	X	X			Need GPU processing due to heavy computation; remove moving objects;
Patete et al. [21]	X				Uses breath artefact compensation
Li et al. [20]				X	Requires expert personnel to select appropriate views
Herrera et al. [13]				X	Depth map quality improved trough calibration procedure

3. 3D reconstruction study

Motivated by the discussed challenges, in the following section we describe a simple 3D reconstruction study using the aforementioned registration, alignment and keypoint extraction methods. Two different experiments have been conducted using both rigid models and non-rigid models. The reconstruction of rigid models under different initial poses of the views allowed the characterization of the tested methodologies, and the reconstruction of the non-rigid counterparts gave insight about their applicability in a biomedical context. Such tests were run in Matlab 2013b and executed in a machine with Intel core i7 @ 3.70 GHz, 8GB RAM processor.

3.1. Data sets

3.1.1. Rigid models

The used rigid data set contains both the Horse and Bunny Models from Stanford Computer Graphics Laboratory[1], sampled to 10% and 50% of the original number of points. Different views were obtained by partitioning the entire point cloud to 10 views, each having overlapping areas with adjacent views. During the registration study, such models were disturbed with artificial initial misalignments in terms of rotation and translation in order to evaluate the robustness of the reconstruction methods. The two rigid models being used in this research were disturbed with artificial rotation from 0 to 35 degrees iterating each 5 degrees, while the translation (here defined as shift factors) are defined to be 1/100, 1/50 and 1/10 of the longest distance inside each point cloud.

3.1.2. Non-rigid models

The non-rigid data set contains two different models: a human head model and a female torso model from eight breast cancer patients. The acquisition of raw RGB-D data from the Kinect follows the protocol established within the context of the PICTURE Project[2]. The Kinect is placed on a tripod at 90 cm from the subject being scanned. The images are taken consecutively at 15 frames per second while the subject performs a 180º rotation between lateral views, with hands on the hips, in front of a homogeneous blue background panel and the region of interest of each model was manually segmented. The ground-truth for the female torso model was obtained using a commercial 3D scanner, and, the ground-truth for the human head was acquired using the Kinect Fusion.

Custom made software was implemented in C/C++ to process the data from the Kinect using the Microsoft Software Development Kit (SDK) v1.8. The calibration data extracted from the SDK was used to create multiple point clouds from the observed scene.

[1] https://graphics.stanford.edu/data/3Dscanrep/
[2] http://vph-picture.eu/

3.2. Methodology

3.2.1. Keypoint extraction

We evaluated the keypoint extraction stage from point clouds using two standard implementations of the Spin Image and the Curvedness algorithm to two adjacent point clouds belonging to the rigid data set. Concerning the usage of Curvedness for the feature detection, the work of Feldmar *et al.* in [22] [was implemented. First, the Gaussian and Mean curvatures for each point were calculated using the Moving Least Squares Algorithm, as implemented by Yang *et al.* [23]. Then, the maximum and minimum k_1 and k_2 curvatures were used to compute the curvedness at each point. A candidate was considered as a feature when it is a top curvedness point for three different neighbourhood scales: 90, 120 and 150 neighbours.

3.2.2. Registration

The registration study was performed following the methodology shown in figure 2. For the non-rigid data, we chose 3 views from the acquired data set and iteratively registered them into the same coordinate basis, using the Global Procrustes ICP [24], with or without previously alignment using PCA. It was proposed to use Generalized Procrustes Analysis to achieve the minimization of the cost function during the registration. While the stop criterion of the original algorithm was based on the number of iterations, we changed it slightly to represent both error-based and finite iteration stop criterion. New conditions are presented as follows: (1) The point clouds being registered remain almost unchanged in two consecutive iterations (the mean square error between consecutive poses of the point clouds is below 1×10^{-5} m); or, if the first condition is not met; (2) A maximum of 300 iterations has been reached.

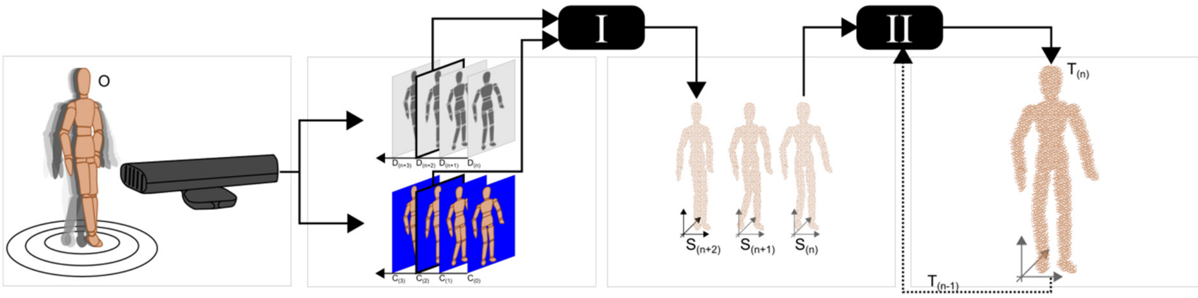

Fig. 2. The general workflow used in this work in order to evaluate reconstruction of non-rigid data. On the left, multiple views of a patient (O) were taken using a Microsoft Kinect. The depth (D(i)) and colour (C(i)) data is used to create point clouds S(i) of the model (I) that are then merged iteratively (II) towards a reconstructed model T.

Considering the rigid data registration, we discarded the first stage of the above presented pipeline and took the artificially misaligned views into account for the registration, simulating different acquisition set ups.

3.3. Evaluation

To evaluate the similarity between reconstructed and reference models, Euclidean distance measurements, such as Hausdorff Distance or Root Mean Square Error-related metrics, are commonly used [7]. Since the point clouds to be compared are not necessarily in the same coordinate system it is often necessary to pre-align them to eliminate distances between models, so that remaining distance between point clouds indicates the accuracy of used method. Such pre-alignment process can be performed using standard registration methods; however it is not guaranteed to pre-align models correctly and therefore the evaluation depends on both reconstruction and evaluation alignments.

Other methods are also proposed to evaluate two models by finding the norm distances form the origin of their coordinate, which use the norm of the difference between translation vectors. Same proposal is presented to find the difference between directional angle and rotation angle as the discrepancy of models' orientations [7]. Hence models are reconstructed in different coordinates, such metric puts more overload on system. Besides, misevaluation in symmetric models is probable since more than one directional angle may be found.

Using pose-invariant metric can solve the aroused issue, since point clouds can compared in any positions. Here, we designed an evaluation metric using distance measurements between histograms to compute the similarity of the models. To perform the evaluation, Delaunay Triangulation (DT) of both reference and reconstructed models are extracted. Two matrices are then generated and each one is filled by pairwise distances between all vertices of each DT. Following the conversion of each matrix to a histogram, the difference of the models is computed using cross-correlation of the histograms. Figure 3 depicts how the matrices are generated and filled by DT of each model. Regarding cross-correlation, as values approaches 1, it is concluded models are more similar, hence the distance between point clouds are less anticipating better reconstruction output. In reverse, near -1 value indicates mismatch, while 0 shows that there is no meaningful similarity between point clouds.

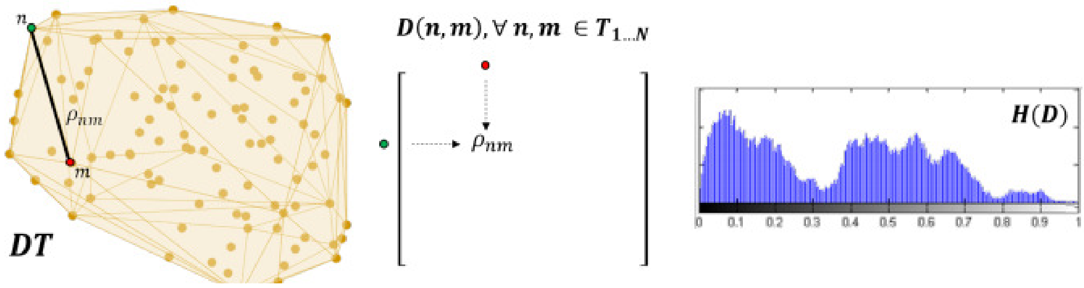

Fig. 3. Computing proposed metric by the histogram of DT of a point cloud

Besides, common methods such as Chi-square and Earth Movers distances can also be applied, but forasmuch as there is no significant deference between results obtained by them, they have been ignored in the metric to avoid redundancy. In this study, we evaluate our results using both the Hausdorff Distance and the designed Cross-Correlation measure.

4. Results and discussions

4.1. Inlier selection: RANSAC

The RANSAC inlier selection was tested using different point clouds. It was observed that for views with different information or with irregular shapes, it is very difficult to correctly define a sub-surface of inliers that correctly describes the model while being present in consecutive views. This makes the registration more prone to failure. On the other hand, the algorithm is not appropriate when time is critical because a high computational cost is required for a good performance

4.2. Keypoint extraction

4.2.1 Spin Images

The Spin Image algorithm was tested and, although good results can be found, they strongly rely on the resolution of the method. This means that, in order to have a good performance, the preliminary stage in which a triangular mesh is created needs to have a great number of elements, thus making the response time of the method not appropriate to be used under feasible time. Nevertheless, the methodology is expected to have fair results if the meshes for each view are already available. Such results suggest that a keypoint-based methodology should not be used, since it would greatly rely on the existence or not of features on the object of interest and the body, as previously discussed, is a nearly-featureless structure.

4.2.2 Curvedness

The results in figure 4 show that, although the represented views are consecutive and high-overlapping, the candidate point distribution is not similar, thus leading to wrong correspondences. Using such correspondences will lead the whole process of registration to be fail.

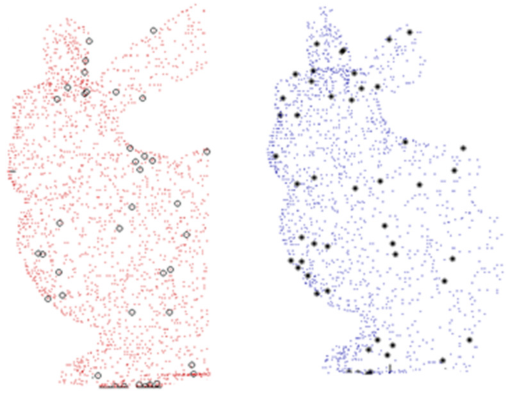

Fig. 4. Corresponding points from curvedness algorithm by [9]; although views have significant overlap, wrong points are selected as correspondences.

4.3. 3D reconstruction: rigid data

4.3.1. Cross-Correlation and Hausdorff Distance

The Cross-Correlation results in figure 5 show that standard methodologies are capable of reconstructing the models well when low initial misalignment levels are found. However, the results start to diverge early (ICP for 5 degrees, PCA+ICP for 10 degrees). The divergence point occurs very soon and avoids the usage of such methods for low-overlapping point clouds with very different initial pose. The problem is especially severe when high artificial translations and rotations exist, since the common methods seem to compensate rotation or translation separately, however not being able to produce a good output when both artificial misalignments exist. This avoids the usage of the Kinect on a medical environment and restricts its usage in low-controlled conditions and by non-experienced users.

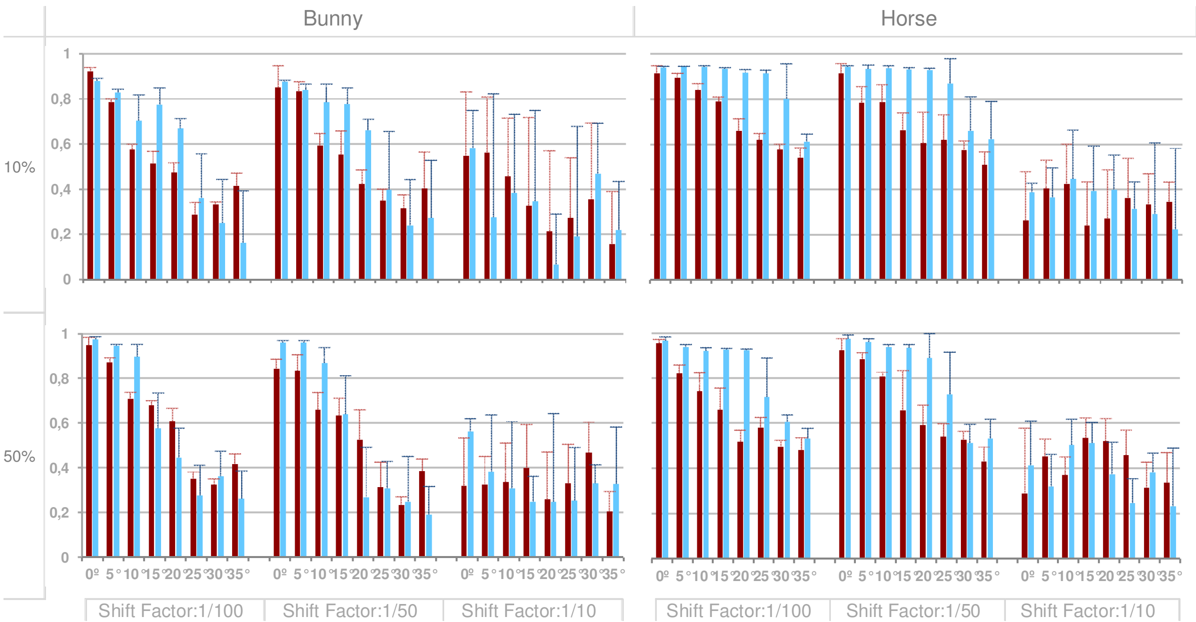

Fig. 5. The Cross-Correlation values [■ ICP; ■ PCA+ICP] shown as means with error bars depicting standard deviation (μ+σ). Both Bunny and Horse datasets results are presented for the sampling percentages of 10% and 50% of the points. A higher correlation corresponds to a better reconstruction. For each shift factor the results for a rotational noise from 0 to 35 degrees are shown.

In addition, results are presented in figure 6 using Hausdorff distances as metric. In both situations when registration is performed either with or without coarse step, as the differences between views increases, quality of reconstruction decreases. But as it was expected, both metrics show positive impression of coarse registration step on the quality of reconstructed model.

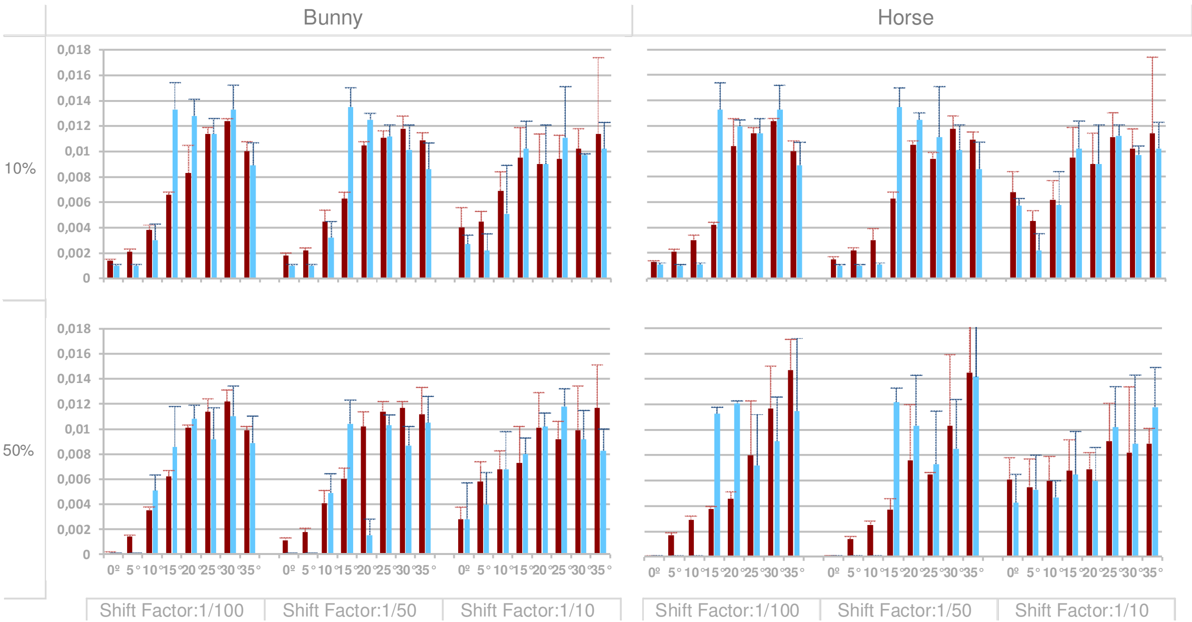

Fig. 6- The Hausdorff Distance values [■ ICP; ■ PCA+ICP] shown as means with error bars depicting standard deviation (μ+σ). Both Bunny and Horse datasets results are presented for the sampling percentages of 10% and 50% of the points. A higher Hausdorff Distance corresponds to a worse reconstruction. For each shift factor the results for a rotational noise from 0 to 35 degrees are shown.

For both models, it is possible to observe higher errors for the registration of views down sampled to 50% of points than to 10%. This can be explained by the easier propagation of errors on the alignment. If an ambiguous point-to-point correspondence is found, the rest of the correspondences are more prone to be erroneous. Such results suggest the need for a robust downsampling method that computes the alignment based on robust keypoints.

Finally, the usage of a coarse alignment stage (PCA) prior to the fine registration is desirable, since lower errors are found. This was expected, since the coarse alignment is intended to align the views in such way that the fine registration converges to the best solution possible.

4.3.2. Efficiency evaluation with respect to time

The efficiency evaluation results are shown in table 2. As expected, increasing the number of points, the registration takes more time. On the other hand, it is possible to observe that the PCA stage does not had any relevant computational complexity to the methodology, increasing the processing time in a feasible amount. A small increase on the processing time yields more accurate results as previously stated. Nevertheless, an even more efficient methodology should still be developed.

Table 2. The performance evaluation of different methodologies on the registraton of rigid models. The presented values are the average for 5 run under all specified conditions.

Points	**Methods**	**Time (sec) $\mu(\pm\sigma)$/view**		**# Iteration $\mu(\pm\sigma)$/view**
		Coarse registration	Fine registration	Fine registration
10%	ICP	-	12.568 (± 0.889)	199.613 (± 1.878)
	PCA+ICP	1.017 (± 4×10^{-4})	10.933 (± 1.977)	199.613 (± 1.878)
50%	ICP	-	46.133 (± 22.767)	115.134 (± 65.89)
	PCA+ICP	0.079 (± 0.004)	48.284 (± 21.701)	120.501 (± 64.867)

4.4. 3D reconstruction: non-rigid data

4.4.1 Point cloud generation

As already mentioned, the Kinect data presents different types of noises. In our acquisition system errors are notably observed at the edges of represented objects. This type of spatial noise, that Malick *et al.* [15] name lateral noise, creates irregular edges in a representation of objects which silhouettes are in reality smooth. It also originates along the contour of a given foreground, image elements to be expanded to background and/or the other way around, as seen in figure 7. Furthermore, lateral noise seems to depend on the spatial context of an object within the frame. A failed output triangulation over the border-adjoining background object was mentioned as a plausible cause by Mallick *et al.* [15], however, the authors do not present a usable solution for this problem. They lead not only to inaccurate depth boundaries or reduced lateral resolution but particularly when the depth map is mapped to the RGB image, visible artefacts can be observed.

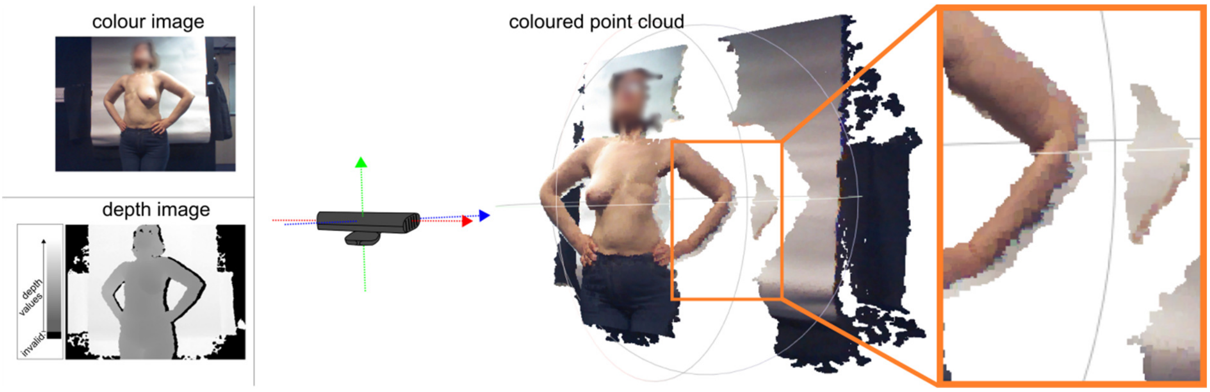

Fig. 7. Example of generated coloured point cloud from the Kinect RGB-D data, and highlighted (orange rectangle) presence of lateral noise.

4.4.2 Reconstruction of the female torso

Concerning non-rigid data, the female torso and the human head models were reconstructed. The output of methods is visually compared figure 8. The results indicate that reconstruction has not been performed well, since relevant structures of the breast, such as the nipples and the breast contour appear duplicated and noisy. Additionally, the texture of the reconstruction is highly inhomogeneous, meaning that the raw RGB-D data was affected by the illumination conditions the acquisition set up. The usage of PCA as coarse alignment, prior to the fine registration, suggests an improvement on the quality of the reconstruction. Nevertheless, the reconstruction is still not optimal as observable, meaning that a new coarse registration methodology could overcome the presented challenge.

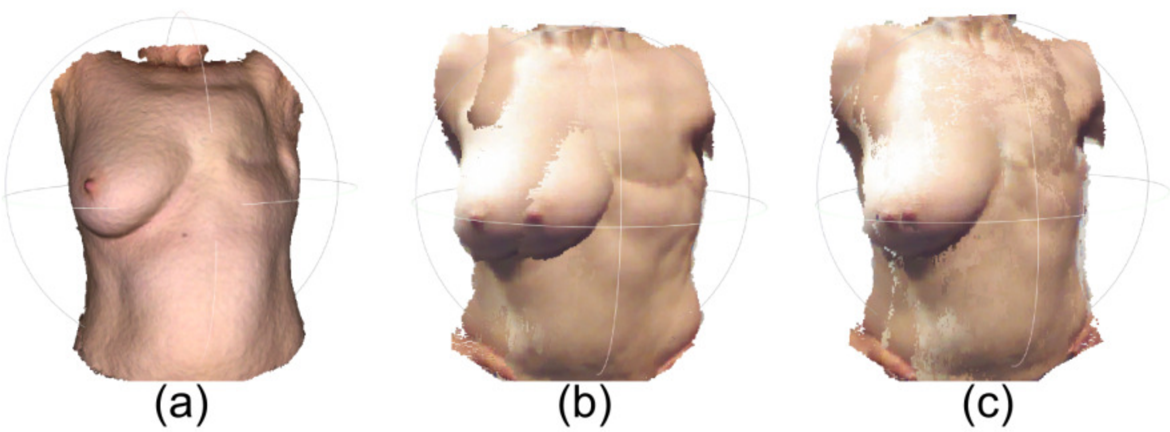

Fig. 8. Reconstruction of the female torso: (a) ground truth data by laser scanner, (b) reconstruction by ICP without coarse step and (c) reconstruction by PCA as coarse step and ICP

4.4.3 Reconstruction of the human head

As for the female torso, the reconstruction of the human head (shown in figure 9) is highly deformed and noisy. It is possible to observe the effect of the previously discussed lateral noise in both the reference (noisy texture on the Kinect Fusion reconstruction) and the reconstructed models (presence of grey noise on the face of the subject being reconstructed). The current registration pipeline seems able to approximate the views, however not being capable of merging them into an ideal reconstruction.

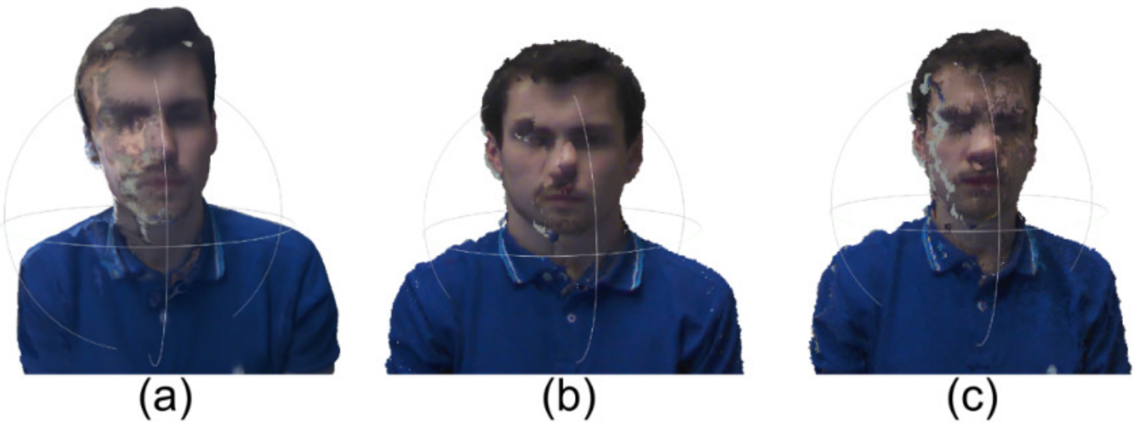

Fig. 9. Reconstruction of the human head, (a) ground truth data from Kinect Fusion, (b) reconstruction only by ICP and (c) Coarse registration (PCA) followed by ICP

5. Conclusion and future work

3D Body scanning is a technique which can be accompanied in biomedical procedures in order to ease the cure of diseases. However, requirements such as cost, timing and proficiency of users play determinant role in the usage of this technique. The advance of low-cost sensors, such as Microsoft Kinect has enhanced the quality of reconstruction.

In this paper, we aimed at analysing the challenges of using such low-cost device in biomedical applications. Considering the pipeline, in the first stage while patient is being scanned by Kinect, challenges such as patient movement and distance between patient and sensor are arose. Hence the output of Kinect is a depth map, regarding to calibration of the device it should be converted to point cloud. Existing challenges include noises caused by device, which have been analysed before.

The quality of reconstruction depends not only on both scanning device and registration steps, but also the characteristics of the scanning object that, generally, within medical context, comprehends bare body parts that can be regarded as near featureless, deformable or dynamic. Other than that, when the aligning single views present high rotational and/or translation differences from each other, common single stage registration methodologies often fail.

Due to registration algorithm, as it was proposed in other researches, using two-step registration procedure reconstructs better models. This claim was proved by conducting experiments using rigid models. It was shown that by increasing artificial misalignment, single step ICP fails, but in the presence of PCA as coarse registration procedure, better results were obtained. Nevertheless, the obtained results are not enough when dealing with high rotational and translational misalignments between views.

Performing accurate reconstruction requires methods which can withstand high degrees of rotational and translation noise. The importance of coarse registration stage is clarified when fine stage is too sensitive to views posture and outliers. Common coarse methods' requirements are not satisfying the data provided by low-cost sensors; hence they require large rigid overlapping areas, or reconstruction takes a lot of time. As the size of point cloud grows, common methods fail to perform the reconstruction task since they need to evaluate all the points to find correct correspondences.

Following the premises that (1) a coarse alignment prior to the registration is desirable but still not enough and (2) down-sampled versions of the views lead to lower errors, the obtained results suggest that a new coarse registration methodology is needed. Such methodology should take into account morphological information of the point cloud to choose robust keypoints (dealing in this way with the featureless nature of the human body) based on which the coarse alignment can be done. Additionally, colour information can be taken account to guide the reconstruction and guarantee that the texture of the reconstructed point clouds is reasonably close to the original information.

Moreover, regarding the evaluation metrics, Euclidean distances as metrics present feeble performance and are inconsistent with visual output since they require both comparing models to be completely aligned. To perform evaluation of the registration, metrics that posed invariant are required and should be searched for. The inclusion of normal vector, angles and distance information could be essential to develop a robust metric. Finally, future work must also focus on the point cloud generation stage, developing software that filters the generated point cloud towards the elimination of the effect of lateral noise and avoiding the presence of background information on the models.

Acknowledgments

The work leading to these results has received funding from the European Community's Seventh Framework Programme [grant number FP7600948]. This work was also financed by the European Regional Development Fund (ERDF) through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the Fundação para a Ciência e Tecnologia (FCT) – Portugal within project PTDC/SAU-ENB/114951/2009. The second author would like to thank FCT – Portugal for the financial support for the Ph.D. [grant number SFRH/BD/97698/2013]. We acknowledge Professor Keshtgar of the Royal Free Hospital in London for providing data.

References

[1] Z. Zhang, "Microsoft Kinect Sensor and Its Effect", *IEEE MultiMedia*, vol. 19, no. 2, 2012, pp. 4-10, http://dx.doi.org/10.1109/MMUL.2012.24.

[2] H. P. Oliveira, J. S. Cardoso, A. T. Magalhães and M. J. Cardoso, "A 3D low-cost solution for the aesthetic evaluation of breast cancer conservative treatment", *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 2, no. 2, 2014, pp. 90-106, http://dx.doi.org/10.1007/978-3-642-38628-2_31.

[3] H. Zolfagharnasab, J. S. Cardoso and H. P. Oliveira, "3D Breast Parametric Model for Surgery Planning- a Technical Review", *in Proceedings of 22nd Portuguese Conference on Pattern Recognition*, Lisbon, 2013, pp. 85-86.

[4] H. P. Oliveira, P. Patete, G. Baroni and J. S. Cardoso, "Development of a BCCT Quantitative 3D Evaluation System Through Low-Cost Solutions", *in Proceedings of the 2nd International Conference on 3D Body Scanning Technologies*, Lugano, Switzerland, 2011, pp. 16-27, http://dx.doi.org/10.15221/11.016.

[5] H. P. Oliveira, M. D. Silva, A. Magalhaes, M. J. Cardoso, J. S. Cardoso, "Is Kinect Depth Data Accurate for the Aesthetic Evaluation after Breast Cancer Surgeries?", *in Proceedings of the 6th Iberian Conference on Pattern Recognition and Image Analysis*, Funchal, Portugal, 2013 pp. 261-268, http://dx.doi.org/10.1007/978-3-642-38628-2_31.

[6] P. Besl and N. McKay, "A method for registration of 3-d shapes", *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 2, no. 114, 1992, pp.239–256, http://dx.doi.org/10.1109/34.121791.

[7] J. Salvi, C. Matabosch, D. Fofi and J. Forest, "A review of recent range image registration methods with accuracy evaluatio*n", Image and Vision Computing*, vol. 25, no. 5, 2007, pp. 578-596, http://dx.doi.org/10.1016/j.imavis.2006.05.012.

[8] A. Johnson, "Spin-images: A Representation for 3-D Surface Matching", PhD dissertation, Robotics Inst., Carnegie Mellon Univ., 1997.

[9] H. T. Ho, "3D Surface Matching from Range Images using Multiscale Local Features", PhD dissertation, Depar of Electrical and Electronic Eng., University of Adelaide, 2009.

[10] D. H. Chung, I. D. Yun and S. Lee, "Registration of multiple-range views using the reverse-calibration technique", *Pattern Recognition*, vol. 31, no. 4, 1998, pp. 457–464, http://dx.doi.org/10.1016/S0031-3203(97)00063-0.

[11] K. Khoshelham and S. Elberink, "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications", *Sensors*, vol. 12, no. 2, 2012, pp. 1437–1454, http://dx.doi.org/10.3390/s120201437.

[12] C. Zhang and Z. Zhang, "Calibration between depth and color sensors for commodity depth cameras", *in IEEE International Conference on Multimedia and Expo (ICME)*, 2011, pp. 1-6, http://dx.doi.org/10.1109/ICME.2011.6012191.

[13] D. Herrera C., J. Kannala and J. Heikkilä, "Accurate and Practical Calibration of a Depth and Color Camera Pair*", Proceedings of the 14th international conference on Computer analysis of images and patterns (CAIP),* vol. 2, Seville, 2011, pp. 437-445, http://dx.doi.org/10.1007/978-3-642-23678-5_52.

[14] J. Han, L. Shao, D. Xu and J. Shotton, "Enhanced Computer Vision with Microsoft Kinect Sensor: A Review*", IEEE Transactions on Cybernetics*, vol. 43, no. 5, 2013, pp. 1318-1334, http://dx.doi.org/10.1109/TCYB.2013.2265378.

[15] T. Mallick, P. Pratim Das and A. K. Majumdar, "Characterizations of Noise in Kinect Depth Images: A Review*", IEEE Sensors Journal*, vol. 14, no. 6, 2014, pp. 1731-1740, http://dx.doi.org/10.1109/JSEN.2014.2309987.

[16] M. Strickland, J. Tremaine, G. Brigley and C. Law, "Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field", *Canadian Journal of Surgery*, vol. 53, no. 3, 2013, pp. E1-E6, http://dx.doi.org/10.1503/cjs.035311..

[17] A. Weiss, D. Hirshberg and M. J. Black, "Home 3D body scans from noisy image and range data", *in IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1951-1958, http://dx.doi.org/10.1109/ICCV.2011.6126465.

[18] J. Tong, J. Zhou, L. Liu, Z. Pan and H. Yan, "Scanning 3D Full Human Bodies Using Kinects", *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 4, 2012, pp. 634-640, http://dx.doi.org/10.1109/TVCG.2012.56.

[19] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison and A. Fitzgibbon, "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera*", in ACM Symposium on User Interface Software and Technology*, 2011 pp. 559-568, http://dx.doi.org/10.1145/2047196.2047270.

[20] H. Li, E. Vouga, A. Gudym, L. Luo, J. T. Barron and G. Gusev, "3d self-portraits", *ACM Transactions on Graphics (TOG),* vol. 32, no. 6, 2013, pp. 187, http://dx.doi.org/10.1145/2508363.2508407.

[21] P. Patete, M. Riboldi, M. Spadea, G. Catanuto, A. Spano, M. Nava and G. Baroni, "Motion compensation in hand-held laser scanning for surface modeling in plastic and reconstructive surgery", *Annals of Biomedical Engineering*, vol. 37, no. 9, 2009, pp. 1877-1885, http://dx.doi.org/10.1007/s10439-009-9752-8.

[22] J. Feldmar and N. Ayache, "Rigid, affine and locally affine registration of free-form surfaces", *International journal of computer vision*, vol. 18, no. 2, 1996, pp. 99-119, http://dx.doi.org/10.1007/BF00054998.

[23] P. Yang and X. Qian, "Direct computing of surface curvatures for point-set surfaces", *in the Proceedings of the Eurographics Symposium on Point-Based Graphics*, 2007, pp. 29-36.

[24] R. Toldo, A. Beinat and F. Crosilla, "Global registration of multiple point clouds embedding the generalized procrustes analysis into an icp framework", *in the Proceedings of the 3DPTV Conference*, 2010, pp. 1-8.