AUTOMATING 3D RECONSTRUCTION PIPELINE BY SURF-BASED ALIGNMENT

*Maurício Pamplona Segundo, Leonardo Gomes, Olga Regina Pereira Bellon, Luciano Silva**

IMAGO Research Group, Universidade Federal do Paraná, Brazil

ABSTRACT

In this work, we automate a 3D reconstruction pipeline using two SURF-based approaches. The first approach uses SURF correspondences to pre-align multiple 3D scans from a same object without requiring manual labor. The second approach uses SURF correspondences to calibrate high resolution color images to 3D scans in order to improve the texture quality of the final 3D model. Both approaches succeeded in more than 95% of the test cases and were able to automatically identify incorrect results. Our pipeline has been widely used in several projects of cultural heritage and the proposed improvements are very important to allow its use in large collections. The proposed approaches were favorably compared to other methods and have been successfully applied in cultural heritage preservation of sculptures located in a UNESCO World Heritage site.

Index Terms— 3D reconstruction, cultural heritage, pairwise registration, texture calibration.

1. INTRODUCTION

The 3D reconstruction of objects and its application on cultural heritage preservation has received considerable attention in the last decade [1–4]. We are currently engaged in reconstructing Baroque sculptures from *Antônio Francisco Lisboa*, *"O Aleijadinho"*, from a UNESCO World Heritage site. These large soapstone sculptures are located outdoors and therefore are subjected to weathering, vandalism and other degradation factors. For this reason this research project aims to preserve the current geometry and texture of these sculptures in order to support future restoration processes, digital art preservation, social an digital inclusion through virtual museums, and so on.

This work is based on the 3D reconstruction pipeline described in [4–6], which is composed of six main stages:

Acquisition: it is performed by a sub-system that automatically controls both Konica Minolta VIVID 910 laser scanner and a Canon EOS 5D digital camera. The laser scanner captures the color and the geometric information (called I_s), and the digital camera captures a high resolution color image (called I_c).

Pairwise registration: it places all I_ss in a same coordinate system. To this end, I_ss pairs with overlapping regions are aligned by a variant of the Iterative Closest Point algorithm (ICP) [4] until there is a path of aligned images between every I_ss pair.

Global registration: Pulli's algorithm [8] is employed to homogeneously spread the error of the previous stage over the entire model to obtain a better global alignment.

Integration: a single volumetric model is created from the set of aligned I_ss [6], and holes are filled using the diffusion algorithm [9].

Mesh generation: the Marching Cubes algorithm [10] is employed to create manifold meshes.

Texturing: each I_s is projected into its respective I_c to obtain a realistic texture of the reconstructed object [5].

This pipeline has been successfully employed on indigenous artworks, small statues and natural assets [5]. However, the 3D reconstruction of large-scale objects presents some challenges that have required some modifications in the original pipeline [11]. The three most important improvements introduced in our pipeline and presented in this paper are the automatization of the following steps: both selection of overlapped views and pre-alignment of views, previously performed manually [11] which was a highly exhaustive task; and calibration between scanner and camera, previously recomputed when moving them around the object to capture new views. It works for small movable objects, but not for large fixed objects.

We investigated a number of pre-alignment methods proposed in the literature, including: Point Signatures [12], principal axes [13], scanning sequence [14] and coplanar points [15]. However, the main limitation of these approaches is that they are highly pose dependent [13, 14] or very computationally expensive [12, 15]. Considering the calibration problem, we evaluated different solutions: manual correspondences [16], edges [17], object silhouettes [18] and mutual information [19]. However these approaches are not applicable to large objects [16, 18] or require a good initialization [17, 19].

In this work, we propose a novel approach based on the Speeded Up Robust Features (SURF) [20] to solve both problems. First, we use SURF correspondences of a set of I_ss

* The authors thank UFPR, IPHAN, CNPq, and UNESCO for financial support.

from different views of the object to find overlapped pairs to be pre-aligned. Afterward, for each I_s and I_c we use SURF correspondences to obtain a transformation matrix that projects I_s into I_c. The pre-alignment stage was performed using 3D invariant features in other works [21], but such features are not applicable to the calibration problem.

2. AUTOMATING 3D RECONSTRUCTION

The SURF approach [20] extracts and describes interest points of an image in a fast, robust and reliable way. The SURF descriptor is both scale and rotation-invariant and may be used to find corresponding points in a pair of images.

This approach has been successfully applied to the image registration problem considering images taken from very different points of view [20], which is what makes it ideal for solving both encountered problems in the 3D reconstruction pipeline. It is also faster than other similar approaches, allowing the pre-alignment stage to be performed in real-time and identification of unscanned regions during the acquisition stage.

2.1. Pre-alignment of overlapped I_ss

To pre-align a pair of I_ss, first we extract SURF keypoints (k_s) from the color information of these images, then we assign 3D coordinates to each k_s using the geometric information, as illustrated in Fig. 1(a). After that, we obtain point correspondences by matching k_s descriptors of the first image against the ones of the second image, as shown in Fig. 1(b).

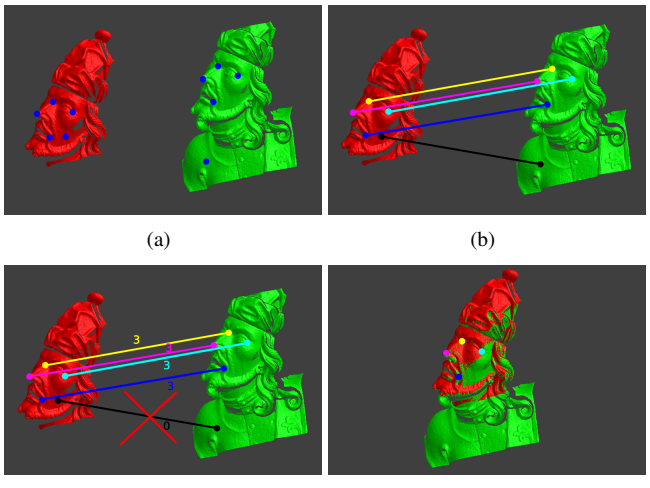

Fig. 1. Pre-alignment approach: (a) feature points; (b) correspondences; (c) consistency evaluation and overlapping detection; and (d) pre-alignment result.

Incorrect correspondences are filtered out by analyzing their spatial distribution. To this end, the distance between two points in the first image and the distance between their correspondent points in the second image should be equal, taking into account the scanner's accuracy. A point correspondence is valid if it satisfies the distance restriction with at least three other point correspondences, as shown in Fig. 1(c).

If there are three or more valid correspondences in a pair of I_ss, we assume they are overlapped. Then, we apply the RANSAC algorithm [22] over the set of valid correspondences to find the best transformation between the I_ss. The result of the pre-alignment using the proposed approach is shown in Fig. 1(d).

2.2. Calibration between scanner and I_cs

The calibration between scanner and I_cs starts by extracting SURF keypoints from the I_c (k_c), the k_s are already available, as illustrated in Fig. 2(a), and finding point correspondences between scanner and I_cs, as shown in Fig. 2(b).

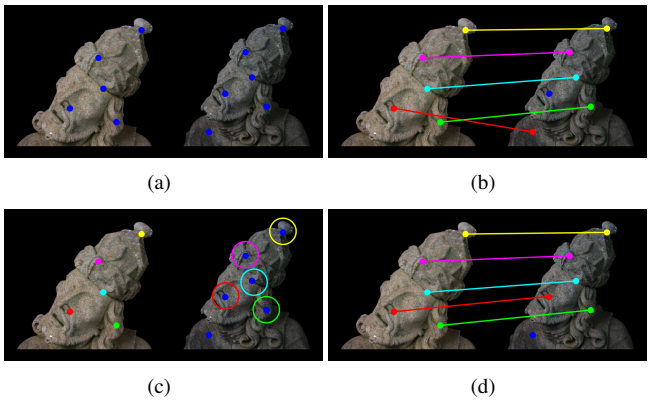

Fig. 2. Calibration approach: (a) feature points; (b) correspondences; (c) limited searching area; and (d) rectified correspondences.

To project the I_s into the I_c, we employed a camera model with seven parameters: translation and rotation in X, Y and Z axes, and focal length. Given four point correspondences, we use the Levenberg-Marquardt minimization approach to compute these parameters. The RANSAC algorithm is applied to find the subset of point correspondences that projects the maximum number of k_s into their correspondent k_c.

After that, the obtained camera parameters provide a rough estimate of the k_s location in the I_c. We use this estimation to rectify the correspondences by reducing the searching area to a spherical region around the estimated location, as shown in Figs. 2(c) and 2(d). Then, we recomputed the camera parameters using the same approach over the rectified correspondences to obtain the final calibration between scanner and I_cs, shown in Fig. 3.

(a) (b)

Fig. 3. Calibration result: (a) I_c and (b) the projected I_s.

3. EXPERIMENTAL RESULTS

Our experiments were performed using images from five large sculptures, each one about two meters tall, shown in Fig. 4.

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

Fig. 4. Reconstructed sculptures: (a) Daniel, (b) Ezekiel, (c) Hosea, (d) Joel and (e) Jonah.

Daniel, Ezekiel, Hosea, Joel and Jonah are composed by 178, 105, 117, 161 and 156 views, respectively. Each view is composed by a 640×480 range image and a 640×480 color image both acquired by the laser scanner, and a 4368×2912 color image acquired by the digital camera.

3.1. Pre-alignment results

In our first experiment we evaluated all possible combinations of views from a same object, and the proposed approach was able to find overlaps and obtain a correct pre-alignment in about 95% of the tested views. In most of them, the overlapping area is between 15% and 30% of the image. A pre-alignment is only considered correct if it is accurate enough for ICP to converge to the global minimum before 50 iterations. These results show the robustness of the proposed approach to low overlap, which is not common in the literature. Failures are mainly due to the lack of proper k_s or insufficient overlap. However, these failures are automatically detected by analyzing the resulting Root Mean Square Error (RMSE) in the ICP alignment. If the RMSE is below a threshold consistent with the scanner's accuracy, the pre-alignment is considered correct. Then, we manually generate the pre-alignment when necessary.

In our second experiment we manually selected 1,027 pairs of overlapped images in order to compare the proposed approach against three other state-of-the art works: principal axes [13], scanning sequence [14] and coplanar points [15]. These works were chosen because they can solve the pre-alignment problem in real-time, which is necessary to identify unscanned regions during the acquisition stage. Table 1 shows the percentage of pairs successfully aligned, the average RMSE in millimeters of the correct pre-alignments, and the average runtime in seconds for each tested approach.

Table 1. Comparison between pre-alignment approaches.

Approach	Success	RMSE	Time
Scanning sequence	23%	45.76	0.795
Principal axes	29%	25.62	**0.007**
Coplanar points	34%	21.42	1.169
Proposed approach	**74%**	**7.97**	1.608

As shown, the proposed approach outperforms the other three approaches in success rate and pre-alignment accuracy. Despite it presents the highest average runtime, it is still acceptable for real-time applications.

3.2. Calibration results

The proposed calibration approach was able to project I_ss into their respective I_cs in 98.6% of the tested views. The average runtime is about three minutes, which is acceptable since this stage is performed off line and has no influence on the acquisition stage. Calibration failures are mainly due to the lack of correct correspondences, but these failures are also automatically detected by analyzing the number of k_s correctly projected into the I_c. To this end, a calibration result is only considered correct if this number is above a pre-defined threshold (*i.e.* 10 is a safe threshold). Thus, no visual inspection is required to find incorrect calibrations, which is an advantage in comparison to works based on mutual information that have no control over the final result.

Next, we manually marked an average of 27 point correspondences in each view of the sculpture named Joel to be used as a ground truth. First, we used these correspondences to compute a reference calibration for each view. The average reprojection error of the ground truth correspondences using the reference calibration was about 4.87 pixels. Then, we used these correspondences to compute the reprojection error using the calibration obtained by the proposed approach, and we obtained an average error of 7.38 pixels. The difference of 2.5 pixels between manual and automatic approaches does not expressively affect the quality of the final object texture since the resolution of the I_cs is always higher than the resolution of the final texture.

4. CONCLUSION

We have presented our approaches for automatic pre-alignment of overlapped I_ss and automatic calibration of scanner and I_cs using SURF correspondences, which were developed to automate our 3D reconstruction pipeline.

The proposed approaches were evaluated using images acquired from five large sculptures. The manual work required in the original reconstruction pipeline was reduced in more than 95% of the cases. Furthermore, both approaches are able to automatically identify incorrect results thus eliminating the need for visual inspections to perform this task. The performance of the proposed approaches was mainly affected by the lack of proper keypoints, insufficient overlap or incorrect correspondences. In order to improve performance we can combine others features detector approaches to find new correspondences when SURF alone fails.

We automatically detect and pre-align overlapped views in 95% of tested images. The proposed approach outperformed other state-of-the-art works in the literature in both efficiency and accuracy and satisfies real-time requirements. This approach was encapsulated in our reconstruction tool to show all scanned views in a same coordinate system and allows monitoring the acquisition progress.

Our calibration approach correctly projected I_ss to their respective I_cs in more than 98% of the tests. Its accuracy is comparable to manual calibration and the final result can be automatically classified as correct or incorrect, unlike mutual information-based approaches.

Finally, the proposed approaches were successfully applied to the reconstruction of important Baroque sculptures from *"O Aleijadinho"* in a UNESCO World Heritage site, contributing to the preservation of Brazilian cultural heritage.

5. REFERENCES

[1] K. Ikeuchi, T. Oishi, J. Takamatsu, R. Sagawa, A. Nakazawa, R. Kurazume, K. Nishino, M. Kamakura, and Y. Okamoto, "The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects," *IJCV*, vol. 75, no. 1, pp. 189–208, 2007.

[2] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk, "The digital michelangelo project: 3D scanning of large statues," in *SIGGRAPH*, 2000, pp. 131–144.

[3] G. Pavlidis, A. Koutsoudis, F. Arnaoutoglou, V. Tsioukas, and C. Chamzas, "Methods for 3D digitization of cultural heritage," *J. Cultural Heritage*, vol. 8, no. 1, pp. 93–98, 2007.

[4] A. Vrubel, O. R. P. Bellon, and L. Silva, "A 3d reconstruction pipeline for digital preservation," in *CVPR*, 2009, pp. 2687–2694.

[5] B. T. Andrade, O. R. P. Bellon, L. Silva, and A. Vrubel, "Digital preservation of brazilian indigenous artworks: Generating high quality textures for 3d models," *J. Cultural Heritage*, 2011, vol. 12, pp. 05.002.

[6] J. Santos Jr., O.R.P. Bellon, L. Silva, and A. Vrubel, "Improving 3d reconstruction for digital art preservation," in *16th ICIAP*, 2011, vol. 6978, pp. 374–383.

[7] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *3DIM*, 2001, pp. 145–152.

[8] K. Pulli, "Multiview registration for large data sets," in *3DIM*, 1999, pp. 160–168.

[9] J. Davis, S. R. Marschner, M. Garr, and M. Levoy, "Filling holes in complex surfaces using volumetric diffusion," in *3DPVT*, 2002, pp. 428–861.

[10] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," in *SIGGRAPH*, 1987, pp. 163–169.

[11] B. T. Andrade, C. M. Mendes, J. O. Santos Junior, O. R. P. Bellon, and L. Silva, "3D preserving XVIII century barroque masterpiece: Challenges and results on the digital preservation of Aleijadinho's sculpture of the prophet Joel," *J. Cultural Heritage*, 2011, vol. 12, pp. 05.0003.

[12] C. S. Chua and R. Jarvis, "Point signatures: A new representation for 3d object recognition," *IJCV*, vol. 25, no. 1, pp. 63–85, 1997.

[13] N. M. Alpert, J. F. Bradshaw, D. Kennedy, and J. A. Correia, "The principal axes transformation–a method for image registration," *J. Nuclear Medicine*, vol. 31, no. 10, pp. 1717–1722, 1990.

[14] P. Pingi, A. Fasano, P. Cignoni, C. Montani, and R. Scopigno, "Exploiting the scanning sequence for automatic registration of large sets of range maps," *Computer Graphics Forum*, vol. 24, no. 3, pp. 517–526, 2005.

[15] D. Aiger, N. J. Mitra, and D. Cohen-Or, "4-points congruent sets for robust surface registration," *ACM Trans. Graphics*, vol. 27, no. 3, pp. 1–10, 2008.

[16] T. Franken, M. Dellepiane, F. Ganovelli, P. Cignoni, C. Montani, and R. Scopigno, "Minimizing user intervention in registering 2d images to 3d models," *The Visual Computer*, vol. 21, no. 8, pp. 619–628, 2005.

[17] R. K. Ko, K. Nishino, Z. Zhang, and K. Ikeuchi, "Simultaneous 2d images and 3d geometric model registration for texture mapping utilizing reflectance attribute," in *ACCV*, 2002, pp. 99–106.

[18] Y. Iwashita, K. Ryo, T. Hasegawa, and K. Hara, "Fast alignment of 3d geometrical models and 2d color images using 2d distance maps," in *3DIM*, 2005, pp. 164–171.

[19] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *IJCV*, vol. 24, no. 2, pp. 137–154, 1997.

[20] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *CVIU*, vol. 110, no. 3, pp. 346–359, 2008.

[21] F. Bonarrigo, A. Signoroni, and R. Leonardi, "A robust pipeline for rapid feature-based pre-alignment of dense range scans," in *ICCV*, 2011, vol. 24, pp. 2260–2267.

[22] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.