



HARVEST4D

HARVESTING DYNAMIC 3D WORLDS FROM COMMODITY SENSOR CLOUDS

Publications for Task 4.2

Deliverable 4.21

Date: 2015-07-08
Grant Agreement number: EU 323567
Project acronym: HARVEST4D
Project title: Harvesting Dynamic 3D Worlds from Commodity Sensor Clouds

Document Information

Deliverable number	D4.21
Deliverable name	Publications for Task 4.2
Version	1.0
Date	2015-07-08
WP Number	4
Lead Beneficiary	TUD
Nature	R
Dissemination level	PU
Status	Final
Author(s)	TUD

Revision History

Rev.	Date	Author	Org.	Description
0.1	2015-05-12	Tobias Plötz	TUD	First draft
1.0	2015-07-08	Tobias Plötz	TUD	Final version

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

1	Executive Summary	1
1.1	Introduction	1
1.2	Publications	1
2	Description of Publications	2
2.1	Overview	2
2.2	Fast and simple automatic alignment of large sets of range maps	2
2.3	Registering images to untextured geometry using average shading gradients.....	3
2.4	Global refinement of image-to-geometry registration for color projection.....	4
2.5	Other Results.....	5
3	References.....	5
4	Appendix	6

1 EXECUTIVE SUMMARY

1.1 INTRODUCTION

This deliverable describes the publications that resulted from Task 4.2, and how they fit into the work plan of the project.

The objective of Task 4.2 is to develop algorithms for bringing data coming from different modalities into a single coherent coordinate frame. This entails the fusion of multiple range maps to create a 3D model of the world as well as registering images to the model, enabling information transfer between the 2D images and 3D geometry. For example the registered images can be used to infer material properties of the real surface using the techniques developed in Work Package 7.

So far there are three publications that are mainly attributable to Task 4.2, two of which can be found in the appendix of this deliverable and one in the restricted section of the webpage. Additionally, there are two publications which are related to this deliverable. These are only mentioned but not discussed here since they stronger belong to other deliverables. The latter, related papers are available on the Harvest4D webpage or in other deliverables.

1.2 PUBLICATIONS

The following publications can be found in the appendix:

- Matteo Dellepiane, Roberto Scopigno
Global refinement of image-to-geometry registration for color projection.
Digital Heritage 2013 Int. Conf. Proc., pp 39-46, Nov. 2013
- Paolo Pingi, Massimiliano Corsini, Fabio Ganovelli, Roberto Scopigno
Fast and Simple Automatic Alignment of Large Sets of Range Maps.
Elsevier Computer & Graphics, Volume 47, April 2015, Pages 78–88

At the time of delivery of this deliverable, the following working paper still contains original unpublished work. Therefore, it can only be accessed through the restricted section of the webpage (for papers under submission, conditionally accepted papers, etc.):

- Tobias Plötz, Stefan Roth
Registering Images to Untextured Geometry using Average Shading Gradients.
submitted to ICCV 2015

The following related publications can be found on the webpage or in other deliverables:

- Murat Arikan, Reinhold Preiner, Claus Scheiblauer, Stefan Jeschke, Michael Wimmer
Large-Scale Point Cloud Visualization through Localized Textured Surface Reconstruction.
IEEE Transactions on Visualization & Computer Graphics, 2014.
- Murat Arikan, Reinhold Preiner, Michael Wimmer
Multi-Depth-Map Raytracing for Efficient Large-Scene Reconstruction.
IEEE Transactions on Visualization & Computer Graphics, 2015.

2 DESCRIPTION OF PUBLICATIONS

2.1 OVERVIEW

The main objective of Task 4.2 is to develop registrations algorithms that enable the alignment of data coming from different modalities to a single 3D model of the world. Once the data, say images, is registered it could be exploited in a multitude of application scenarios ranging from texturing the 3D geometry to temporal change detection or refinement of the existing geometry.

In order to have a common coordinate system within which we can align the incoming data, we need to build a 3D model of a scene. One way to do that is by fusing multiple range scans, each showing a different part of the scene, into a single point cloud. This registration task is tackled in the paper “Fast and simple automatic alignment of large sets of range maps” [Pigni et al. 2015]. The resulting point cloud often lacks texture information. Hence it is beneficial to align colored images with the geometry to infer texture information. The second paper of this task, “Registering images to untextured geometry using average shading gradients” [Plötz & Roth, 2015], proposes a solution to this challenging cross-model registration problem, building on the insights collected from the earlier work in Task 4.1 [Plötz et al. 2014]. The proposed fully-automated algorithm produces registration hypotheses that may still lack the accuracy to transfer high-quality texture information from the image to the 3D model. The hypotheses, however, can be used as initialization for the algorithm presented in the third paper, “Global refinement of image-to-geometry registration for color projection” [Dellepiane & Scopigno, 2013], that refines the alignment of multiple images for the purpose of texturing the 3D model. Together, the results of this task enable the integration of images and range scans into a single model of the world in a fully automated fashion – hence catering to the application scenarios of the Harvest4D project.

2.2 FAST AND SIMPLE AUTOMATIC ALIGNMENT OF LARGE SETS OF RANGE MAPS

In Harvest4D, incoming data should be registered to a common, potentially time-varying model of the 3D world. Hence, it is necessary to build this 3D model in the first place. 3D scans are a popular approach to acquire 3D geometry information. However, for larger scanning campaigns several range scans covering parts of the scene are necessary and manually aligning them

becomes intractable. This paper proposes a method to register large sets of range scans in an automated fashion within a few minutes and thus enables the creation of 3D reference models to which newly arriving data from other modalities can be aligned later on.

The algorithm iteratively selects a pair of range maps that get aligned to each other, exploiting a compact and GPU-friendly descriptor specifically designed for the alignment of range scans. The pairs are selected based on their maximal pixel-wise correlation considering all possible translations of the range maps, which can be efficiently computed using the Fast Fourier Transform. This procedure is repeated until the pairs of aligned scans form a spanning tree and hence for each pair of range scans the relative translation and rotation is known. To avoid false registrations the paper proposes an efficient method of validating the alignment of two range scans. Finally, the global registration can be refined using ICP to get a high-quality point cloud. An exemplary result is shown in Figure 1.

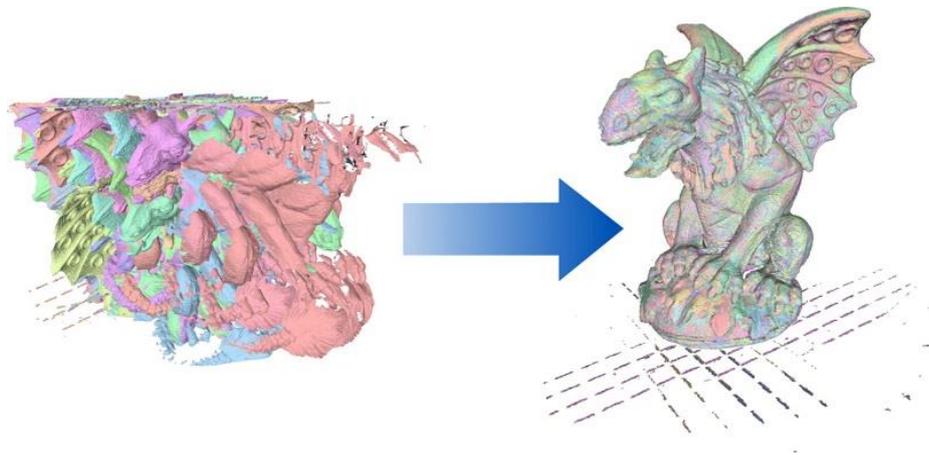


Figure 1: Range maps before the alignment (left) and registration results obtained after the computation with the proposed methods (right).

2.3 REGISTERING IMAGES TO UNTEXTURED GEOMETRY USING AVERAGE SHADING GRADIENTS

The core of this task is to use a 3D model of the world, e.g. acquired through the method described above, and integrate information coming from different sources of data. Before information transfer is possible, the incoming data needs to be aligned to the 3D model. This paper deals with the challenging multimodal registration problem of aligning images to an untextured 3D model. While most existing methods require human interaction in order to get a careful initialization for the registration process, the paper proposes a fully automated algorithm as user interaction is not feasible for large streams of incoming data, which Harvest4D is aiming at.

Registering images and untextured 3D geometry is challenging for two reasons. First, the two modalities live in spaces of different dimensionality – 2D vs. 3D – and second, the image shows the 3D model with its real texture as well as under unknown lighting conditions. The paper deals with the first issue by representing the 3D world with a set of renderings. From these renderings



newly developed average shading gradients are extracted that correspond well to gradients on the image due to shading changes, thus addressing the second issue. The missing texture information on the 3D model calls for a robust registration process. Hence, first coarse registration hypotheses are found that are refined and verified later on. The algorithm is able to register images showing 3D models of varying scales without the need of any user input, and also is able to indicate when a registration could not be performed successfully. Figure 2 shows some example results of the registration process.



Figure 2: Example results of the fully automated image-to-geometry registration. The input image is shown on the left and the best scoring verified hypothesis on the right, respectively.

2.4 GLOBAL REFINEMENT OF IMAGE-TO-GEOMETRY REGISTRATION FOR COLOR PROJECTION

The method described above can be used to register images to an untextured model of the 3D world. Having done so, a natural example of cross-modal data integration in the context of Harvest4D is to use the algorithms developed in work package 7 to transfer color information or other material properties from the image to the model. However for color transfer, the initial registration lacks the desired accuracy and hence the third paper in this task is concerned with refining an existing registration of a set of images in order to colorize the model. A core part of the method itself is the cross-modal fusion of geometric and color information into a single data representation. As a result of the algorithm, the quality of the color projection is improved, especially when dealing with small details, as can be seen in Figure 3.

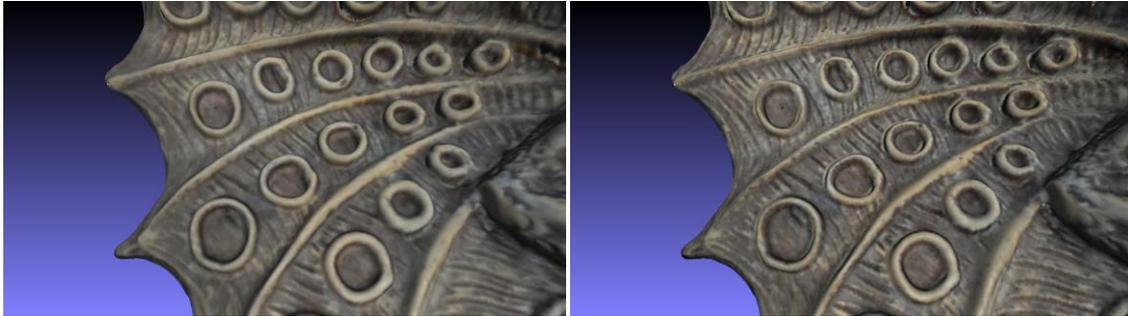


Figure 3: Color project before (left) and after (right) the global refinement.

Beginning from an initial estimate of the camera pose for each image in the input set, the proposed method iteratively selects an image and refines its registration by maximizing the Mutual Information – a statistical measure of dependency – between the image and a rendering of the 3D model. The rendering fuses both color information from the other images as well as geometric information from the model to achieve maximal accuracy of the refined registration. The refinement is repeated for all nodes.

2.5 OTHER RESULTS

Beyond these three papers, two papers [Arikan et al. 2014, Arikan et al. 2015] mainly attributed to task 8.1 have a connection to this deliverable. While the papers in this deliverable deal with the registration of individual images to a given geometry, they do not solve the problem of which image to select in areas where multiple images overlap. This is handled by the two mentioned publications in 8.1: the methods presented there assume already registered images, and solve an optimization problem to select for each pixel on screen the image leading to the highest overall image quality.

3 REFERENCES

- Matteo Dellepiane, Roberto Scopigno
Global refinement of image-to-geometry registration for color projection.
 Digital Heritage 2013 Int. Conf. Proc., pp 39-46, Nov. 2013
- Paolo Pingi, Massimiliano Corsini, Fabio Ganovelli, Roberto Scopigno
Fast and Simple Automatic Alignment of Large Sets of Range Maps.
 Elsevier Computer & Graphics, Volume 47, April 2015, Pages 78–88
- Tobias Plötz, Thorsten Franzel, Stefan Roth
Towards edge-based representations for image-to-geometry registration
 Working Paper, TU Darmstadt, 2014
- Tobias Plötz, Stefan Roth
Registering Images to Untextured Geometry using Average Shading Gradients.
 submitted to ICCV 2015

4 APPENDIX

The following pages contain all the publications that are directly associated with this deliverable. Other publications referenced in this deliverable can be found in the public Harvest4D webpage (for already published papers), or in the restricted section of the webpage (for papers under submission, conditionally accepted papers, etc.).

Global refinement of image-to-geometry registration for color projection

Matteo Dellepiane and Roberto Scopigno
Visual Computing Laboratory ISTI - CNR
Via G. Moruzzi 1, 56124 Pisa (PI), Italy
Email: *surname@isti.cnr.it*

Abstract—The management, processing and visualization of color information is a critical subject in the context of the acquisition and visualization of real objects. Especially in the context of Cultural Heritage, artifacts are so complex or hard-to-handle that the appearance information has to be extracted from a set of images.

The images usually have to be registered to the 3D model of the objects, in order to transfer the needed information. Hence, the problem of image-to-geometry registration has been thoroughly studied by the Computer Graphics and Computer Vision community. Several methods have been proposed, but a fully automatic and generic solution is still missing. Moreover, small misalignments often lead to visible artifacts in the final colored 3D models.

In this paper, we propose a method to refine the alignment of a group of images which has been already registered to a 3D model. Taking advantage of the overlapping among the images, and applying a statistical global method based on Mutual Information, the registration error is distributed among all the elements of the dataset. Hence, the quality of color projection is improved, especially when dealing with small details.

The method was tested on a number of heterogeneous Cultural Heritage objects, bringing to a visible improvement in the rendering quality. The method is fully automatic, and it does not need powerful hardware or long processing time. Hence, it represents a valid solution for a wide application on CH artifacts.

I. INTRODUCTION

Accurate registration of images on geometry is an important kernel operation among the technologies for Cultural Heritage. This is related to a number of different applications, and is often needed due to the necessity to integrate data coming from different acquisition devices.

The main application is related to the projection and visualization of color information on a 3D model, but registered images are important also in the fields of material properties estimation, referenced images navigation, annotation for monitoring or restoration.

One of the main problems is how to cope with the small misalignments remaining after the registration process. These errors can be due both to the methods used, and to the quality of the initial datasets, and they can result in annoying artifacts.

In this paper we proposed a refinement technique which aims at removing the above cited errors, by applying a global approach where the camera parameters are slightly modified in order to obtain a perfect projection. The method starts from a set of images aligned to a 3D model. It uses Mutual Information to refine each image by taking into account

the corresponding projection of all the other images on the model. The camera parameters are refined until convergence, so that the alignment error is distributed among the images, and globally minimized. One of the advantages of the method (which is totally automatic) is that no assumption about the quality of the initial dataset is made: the method will try to obtain the best projection regardless of the quality of both 2D and 3D data. The method was tested on a number of artifacts, showing that the quality of the color projection is greatly improved, especially when dealing with fine decorative details.

II. RELATED WORK

The problem of image-to-geometry registration has been thoroughly studied, and a number of different solutions have been proposed. In the following we tried to divide the methods in groups. We start from the assumption that the images have been taken in a different moment w.r.t. the geometry acquisition, hence we do not take into account methods based on co-located cameras [1], [2], [3].

Semi-automatic methods. A very robust approach is based on setting several 2D-3D correspondences: the correspondences are then used to estimate the camera parameters using a minimization algorithm. Although new procedures to speed up the process were proposed [4], the approach can be very time-consuming, especially when tens or hundreds of images need to be aligned. Automatic planning of the images required could minimize image acquisition and remove the need for registration [5], but this approach can only be used in controlled environments.

Features and Silhouette-based methods. The geometric features of the object can be used to find the registration of images. Features can be points, lines, rectangles [6], edge intensities [7], or the silhouette of the object [8], [9], [10]. However, these methods rely on the presence of these features on the object, hence they could be best applied on peculiar types of artifacts (e.g. architectural scenes, if using clusters of orthogonal lines).

Color-based methods. Another feature that can be used (if present) is the reflectance value (laser intensity) or color information that some 3D scanners acquire. This helps the feature extraction and the establishment of correspondences [11]. Yang et al. [12] made their co-located camera approach robust using this method. Wu et al. [13] exploited color information to align two 3D scenes even from significant viewpoint changes.

Statistical methods. Other approaches try to catch the non-linear correlations between the image and the geometric properties of the target surface. A measure which is extensively used in medical imaging (see [14] for a survey) is called Mutual Information (MI). It was pioneered by Viola and Wells [15] and by Maes et al. [16]. Viola and Wells [15] suggested to compare the gradient variations of the image and a rendering of the 3D models showing the surface normals. Corsini et al. [17] extended this algorithm by including other geometric properties, such as ambient occlusion and reflection directions, in the alignment algorithm. Cleju et al. [18] also extended Viola and Wells’s work to align more than one image simultaneously. We propose a similar approach to refine the global registration, based on a different optimization framework.

Multi-view methods. Almost all the above mentioned methods are based on the alignment of a single image on the geometry. A more recent group of works rely on the fact that if a group of images has to be aligned on a model, it is possible to take advantage also on the relations between images. Several works exploit Structure From Motion (SFM) during 2D/3D registration process, like those of Zhao et al. [19], Stamos et al. [20] (which is an extension of the work of Liu et al. [21]), Zheng et al. [22], Pintus et al. [23], and Corsini [24].

Refinement of existing registrations. Our method aims at refining an initial registration of a group of images. Hence the results of any of the above methods could be a starting point for the refinement. Other methods aimed at improving the color projection by modifying the initial image set, using for example Optical Flow methods [25], [26]. These methods can obtain extremely accurate results, but they are limited by the resolution of the images, or by the need of advanced hardware and very long processing time. Our method aims at improving the camera parameters without modifying the rest of the dataset. We make use of Mutual Information, but on a more global and interconnected way w.r.t. the usual statistical methods.

III. GLOBAL REFINEMENT USING MUTUAL INFORMATION

The input of our method is composed by:

- A 3D model
- A set of images
- A set of camera parameters associated to each image of the dataset. In our case, each camera is defined by seven parameters: three for camera position, three for camera orientation, and one for the focal length (see next Section for details).

The assumption is that an initial registration of the images on the 3D model is already provided. The registration can be obtained using any of the above mentioned methods, or it could come from external systems (like multi-view stereo reconstruction tools).

The goal of the method is to modify the camera parameters of all the images so that the global Mutual Information will be maximized. This means that each of the images will have the maximum Mutual Information Value w.r.t. all the other images projected on the 3D model. In order to do this, we will treat the system of registered images as a graph, and

we will try to distribute the alignment error in the graph by improving all its nodes.

This approach has several points in common with the one by Pulli [27], which aimed at the global refinement of groups of range maps. It obtained this by treating the system of range maps as a graph, and improving the alignment of a range map at a time, trying to distribute the alignment error in a balanced way.

A. Using Mutual Information to align a single image

Mutual Information (MI) measures the information shared by two random variables A and B . The Mutual Information \mathcal{MI} between two images I_A and I_B can be defined as:

$$\mathcal{MI}(I_A, I_B) = \sum_{(a,b)} p(a,b) \log \left(\frac{p(a,b)}{p(a)p(b)} \right) \quad (1)$$

where $p(a)$ ($p(b)$) is the probability that the value of the pixel I_A (I_B) is a (b) and $p(a,b)$ is the joint probability of the event (a,b) . The joint probability distribution can be easily estimated by evaluating the joint histogram of the two images and then dividing the number of occurrences of each entry by the total number of pixels. A joint histogram is a bi-dimensional histogram made up of $N \times N$ bins; the occurrence (a,b) is associated with the bin (i,j) where $i = \lfloor a/m \rfloor$ and $j = \lfloor b/m \rfloor$ and m is the width of the bin. This value can be seen as an expression of the *nonlinear correlation* between the variables A and B .

The image-to-geometry registration problem in this case can be defined as the estimation of the camera parameters that produce a rendering I_B of the 3D model that maximizes MI with respect to the image to align (I_A). The generation of the rendering is the main issue to be solved, since generally there’s a lack of knowledge of not only the color and materials of the object but also of the lighting conditions. At the same time, since the Mutual Information expresses a correlation between the images, the photorealism is not a requisite: it is important that the rendering contains a high amount of information ”in common” with (I_A). Corsini et al. [17] proposed several rendering types following this aim. In particular, they showed that ambient occlusion correlates well since the occluded parts of the geometry often correspond with the dark parts in the real image due to the poor illumination arriving at these points, and normal maps are strongly correlated with more directional illumination.

In this context the registration can be formalized as an optimization problem in a 7D space:

$$\begin{aligned} \mathcal{C}^* &= \arg \max_{\mathcal{C} \in \mathbb{R}^7} \mathcal{MI}(I_A, I_B(\mathcal{C})) \\ \mathcal{C} &= (t_x, t_y, t_z, \theta_x, \theta_y, \theta_z, f) \end{aligned} \quad (2)$$

where f is the focal length, (t_x, t_y, t_z) and $(\theta_x, \theta_y, \theta_z)$ define the position and orientation of the camera, I_A is the image to align and I_B is the rendering of the 3D model. Obviously, I_B depends on the camera parameters (\mathcal{C}). The Equation (2) can be solved by a non-linear optimization algorithm such as NEWUOA [28].

B. Extension to groups of images

The proposed method aims at extending the single image registration problem to a group of images. Instead of using a pure rendering of the 3D model, the goal is to take advantage of the overlaps among the projections of the images on the surface of the model. In this way, the best alignment will be reached when all the images will project the same color details on the same part of the surface.



Fig. 1. An example of the images used to calculate the value of an arc. Top, the original image. Bottom: a rendering of another image projected on the top image plane. The parts which are not covered by the image are shown using a "combined" (normal map + ambient occlusion) rendering.

1) *Representing the images as a graph*: In order to handle the connections among all the elements of the registration project, it's necessary to encode them in a structure. Hence, the registration project is represented as a graph. The nodes of the graph are represented by each image of the dataset. The nodes are connected through arcs, and each arc is associated to a weight. The value of the weight between an image I_1 and another image I_2 , indicated with $w(I_1, I_2)$, corresponds to

$$w(I_1, I_2) = \mathcal{MI}(I_1, \text{proj}(I_2, I_1))OV(I_2, I_1) \quad (3)$$

where the first term is the MI calculated between the image I_1 and the projection of the image I_2 on the image plane of I_1 . The projection is generated by projecting I_2 on the 3D model, and then generating a rendering from the point of view of I_1 . The parts of the 3D model which are not covered by I_2 are represented using the *combined rendering* (ambient occlusion + normals map) proposed by Corsini et al. [17]. Figure 1 shows an example of a couple of images used for the calculation of the arc weight.

The value of the arc is also weighted by OV , which represents the amount of overlap between the images. This is the ratio

between the pixels on I_1 image plane which is covered by I_2 , and the total number of pixels covered by the 3D model. This term aims at giving a more important role to the images which share a bigger common projection surface. At the same time, the arc between two nodes is created only if the value of OV is bigger than 0.2: this simplifies the graph structure and prevents artifacts coming from images which share very small common projection surfaces.

According to this definition, the graph related to each dataset analyzes each image, and creates an arc for each couple of images where there is enough overlap. The result of the building phase is a weighted directed graph, since $w(I_1, I_2)$ is usually different from $w(I_2, I_1)$. It is interesting to note that the 3D model is not directly represented in the graph, but it plays the role of a "medium" due to the projections involved.

2) *Graph-based registration refinement*: The refinement of the graph is obtained following a similar procedure to the one used by Pulli [27] for the alignment of range maps. In that method, the refinement is reached by considering one node at a time, and using Hausdorff distance as the value to minimize. In our case, we used Mutual Information as a value to be maximized.

The refinement loop follows these steps:

- **Node selection**: among the nodes which were not already refined, the one with more connections with already refined nodes is chosen. When more than one node has the same number of connections with refined nodes, the one with the biggest largest number of entering arcs is chosen.
- **Node refinement**: the refinement is obtained by maximizing the MI (Equation 2) between the image associated to the node and a rendering of the 3D model where all the images associated with connected nodes are projected on the geometry. Figure 2 shows an example of a rendering used for the refinement. Since several images can project onto the same portion of geometry, the color assigned to the pixels is a combination of the contributions of all the images. The contribution is weighted by the value of the arc connecting the node image and each image which projects on the model. This approach aims at having the other images of the dataset as a guide for the chosen node to converge to a common alignment. If portions of the geometry are not covered by any other image in the set, *combined rendering* is used.
- **Node update**: when the maximization procedure ends, the node is labeled as *refined*, and the graph is updated (all the weights of the arcs involving the node are recalculated). The procedure goes back to step 1, until all the nodes are refined.

When all the nodes have been refined, it is possible to start the refinement again to further improve the global alignment. The procedure should be able to converge until all the camera parameters associated to the images are not modified anymore. See Figure 3 for two examples of the improvements of the registration of images.

Comparing two cameras is not trivial: one of the most reliable methods is to compare the projection on the image planes

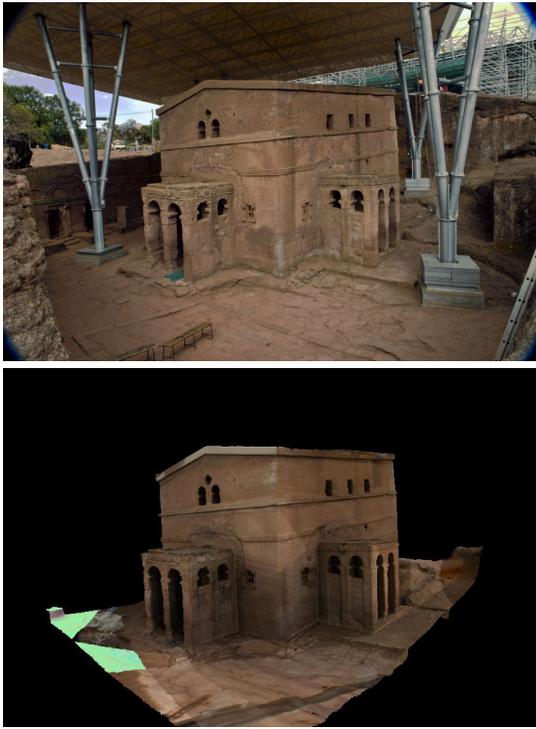


Fig. 2. An example of the renderings used for the refinement of camera parameters. Top: the image associated to the node to refine. Bottom: a rendering obtained by projecting all the other images on the 3D model. The small portions not covered by images are rendered using the "combined" rendering.

of several samples. For this reason, in order to measure the variation in the camera system after a refinement step, a group of N 3D sample points $X = x_1, x_2, \dots, x_N$ is extracted from the 3D model. The average variation of cameras is calculated as:

$$Var = \frac{\sum_{i=1}^N \sum_{j=1}^M |pro(x_i, C_j^{After}) - pro(x_i, C_j^{Before})|}{MN} \quad (4)$$

where $pro(x_i, C_j^{After})$ is the projection of the point x_i on the image plane of C_j^{After} , the camera associated to image I_j . The average of the variation in pixel of the projection of the sample points on the image planes before and after the refinement gives a reasonable estimation of the amount of perturbation applied.

Hence, at the end of the refinement step, the Var value is calculated. If it is below a pre-defined threshold, or the maximum number of iteration has been reached, the refinement stops. In the examples shown in the Results Section, the number of samples used was 5000, the Var threshold was set to 1.2 pixels, and the maximum number of iterations was set at 5.

3) *Selective refinement of nodes*: One of the limitations of the statistical approach is that the Mutual Information is a pure number. Hence, it is not possible to compare its value between couples of images. Hence, one of the limitations of the proposed system is that an image which exhibits a misalignment could "guide" all the others to a sharp, but uncorrect color projection (see Figure 4).



Fig. 3. Two examples of the refinement of single images after the application of our method, showing an image with the 3D model in transparency. In the first one (first and second row) small details are better aligned. In the second one (third and fourth row) the silhouette is better matched.

In order to cope with this potential problem, the proposed system gives the possibility to the user to indicate some *anchor nodes*, which are associated to images which already have an accurate registration. In this case, the nodes will always be considered as *refined*, and their role will be to "guide" the other images to a more accurate registration. This modality gives the possibility to handle the datasets where only a few images are misaligned: the time needed for the refinement process will be much smaller, and the user will have control on the registration procedure.

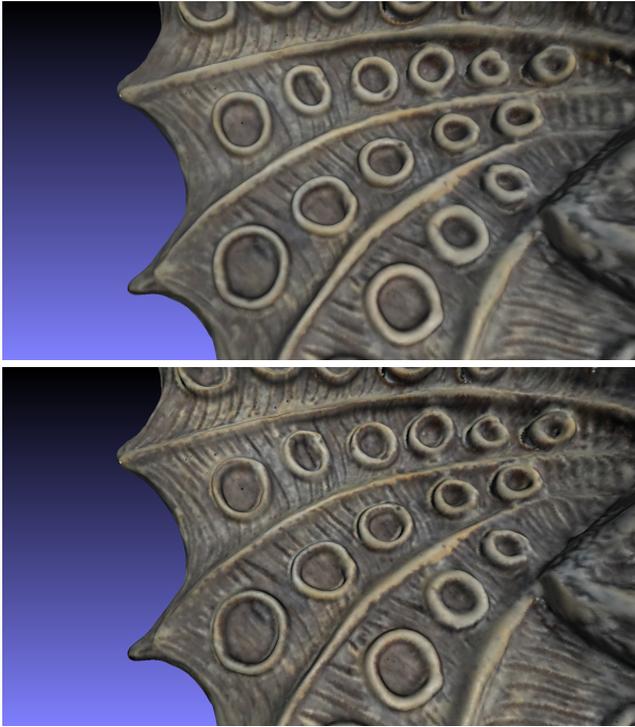


Fig. 4. An example of a possible limitation of the approach. Top: a portion of the model before refinement. Bottom: the same portion after the refinement. Although the color is sharper, the alignment w.r.t. the geometric features (e.g. the circles) is less accurate. This is because a misaligned image "guided" the others to a misaligned common position.

IV. RESULTS

The proposed system was tested on a number of real cases covering a wide range of possible objects, from small artifacts to architectures, and a variety of datasets, from a few to tens of images, with varying quality of both 2D and 3D data. All the objects are examples of Cultural Heritage, and they were acquired using 3D Scanners. Photographic datasets were mostly acquired with the purpose of color projection.

The initial registration of the images was obtained using a single image registration method, the one based on MI, proposed by Corsini [17] and implemented in MeshLab [29]. In the case of *Formella* and *Maryam Church* datasets, the registration was obtained using an evolution of Corsini et al, *Mutual Correspondences* [30], which gives the possibility to the user to guide the registration process with a simple interaction. All the processing was performed on an Intel Core i7 CPU, with 24 GB RAM and an NVidia GeForce GTX 560 Ti.

The datasets are described in Table 5. The Table describes the physical size of the object, the complexity of the 3D model, the amount and resolution of the images, the processing time needed for the convergence of global refinement. From a general point of view, the method proved to be applicable regardless of the physical size of the object and of the complexity of the dataset. The time needed for the refinement is partially dependent on the size of the 3D model, and on the number and resolution of images, but the initial misalignment and the amount of overlap among the images play a critical role. Anyway, the processing time is reasonable

even in the case of very big datasets (hundreds of images), also considering that the refinement step is an *una tantum* operation, needed just before the color projection.

Figure 6 shows the results of the registration refinement. The color was transferred from the images to the 3D Model using the Masked Photo Blending approach [31], which is a method which blends the contribution of all the images during color projection. This method proved to be very robust, being able to deal also with very complex datasets. Nevertheless, one of the limitations is that small misalignments could introduce *aliasing* (or *ghosting*) effects on the colored model. This is due to the fact that color details are projected in slightly different positions on the surface of the 3D model. Figure 6 shows a rendering of the colored 3D models obtained with a color projection applied before and after the global refinement. We can observe that, in general, the quality of the color information is clearly improved. In the case of small and medium size objects (*Gargoyle* and *Formella*) the method is able to recover the fine details of the decoration of the objects.

In the case of bigger objects, the method recovers also quite big starting misalignments, regardless of the number of images taken into account. In the case of the *Maryam Church*, given the low number of images and the simple shape of the object, the color was encoded in a Texture (all the other objects are represented using color-per-vertex). Also in this case, there was an improvement of the fine color details. In the case of the *Cathedral*, the method proved to be able to handle a massive amount of images, that usually produce blurry color, due to the accumulation of misalignments. Also in this case, most of the color detail was recovered, although part of the finer elements was lost. This was due both to the quality of the images (average resolution) and to the accuracy of the 3D model, which was not high. Additional snapshots of the results are shown in Figure 7

The proposed method has some limitations. The main one is shared with all the statistic-based registration approaches: if the quality of the dataset is low, or the elements do not share enough information, the method will not be able to converge, and it could lead to the degeneration of the camera parameters estimation. This limitation is generally shared also with the other registration approaches, except some of the ones needing a strong user intervention.

As already mentioned in Section III-B3, the other limit comes from the fact that it is difficult to compare the quality of the registration of the single images. This can bring to a convergence of the refinement obtained by "following" an image which had a lower quality registration. This can be partially solved by applying a selective refinement.

Nevertheless, the proposed method is simple and completely automatic, and it does not need complex hardware and long processing time. It reduces the time that the user needs to spend in order to obtain very accurate image registration. Finally, it helps overcoming the limitations of most of the state-of-the-art color projection tools.

Object	Size (cm)	3D Model (MTri)	N. Images (Resolution)	Processing Time (sec.)
<i>Gargoyle</i>	12	3	11 (3872x2592)	191
<i>Formella</i>	65	5	10 (3872x2592)	127
<i>Neptune</i>	580	10	44 (1728x1152)	740
<i>MaryamChurch</i>	930	3 (with texture)	8 (3872x2592)	112
<i>Abside</i>	3500	9	310 (1936x1296)	3250

Fig. 5. Table of data for the five test cases.

V. CONCLUSION

In this paper, we presented a method for the refinement of image-to-geometry registration. The goal is to improve the quality of an already registered set of images, in order to solve eventual misalignments and improve color projection.

In order to achieve this, the set of images is treated as a graph, and the estimation of the camera parameters is calculated taking into account the projection of all the other images on each image plane. In order to refine the registration, a statistical method based on Mutual Information was implemented. The graph representing the images is refined node-by-node until convergence. The 3D model acts as a simple medium for color projection, because the final goal is to have a registration where all the images project the same color details on the same part of the geometry, regardless of its quality.

The method proved to be robust and reasonably fast. It was tested on a number of Cultural Heritage objects, covering different physical sizes and dataset complexities. All the tests showed an improvement in the color quality. This method proves to be extremely useful especially in the Cultural Heritage field, where most of the times the only way to obtain a basic color information of an object is to transfer it from a set of uncalibrated images.

The future improvements of the method include: the study of mechanisms to prevent the worsening of color quality, in the case of low quality datasets, and the implementation of simple interaction procedures to give the possibility to the user to guide the refinement process.

ACKNOWLEDGMENT

This research work was partly funded by the EU Community FP7 ICT *Harvest4D* project (Grant Agreement 323567). We'd like to thank the people from Zamani Project (<http://www.zamaniproject.org/>) for the Maryam Church data.

REFERENCES

- [1] K. Pulli, H. Abi-Rached, T. Duchamp, L. G. Shapiro, and W. Stuetzle, "Acquisition and visualization of colored 3D objects," in *Proc. of the 14th International Conference on Pattern Recognition (ICPR'98)*, vol. 1. IEEE Computer Society, 1998, pp. 11–15.
- [2] V. Sequeira and J. G. Goncalves, "3D reality modelling: Photo-realistic 3d models of real world scenes," *3D Data Processing Visualization and Transmission, International Symposium on*, pp. 776–783, 2002.
- [3] C. Fröh and A. Zakhor, "Constructing 3D city models by merging aerial and ground views," *IEEE Computer Graphics and Applications*, vol. 23, pp. 52–61, 2003.
- [4] 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. 8-10, pp. 619–628, sep 2005.
- [5] K. Matsushita and T. Kaneko, "Efficient and handy texture mapping on 3D surfaces," *Computer Graphics Forum*, vol. 18, no. 3, pp. 349–358, 1999.
- [6] L. Liu and I. Stamos, "Automatic 3D to 2D registration for the photorealistic rendering of urban scenes," in *CVPR*, vol. 2. IEEE Computer Society, 2005, pp. 137–143.
- [7] P. J. Neugebauer and K. Klein, "Texturing 3D models of real world objects from multiple unregistered photographic views," *Computer Graphics Forum*, vol. 18, no. 3, pp. 245–256, 1999.
- [8] D. G. Lowe, "Fitting parameterized three-dimensional models to images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 441–450, May 1991.
- [9] L. Brunie, S. Lavallée, and R. Szeliski, "Using force fields derived from 3D distance maps for inferring the attitude of a 3D rigid object," in *Proc. of the Second European Conference on Computer Vision (ECCV'92)*. Springer-Verlag, 1992, pp. 670–675.
- [10] H. P. A. Lensch, W. Heidrich, and H.-P. Seidel, "Automated texture registration and stitching for real world models," in *PG '00: Proceedings of the 8th Pacific Conference on Computer Graphics and Applications*. IEEE Computer Society, 2000, pp. 317–326.
- [11] K. Ikeuchi, A. Nakazawa, K. Hasegawa, and T. Ohishi, "The great buddha project: Modeling cultural heritage for vr systems through observation," in *Proc. of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'03)*. IEEE Computer Society, 2003, pp. 7–16.
- [12] G. Y. Gehua Yang, J. Becker, and C. V. Stewart, "Estimating the location of a camera with respect to a 3d model," in *Proc. of the Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM2007)*. IEEE Computer Society, 2007, pp. 159–166.
- [13] C. Wu, B. Clipp, X. Li, J.-M. Frahm, and M. Pollefeys, "3d model matching with viewpoint-invariant patches (vip)," in *CVPR'08*. IEEE Computer Society, 2008, pp. 1–8.
- [14] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
- [15] P. Viola and I. William M. Wells, "Alignment by maximization of mutual information," *Int. J. Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.
- [16] F. Maes, A. Collignon, D. Vandeermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions in Medical Imaging*, vol. 16, pp. 187–198, 1997.
- [17] M. Corsini, M. Dellepiane, F. Ponchio, and R. Scopigno, "Image-to-geometry registration: a mutual information method exploiting illumination-related geometric properties," *Computer Graphics Forum*, vol. 28, no. 7, pp. 1755–1764, 2009.
- [18] I. Cleju and D. Saupe, "Stochastic optimization of multiple texture registration using mutual information," in *Proceedings of the 29th DAGM conference on Pattern recognition*. Springer-Verlag, 2007, pp. 517–526.
- [19] W. Zhao, D. Nister, and S. Hsu, "Alignment of continuous video onto 3d point clouds," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1305–1318, Aug 2005.
- [20] I. Stamos, L. Liu, C. Chen, G. Wolberg, G. Yu, and S. Zokai, "Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes," *Int. J. Comput. Vision*, vol. 78, pp. 237–260, July 2008.
- [21] L. Liu, I. Stamos, G. Yu, G. Wolberg, and S. Zokai, "Multiview geometry for texture mapping 2D images onto 3D range data," in *Proc. of the 2006 IEEE Computer Society Conference on Computer Vision*

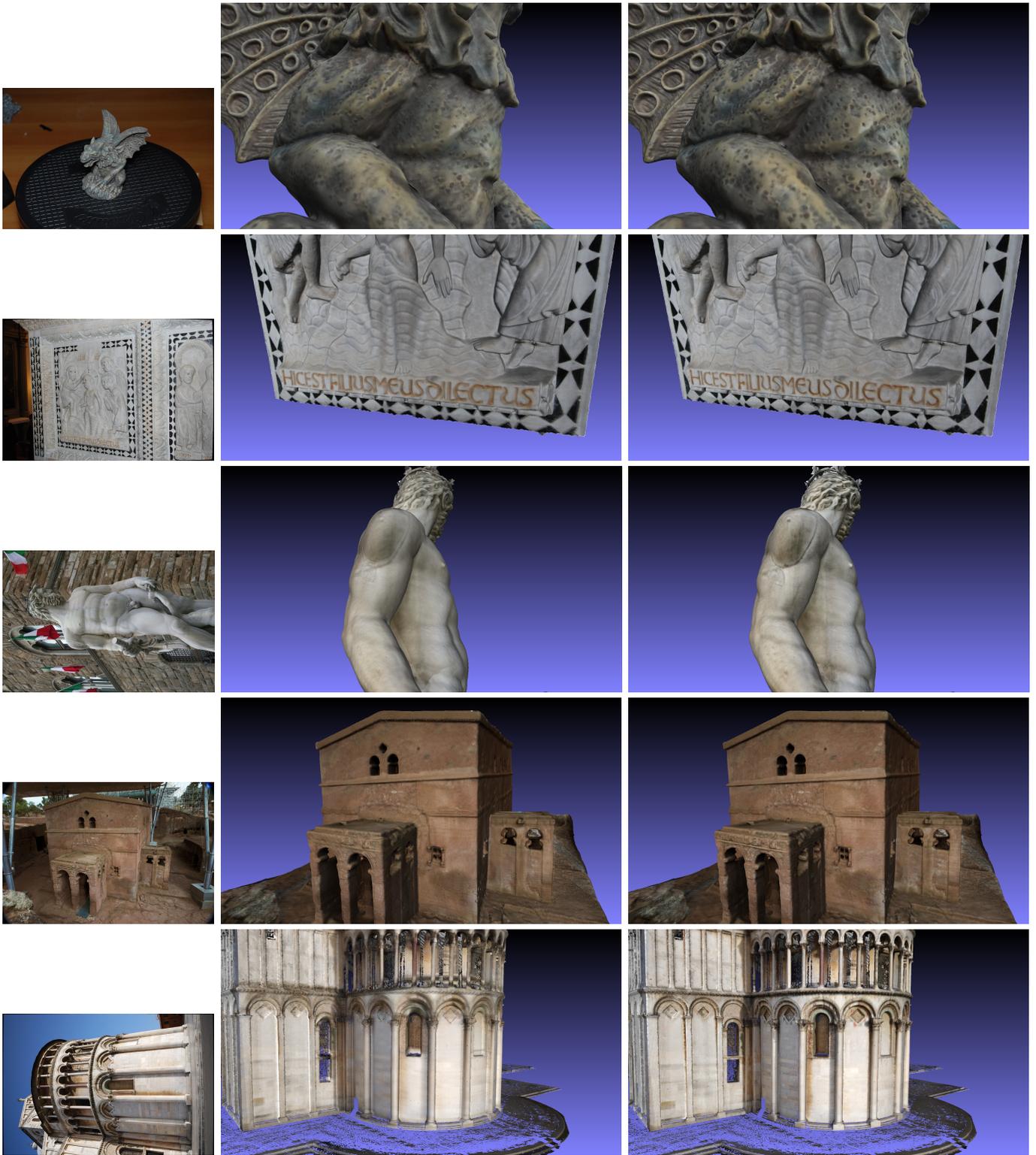


Fig. 6. The test cases. First column: a representative image of the ones used for color projection. Second column: color projection before refinement. Third column: color projection after refinement.

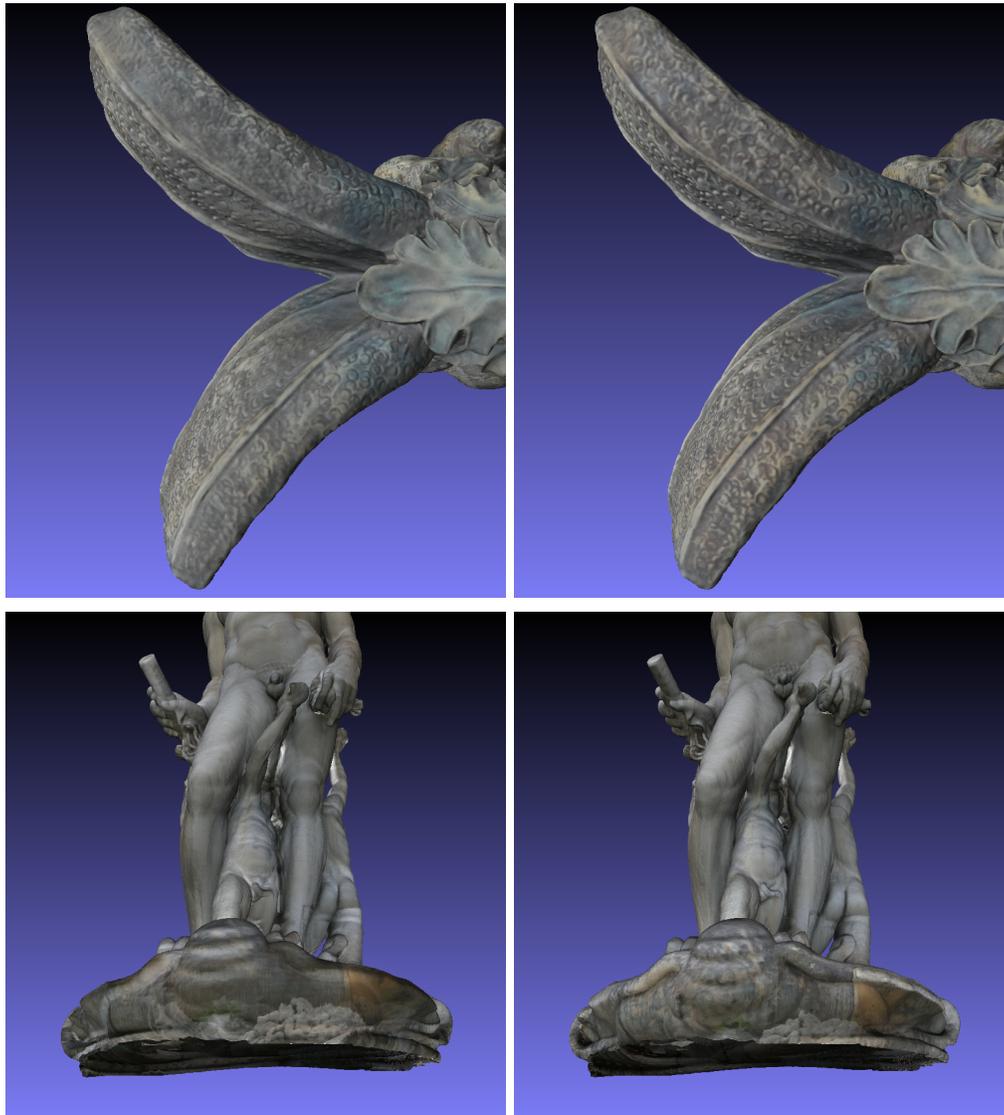


Fig. 7. Snapshots of the color projection before and after refinement. First two rows: a detail of the Gargoyle, before and after refinement. Last two rows: the bottom part of the Neptune, before and after refinement.

- and *Pattern Recognition (CVPR'06)*, vol. 2. IEEE Computer Society, 2006, pp. 2293–2300.
- [22] H. Zheng, I. Cleju, and D. Saupé, “Highly-automatic mi based multiple 2D/3D image registration using self-initialized geodesic feature correspondences,” in *ACCV (3)*, 2009, pp. 426–435.
- [23] R. Pintus, E. Gobbetti, and R. Combet, “Fast and robust semi-automatic registration of photographs to 3d geometry,” in *Proceedings of the 12th International conference on Virtual Reality, Archaeology and Cultural Heritage*, ser. VAST’11. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2011, pp. 9–16. [Online]. Available: <http://dx.doi.org/10.2312/VAST/VAST11/009-016>
- [24] M. Corsini, M. Dellepiane, F. Ganovelli, R. Gherardi, A. Fusiello, and R. Scopigno, “Fully automatic registration of image sets on approximate geometry,” *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 91–111, 2013. [Online]. Available: <http://dx.doi.org/10.1007/s11263-012-0552-5>
- [25] M. Eisemann, B. De Decker, M. Magnor, P. Bekaert, E. de Aguiar, N. Ahmed, C. Theobalt, and A. Sellent, “Floating textures,” *Computer Graphics Forum (Proc. of Eurographics)*, vol. 27, no. 2, pp. 409–418, Apr 2008.
- [26] M. Dellepiane, R. Marroquim, M. Callieri, P. Cignoni, and R. Scopigno, “Flow-based local optimization for image-to-geometry projection,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 3, pp. 463–474, March 2012.
- [27] K. Pulli, “Multiview registration for large data sets,” in *Proc. of the 2nd international Conference on 3-D digital imaging and modeling (3DIM'99)*. Washington, DC, USA: IEEE Computer Society, 1999, pp. 160–168.
- [28] M. J. D. Powell, “Developments of NEWUOA for minimization without derivatives,” *IMA Journal of Numerical Analysis*, vol. 28, no. 4, pp. 649–664, Oct. 2008.
- [29] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, “Meshlab: an open-source mesh processing tool,” in *Sixth Eurographics Italian Chapter Conference*. Eurographics, 2008, pp. 129–136.
- [30] M. Sottile, M. Dellepiane, P. Cignoni, and R. Scopigno, “Mutual correspondences: an hybrid method for image-to-geometry registration,” in *Eurographics Italian Chapter Conference 2010*. EG, Nov 2010, pp. 81–88.
- [31] M. Callieri, P. Cignoni, M. Corsini, and R. Scopigno, “Masked photo blending: mapping dense photographic dataset on high-resolution 3D models,” *Computer & Graphics*, vol. 32, no. 4, pp. 464–473, Aug 2008.



Technical section

Fast and simple automatic alignment of large sets of range maps[☆]Paolo Pingi^{*}, Massimiliano Corsini, Fabio Ganovelli, Roberto Scopigno

Visual Computing Laboratory, ISTI-CNR, Via G. Moruzzi 1, 56124 Pisa (PI), Italy

ARTICLE INFO

Article history:

Received 25 August 2014

Received in revised form

28 October 2014

Accepted 9 December 2014

Available online 20 December 2014

Keywords:

Range maps registration

3D scanning

Automatic registration

ABSTRACT

We present a very fast and simple-to-implement algorithm for the automatic registration of a large number of range maps. The proposed algorithm exploits a compact and GPU-friendly descriptor specifically designed for the alignment of this type of data. This pairwise registration algorithm, which also includes a simple mechanism to avoid to get false positives, is part of a system capable to align a sequence of up to hundreds of range maps in few minutes. In order to reduce the number of pairs to align in the case of unordered range maps we use a prioritization strategy based on the fast computation of the correlation between range maps through FFT. The proposed system does not need any user input and it was tested successfully on a large variety of datasets coming from real acquisition campaigns.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The first step of 3D object acquisition through 3D scanning is to acquire, with some devices such as laser scanners, several range maps from different viewpoints to obtain a complete coverage of the object. After the data acquisition, the range maps have to be aligned and merged, in order to obtain the final surface. In order to perform the alignment, range maps must have at least a partial overlap on the sampled area. Due to shape and material reflectance characteristics, the same surface region is usually acquired several times from multiple points of view, so to avoid leaving portions of the surface unsampled. For these reasons, especially for large objects, many range maps are required to fully cover the object surface.

The goal of the alignment (or registration) process is to estimate the rigid geometric transformations, i.e. roto-translations, that bring all the range maps in a common coordinate system. The range maps registration is usually split into three computational steps. A *coarse registration* step, which produces an initial rough positioning of the range maps by an approximated estimation of the roto-translations. A *fine local* pairwise alignment, usually based on the ICP approach introduced in [1] or its variants, to improve the accuracy of the coarse registration. Finally, a global registration step [2] to ensure the minimization of the global alignment error. If the result of the coarse registration is accurate it is possible to apply directly the global registration step. In these last years many efforts have focused on performing these steps in an unattended

way. Current state of the art algorithms allows for robust automatic coarse registration [3] and improves the results of the fine registration of pair or a set of range maps [4].

Here, we propose a new solution to obtain the coarse registration in short time even for large datasets. Our algorithm is based on a GPU-friendly features descriptor that allows for automatic alignment of a pair of range maps in 200–500 ms, outperforming the current state of the art solutions. In order to avoid false positives, a validation test is performed after each pair alignment. Although in the common practice range maps that are next to each other in the order of acquisition also overlap, we propose a simple prioritization algorithm based on FFT computation for when this assumption cannot be made. This prioritization permits to schedule the pairs of range maps to try for alignment and greatly reduces the total time. The result is an automatic system for the alignment of a very high number of range maps, characterized by

- *Speed*: the proposed GPU-based algorithm permits to align two range maps in about 200–500 ms on a low-end graphics board.
- *Effectiveness*: the system has been tested on several real dataset composed by a large number of range maps obtaining very good results.
- *Robustness and accuracy*: false positive results are very rare. The output alignment of the system is ready for the fine registration step.

2. Related work

Several methods to solve the problem of the automatic registration of range maps were presented in the literature. Here we discuss the ones more closely related to our approach.

[☆]This article was recommended for publication by Andrei Sharf.

^{*} Corresponding author.

E-mail addresses: paolo.pingi@isti.cnr.it (P. Pingi), massimiliano.corsini@isti.cnr.it (M. Corsini), fabio.ganovelli@isti.cnr.it (F. Ganovelli), roberto.scopigno@isti.cnr.it (R. Scopigno).

The algorithms for coarse registration can be divided in two main classes. The first one includes methods that use local feature descriptors to find pairs of corresponding points on the range maps. These methods generally show good convergence rate, but have a high computational cost. One of the first methods of this type was the Spin Images, proposed by Johnson and Hebert [5]. A spin image is generated using oriented points and it is a 2D histogram of the surface around a point. Points that belong to different views of the model and having similar spin image are assumed to be matching points. In [6] Li and Guskov present a method based on detecting a set of salient feature points using a scale-space representation based on a combination of Discrete Cosine Transform and local Discrete Fourier Transform. In Bonarigo et al. [3] the corresponding pairs of points on the scans are selected through a multi-scale analysis approach; once the features are extracted, the most reliable correspondences are matched. A recent evolution of this algorithm [7] is compared with our proposed system in Section 7. Pingi et al. in [8] proposed a method to manage complex scan sets acquired by following a regular scanner pose pattern. The method exploits the scanning sequence to define an initial adjacency graph. The points matching is based on a new shape characterization kernel that focuses on surface normals of adjacent points. Chen and Stamos [9] propose a solution based on the detection and matching of circular features, that has proved to be suitable on large-scale datasets of urban structures. King et al. [10] proposed a method based on detecting and matching keypoints in range images and in intensity images. This approach employs spin images to describe holes in smooth surfaces (range keypoints) and intensity gradient histogram to detect intensity keypoints.

The second class of coarse registration methods includes those that not use features descriptors. An algorithm of this class is the DARCES [11], an approach based on RANSAC scheme [12]. Chua and Jarvis [13] proposed a technique to register $2\frac{1}{2}$ D sensed data points to a model surface represented by another $2\frac{1}{2}$ D model data points. The algorithm selects three reliable points on the sensed surface and find the corresponding three model points using the principal curvatures and the Darboux frames to restrict the search over the model space. Many possible triplets are tested and a heuristic search is used to identify the optimal one. The global technique presented in [14] is based on the correlation of two Extended Gaussian Images (EGIs) in the Fourier domain. Finally, Aiger et al. in [15] present a method for registration of noisy data based on the extraction of all coplanar 4-points sets from a 3D point set that are approximately congruent to a given set of coplanar 4-points. The algorithm uses a RANSAC-like approach to find the optimal solution.

3. Our approach

The proposed system is illustrated in Fig. 1. Given a set $\mathcal{R} = \{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_N\}$ of input range maps, each one is resampled and the corresponding FFT is calculated. The *resampling* operation is very important and it has two purposes. First, it is used to turn the range maps in a power-of-two resolution, necessary to compute the FFT efficiently. Second, in order to achieve *scale invariance* of the descriptor, this resampling allows us to evaluate the FFT on a region of the same physical size. In fact, the resampling is done in such a way to normalize all the range maps to the mean distance of the depth samples in the set (evaluated excluding the borders of the range maps and the depth discontinuities).

The FFT is used by the *prioritization stage*, which computes the *matching order*, that is, the order the system follows in trying to pairwise align the input range maps ($\mathcal{R}_i, \mathcal{R}_j$). Before to accept the

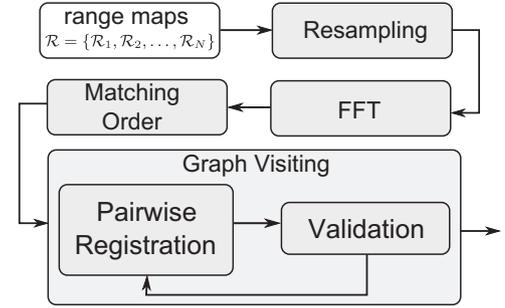


Fig. 1. Overview of the system.

result of a pair alignment, the validation block checks, using some heuristics and an ad hoc GPU-based technique, if the result of the alignment is correct or it is a false positive. In the next we detail the different processing stages of the system. All these stages are implemented in GPU, except the computation of the FFT, for which we employ the FFTW library [16], and minor parts of the pairwise registration algorithm.

4. Matching order

In the general case, we do not know a-priori which range maps overlap and can be aligned together. So, the set of N range maps requires $N(N-1)/2$ alignment attempts to fully cover all the possible pair alignments. Here, we considerably reduce the number of such attempts following a strategy based on a global similarity criterion.

The idea behind this strategy is that, if a range map A is *similar* to the range map B , there is a high probably this pair contains overlapping parts. To explore the similarities among all the range maps in a computationally efficient manner we apply, for each possible pair, the so-called *phase alignment algorithm* originally proposed in [17]. This algorithm permits to align two images, where one is the translated copy of the other, by maximizing the correlation over all the possible translations (t_x, t_y) . By assuming that two depth maps \mathcal{D}_A and \mathcal{D}_B are related by a simple translation, we can write

$$\mathcal{D}_A(x, y) = \mathcal{D}_B(x + t_x, y + t_y) \quad (1)$$

According to the phase alignment algorithm, we can estimate (t_x, t_y) by computing

$$(t_x, t_y) = \frac{F_A^* F_B}{\|F_A F_B\|} \quad (2)$$

where F_A and F_B are the corresponding FFT of the range map, i.e. $F_A = \mathcal{F}(\mathcal{D}_A)$ and $F_B = \mathcal{F}(\mathcal{D}_B)$.

Obviously, if the viewpoint of the range map B is very different from the other one this measure loses significance but we found it a good compromise in terms of computationally efficiency and robustness. We would like to point out that variants of the phase alignment algorithm can be used to account for other geometric transformation than translation, but, some preliminary tests we conducted in this direction (using [18]) suggested us that these variants are not robust in this context.

After identifying the candidate translation (t_x, t_y) , we calculate the measure of similarity $s(\mathcal{D}_A, \mathcal{D}_B)$ as the Normalized Cross Correlation (NCC) between \mathcal{D}_A and \mathcal{D}_B that is the translated version

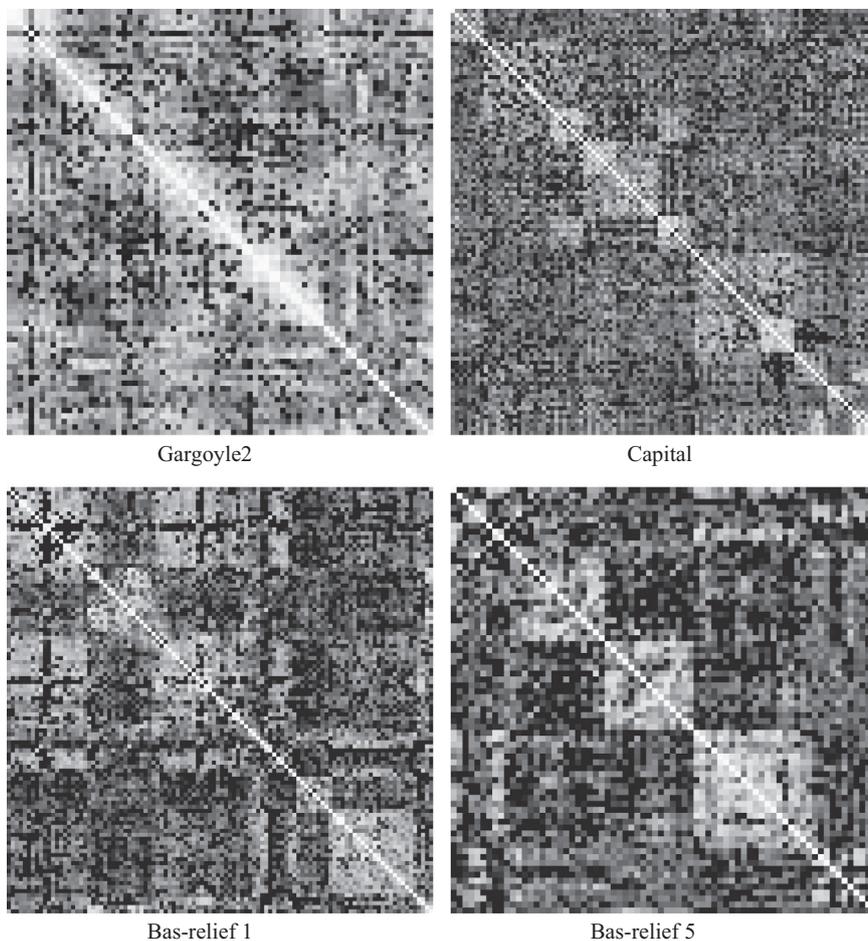


Fig. 2. Examples of the similarity matrix S obtained on some of the datasets used to assess the system performance. (a) Gargoyle2, (b) Capital, (c) Bas-relief 1 and (d) Bas-relief 5.

of \mathcal{D}_B :

$$s(\mathcal{D}_A, \mathcal{D}_B) = NCC(\mathcal{D}_A, \mathcal{D}'_B) = \frac{\sum \mathcal{D}_A(x, y) \mathcal{D}_B(x', y')}{\|\mathcal{D}_A\|^2 \|\mathcal{D}_B\|^2} \quad (3)$$

Note that the NCC is calculated on the entire range map and that the translated coordinates (x', y') are wrapped as what happens in the repeat mode of standard texture wrapping. If the two range maps are related by a simple translation this measure provides a good estimation of the NCC between the overlapping parts. If the two range maps are related by a rotation or by a more complex perspective transformation, the phase alignment algorithm still tends to provide a (t_x, t_y) which maximizes the cross-correlation between A and B , and then the calculation of the NCC in this way is still a reasonable measure of similarity. For each possible pair $(\mathcal{R}_i, \mathcal{R}_j)$ we evaluate this measure of similarity and we build a *similarity matrix* S . An example of S for some of the dataset tested in the results Section is shown in Fig. 2.

In general, we expect that the reliability of this measure of similarity increases as the overlapping part between the two range maps increases. Since our objective is to use S to find a good alignment sequence for the subsequent processing stages this is an advantage. The performance of this measure in the more general context of partial matching should be evaluated against other methods, for example the ones employed in image stitching. Since such techniques are, in general, not suitable for range maps we consider this an interesting matter of future research. Here, as a preliminary study, we report a performance evaluation of s in Section 7.1.2.

5. Graph visiting

After computing S , we associate a graph to it and we visit this graph to align the whole dataset.

The graph representation of the similarity matrix consists of one node for each row (or column, since S is symmetric) and a weighted arc $E(i, j)$ for each entry of $S(i, j)$. In this representation, to align the dataset means to find a *spanning tree*, i.e., a set of arcs $E' = \{(i, j) \mid i \text{ is aligned to } j\}$ such that the graph is *connected*. This is done by visiting the arcs $E(i, j)$ sequentially. For each visited arc, we try the alignment procedure if and only if its two end nodes are not already in the same connected component.

Listing 1 shows the visiting algorithm. Since we want to minimize the attempts of pairwise registration, first, we sort the arcs in descending order w.r.t. the similarity values. Every node belongs to a connected component (initially consisting in the node itself), that is returned by function `Comp (node)`. If the two nodes of the arcs belong to the same component we do not execute the registration procedure because we have already a transformation that aligns them that is obtained as the composition of the transformations in the path connecting the nodes. Otherwise we launch the pairwise registration algorithm (described in Section 6). If the alignment succeeds, the arc is added to the output and the two components are merged. The algorithm ends when either $N - 1$ arcs have been added, which means E' describes a spanning tree, or when all the $N(N - 1)/2$ arcs have been tested. In the latter case we will have a partition of the graph in

components, each one corresponding to a registered group of range maps.

Listing 1: Dataset alignment procedure.

```

Esorted = sortBySimilarityValue(E);
while ((E'.size() < n-1) && (!Es.empty())) {
e = Esorted.pop_front();
if (Comp(e.node1) != Comp(e.node2) {
  if (TryRegistration(e.node1, e.node2)) {
    E'.push_back(e);
    Merge(Comp(e.node1), Comp(e.node2));
  }
}
}

```

6. Pairwise registration algorithm

Given two partially overlapping range maps P and Q , our registration algorithm can be summarized as

1. Select region of interest (\mathcal{I}) on P .
2. Select n random points p_i inside \mathcal{I} .
3. Compute the descriptor D_{p_i} for each p_i and match them in Q .
4. For each possible quadruple of correspondences estimate the rigid roto-translation (\mathcal{M}).
 - 4.1 Evaluate the consensus of \mathcal{M} ($\#C$).
 - 4.2 Store the \mathcal{M} with the the minimum alignment error (e_A) among those with maximum consensus.
5. If $\#C < \#C_{threshold}$ the alignment fails.

All the main steps of the algorithm are calculated on the GPU. In order to provide a clear description of the algorithm, we must first define the descriptor D . The input range maps are sampled regularly as a 2D height fields, this allows us to easily define, for a point p , a regular area of adjacent samples composed by $m \times m$ neighbors pixels. We use as descriptor a matrix in which each element $d_{ij} \in D$ is obtained by the dot product between the normal of p and the normal of p_{ij} . In the implementation of the descriptor we consider an area of 25×25 samples, which corresponds to a physical area of fixed size thanks to the resampling operation applied to each range map. To improve the efficiency in the comparison between descriptors we compare the odd neighborhood pixels only, decreasing the texture accesses to 168 ($13 \cdot 13 - 1$). This does not compromise the discriminability power of the descriptor since usually the normals vary smoothly on the surface. The descriptor just defined is not invariant to rotations around the view direction of the scanning device. This is not a major limitation because the range maps are usually acquired with small or no rotations of the device w.r.t its view axis. In general, as we will see in the Section 7, despite its simplicity this descriptor is fairly robust.

The region of interest \mathcal{I} is calculated taking into account the variance of the descriptor D for each pixel of the range map:

$$\sigma^2(D_p) = \frac{1}{m^2} \sum_{ij} (d_{ij}^p - \mathbf{E}[D_p])^2 \quad (4)$$

where \mathbf{E} is the average of each kernel:

$$\mathbf{E}[D_p] = \sum_{ij} \frac{d_{ij}^p}{m^2} \quad (5)$$

We are interested in the regions with “medium” values of variance. This is because regions with too low variance, i.e. flat regions, have less discriminative power because the values of the descriptor tends to be zero everywhere, and regions with very high variance are likely to correspond to depth discontinuities or to points in proximity of silhouette where, depending on the scanner position, we can have self-occlusion of the surface. According to these requirements we determine a confidence interval of variance $[v_{\min}, v_{\max}]$. Each point which belongs to that interval belongs also to \mathcal{I} (see Fig. 3 for an example).

After finding \mathcal{I} on the range map P , n random points are selected inside this region. In our tests n has been usually set to 30. Since we want to cover a wide part of \mathcal{I} with an uniform density of samples, these points are generated according to a Poisson-disk distribution. In particular, the following procedure is performed: a point is selected randomly, if in a radius r there are other selected points it is discarded. If, after 10,000 attempts, an insufficient number of points are generated the distance constraint r between the points is halved and the process is repeated. In case an insufficient number of points is generated after two repetitions the pair is discarded as not aligned.

For each selected point p on the region of interest $\mathcal{I} \subset P$, we have to find the best matching vertex $q \in Q$. We compare the descriptor D_p with D_q for every $q \in Q$ using the Euclidean distances of its components:

$$d(D_p, D_q) = \frac{1}{m^2} \sum_{ij} (d_{ij}^p - d_{ij}^q)^2 \quad \forall q \in Q \quad (6)$$

we choose as best matching the point having minimum distance $d(D_p, D_q)$. This step produces a set \mathcal{H} of points pairs (p_i, q_j) with $p_i \in P$ and $q_j \in Q$ with $1 \leq i \leq n$, $1 \leq j \leq n$, which could include some false matches. We estimate the final rigid roto-translation matrix \mathcal{M} from \mathcal{H} through an *exhaustive* search; for each combination of k different points pairs in \mathcal{H} we compute $\mathcal{M}_{(p_i, q_j)}$, in this way we obtain $\binom{n}{k}$ matrices. We opt to use an exhaustive approach because the number of point pairs is usually low. So the computational time to test all the possible solutions is negligible. This allows us to avoid potential failures of other approaches (e.g. RANSAC). Then, we find the matrix \mathcal{M} that produces the maximum consensus between the correspondences found. The consensus $\#C$ is given by the number of correspondences that are under a geometric

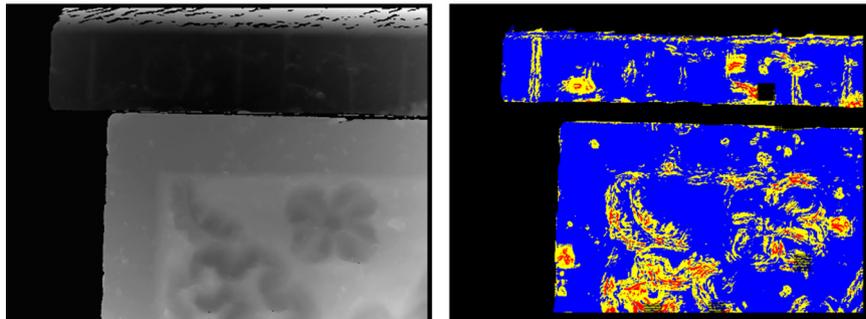


Fig. 3. (Left) Input depth map. (Right) Regions of interests. Red color corresponds to lower variance than the reference interval, blue regions to higher variance. The region of interest is colored in yellow. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

distance threshold (1.5 mm in our implementation) when \mathcal{M} is applied. Between the matrices with the maximum consensus, the one with the minimum alignment error (e_A) is chosen (Fig. 5 shown an example). The alignment error is calculated as

$$e_A(\mathcal{M}) = \frac{1}{\#C} \sum_{(p_i, q_j) \in C} (\|p_i - \mathcal{M}q_j\|^2) \quad (7)$$

where C contains the correspondences under the distance threshold. If the final consensus is less than $\#C_{threshold}$ (set to 6 in our implementation) the alignment fails. We point out that the minimum value of points to determine a roto-translation is $k=3$. Using higher values of k reduces the possibility to find false positives and at the same times reduces the number of possible combinations and thus the computation time. In our experiments we found that $k=4$ gives a good tradeoff between time and robustness.

The algorithm just described is not applied at once on the entire range maps, but iteratively on four different zones of P (see Fig. 4). This strategy does not introduce significative additional computation time but helps to maximize the number of points p_i found on the overlapping region between P and Q . We have a real failure of the alignment when the algorithm fails to find a corresponding quadruple on all the four selection zones.

From an implementation point of view, the only parts implemented in CPU are the generation of the random points p_i and the exhaustive search to find \mathcal{M} .

6.1. Validation

In order to avoid false positive, after computing the transformation matrix \mathcal{M} for a pair, we use the test described in [19] to check if the two range maps overlap in a meaningful way. This is done exploiting the rasterization pipeline. The key idea is that if two range maps are aligned, rendering them produces *almost* the same result on screen (in the overlapping region).

Following this idea, the validation algorithm works assigning two different constant colors to the two range maps and rendering them two times: the first bringing one of them slightly towards the point of view, the second positioning the same range map slightly away from the point of view (see Fig. 6(b)) and counting the number of pixels that changed color between the two renderings. If the alignment is good and the range maps are very close, even a slight displacement along the viewing direction will correspond to a large number of pixels changing color between the two renderings. Fig. 6(c) shows (in black) a map of all the pixels that have different colors in the two rendering. If the number of pixels of the overlapping part is under the 60% of the total the alignment is not accepted.

7. Results

The proposed system has been tested on several real range maps datasets. All the dataset have been acquired through a

triangulation laser Konica Minolta 910 in several acquisition campaigns. A brief description of each dataset follows:

- Urn** This dataset is composed by 17 range maps. It is the dataset used to show the step-by-step processing during the paper. It is “simple” since the relative motion between range maps is mainly translational.
- Gargoyle1** This dataset regards a small test object, a gargoyle statuine, and it is composed by 27 range maps.
- Gargoyle2** This is another version of the same gargoyle statuine with much denser sampling, i.e. 78 range maps have been acquired instead of 27.
- Column** The *Column* dataset is coming from an acquisition campaign of the remains of a late medieval complex, the funerary monument of the emperor Arrigo VII (did in the ambit of an EU project). This dataset is composed by 115 range maps.
- Capital** This medieval capital (Museum of San Matthew in Pisa) is sampled with 103 range maps.
- San Leonardo Church** This dataset is composed by 6 bas-reliefs acquired in the San Leonardo Church in Arcetri. The six different objects represent six corresponding bas-reliefs of the pulpit of the Church. Even if the bas-reliefs are planar object, the acquisition has been done using a scaffolding making the range maps of these objects characterized by very different sampling densities and different angles of viewing.
- Toy Car** Even if our system is designed to work with high quality 3D scans, like the ones usually acquired in 3D scanning campaigns, we use this dataset to test it on range maps with a moderate amount of noise. In fact, since the car acquired is metallic, the resulting scans present a sort of Moiré pattern.

7.1. Impact of the processing stages

In Table 1 we report the detailed results, in term of output and timing, of each stage of the algorithm. These experiments have been conducted on a consumer PC equipped with an Intel Core i7 CPU (3.4 GHz), 24 GB of RAM and a GeForce GTX 760 NVidia graphics card.

If the system produces more than one group of range maps, the number of the range maps in each group is listed, separated by a comma. This is the case of the Gargoyle1 (four groups of 2 range maps each).

7.1.1. Prioritization stage and visiting strategy

The fourth, fifth and sixth columns of the Table 1 allow us to understand the impact and performance of the prioritization stage and of the visiting strategy adopted. The fourth column of the

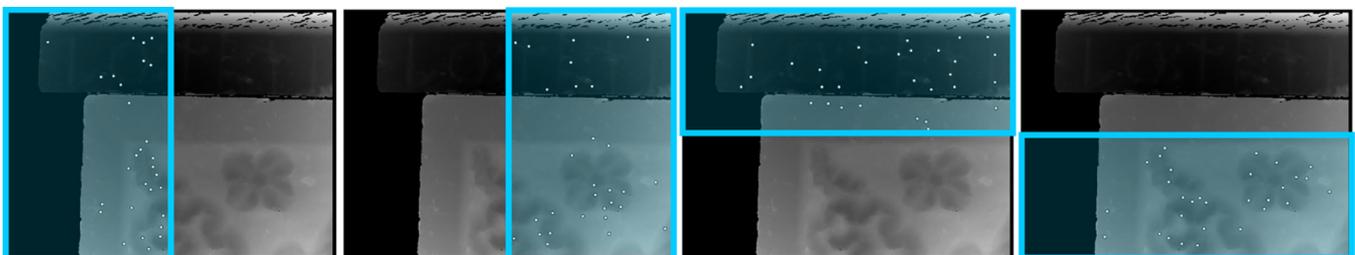


Fig. 4. Selection strategy. The four zones of interest inside which points are selected.

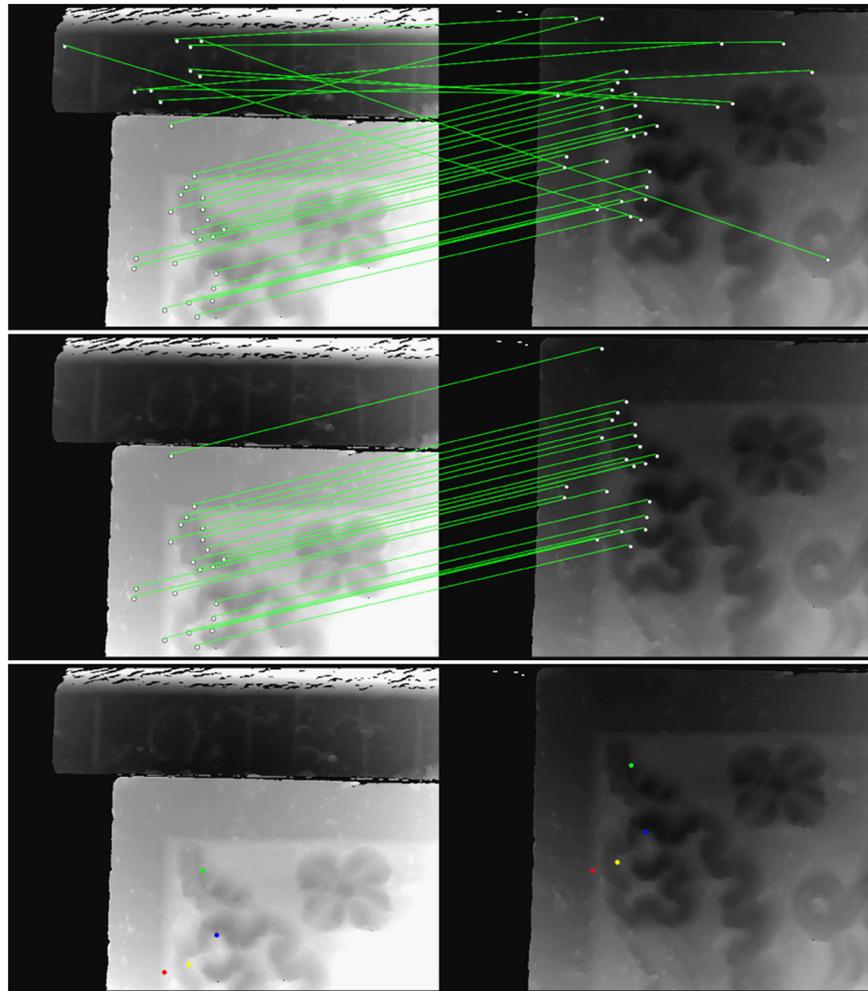


Fig. 5. (Top to bottom) Initial correspondences; correspondences with maximum consensus; best quadruple.

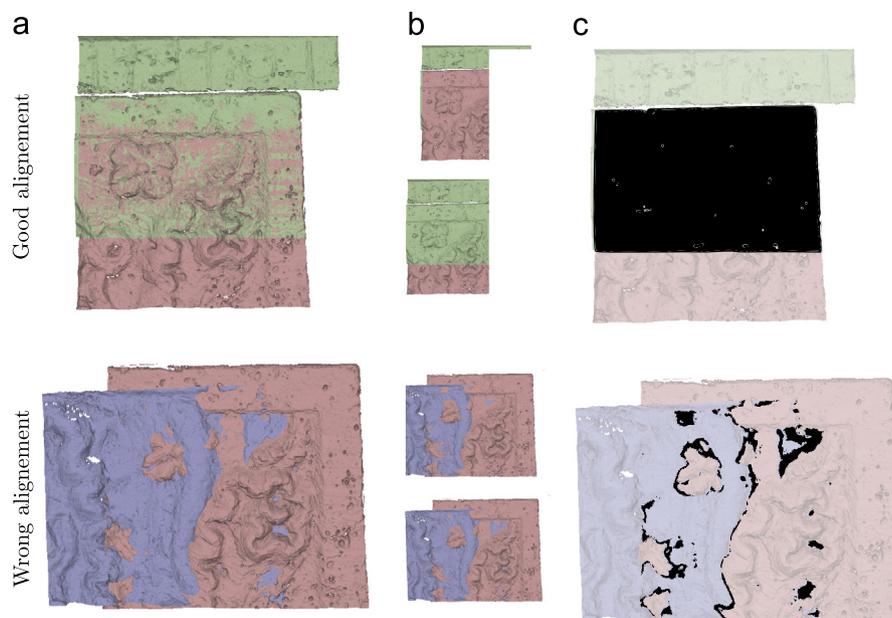


Fig. 6. Validation test. (a) Two pairs of range maps: correctly aligned (Top) and erroneously aligned (Bottom). (b) Two rendering displacing the red range map towards and away from the viewer. (c) Map of the pixels which changed color between the two renderings. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

table shows the number of arcs visited during the overall processing. The percentage is relative to the total number of potential arcs to visit (that is $N(N-1)/2$). Obviously, if the system fails to align all the given range maps in a single group, it means that the 100% of the arcs (all the pairs) were visited, regardless of the value of similarity values computed in the prioritization stage.

The fifth column contains the number of times the pairwise alignment algorithm has been actually performed. We recall that, according to the visiting strategy of the graph described in Section 5, an arc is tested for alignment only if the two corresponding range maps belong to different groups of aligned range maps. This approach, together with the order of the arcs provided by the similarity matrix, is able to greatly reduce the number of pairwise alignments required. For example, for the bas-relief1 we did about the 7% of the potential alignments saving the 93% of the time necessary if we had to test all the pairs.

We note that, in some cases, the similarity matrix S is not able to provide an efficient matching order between the arcs. When this happens, a great number of pairwise alignments need to be tested. This is for example the case of the Column dataset, where each range map is acquired through a rotation step of about 45° around the object, making the correlation between range maps an unreliable measure of similarity. In fact, in this case more than the 70% of the total pairs have been tried.

The processing time to calculate the similarity matrix S , reported in the sixth row, is in general low even in case of many range maps. This underlines that the prioritization stage always worthwhile to be computed for unordered set of range maps.

7.1.2. Evaluation of similarity measure

In the previous section we have discussed some results about the performance of the matching order provided by the prioritization stage together with the visiting strategy adopted. In order to better assess the performance of the similarity measure employed and, hence, the performance of the prioritization stage isolated from the rest of the processing pipeline, we provide here a preliminary analysis about how much the similarity measure proposed is suitable for the task of finding a good matching order between unordered range maps.

In the general case, it is not trivial to quantify the goodness of a certain matching order w.r.t to another, because the final error of different orders can depend also by the subsequent processing steps. Therefore, we decide to evaluate how much the similarity measure is able to identify overlapping range maps. Obviously, higher is the probability to skip non-overlapping range maps, higher is the probability to align the entire set in few runs of the pairwise registration algorithm.

To do this evaluation, we calculate for each possible pair of range maps of a dataset, the amount of overlap between the range

map A and the range map B. The amount of overlap is estimated as

$$O(A, B) = \frac{2\mathcal{A}(A \cap B)}{\mathcal{A}(A) + \mathcal{A}(B)} \quad (8)$$

where $\mathcal{A}(A)$ and $\mathcal{A}(B)$ is the area of the surface of A and B, respectively and $\mathcal{A}(A \cap B)$ is the area of the surface of the overlapping region. This value is estimated with a variant of the validation algorithm applied on the aligned dataset after the ICP refinement (see Section 7.3). This provides us a ground truth on data coming from real acquisition campaigns.

At this point, we rank each possible pair according to its similarity, and we evaluate the number of range maps pair which have a good amount of overlap ($> 20\%$). This evaluation is summarized in Fig. 8 for some of the test dataset. The X axis is the rank number. The Y axis is the number of pairs correctly identified as overlapping. The bluish line regards the results obtained using the similarity measure. The orange line is a reference and represents the number of pair correctly identified as overlapping if the range maps would be uniformly distributed and the pairs randomly selected. For example, considering the Gargoyle2 dataset, at the position 100 of the rank 96 range maps have been correctly identified by the similarity measure while only 44 by the random selection. By examining the graphs reported we can state that s works well in identifying overlapping range maps pairs which are good candidates to be passed to the pairwise registration algorithm.

Concerning the dependency between the amount of overlap and the value of similarity, we can state that we do not have a strong correlation between these two values. Anyway, according to Eq. (3) we expect that low values of overlap ($< 20\%$) do not exhibit high values of s . In fact, as it is possible to note in the 1st and 2nd column of Table 2 the pairs with low overlap have often a low value of s (< 0.4). The percentage of pairs with low s and low overlap w.r.t the total of pairs with low overlap is also reported. Moreover, the average value of similarity of the overlapping pairs (3rd column) is always lower than the ones of non-overlapping pairs (4th column). So, thanks to the ranking we employed, at the end only a small fraction of the non-overlapping pairs are tested by the pairwise registration algorithm.

7.1.3. Pairwise registration

The pairwise registration algorithm proposed is very fast, as expected. As reported in the seventh column of the table, the pairwise registration time ranges from the 150 ms of the Gargoyle to the 674 ms of the Bas-relief 5, with an average time on all the datasets of about 480 ms. The validation stage is included in this time, and it is usually a small fraction of the pairwise registration time. According to the results obtained, the validation operation is very robust and allow us to avoid to obtain false positives.

Table 1
Results obtained on real dataset.

Dataset	Range maps	Aligned	Arcs visited	# Pair alignment	Prioritization stage (s)	Average time (ms)	Total proc. time (s)	Total proc. time (s) ^a
Urn	17	17	45(33.1%)	27(19.8%)	1.5	451	13.7	6.0
Gargoyle1	27	2,2,2,2	351(100%)	351(100.0%)	2.3	150	57.4	55.6
Gargoyle2	78	78	1939(64.6%)	173 (5.8%)	16.9	314	1 m 33	34.7
Column	115	90	6555(100.0%)	4776(72.9%)	38.1	498	40 m 25	37 m 16
Capital	103	101	5253(100.0%)	1307(24.9%)	32.2	434	10 m 9	6 m 13
Bas-relief 1	106	106	1542(27.7%)	398(7.2%)	30.9	559	4 m 16	2 m 34
Bas-relief 2	71	71	1205 (48.0%)	505 (20.3%)	14.1	575	5 m 8	3 m 42
Bas-relief 3	75	73	2775(100.0%)	695(25.0%)	15.8	496	6 m 5	3 m 56
Bas-relief 4	68	67	2278(100.0%)	635(27.9%)	12.9	626	6 m 54	4 m 31
Bas-relief 5	72	72	1252(49.0%)	276(10.8%)	14.6	563	2 m 51	1 m 42
Bas-relief 6	83	83	1605(47.2%)	609(17.9%)	19.4	674	7 m 14	4 m 48
Toy Car	23	23	112(44.3%)	82(32.4%)	1.7	418	37.3	7.5

^a Total processing time obtained by the system using the sequential assumption.

7.1.4. The sequential assumption

As demonstrated by the experimental results reported, the prioritization stage is able to save a considerable amount of work. In order to further improve the performance of the system, we implemented a version where the similarity matrix is adjusted, assuming that, if two images have been acquired sequentially with high probability they are overlapping ones. In other words, sequential range maps are good candidate to be matched. We call this *sequential assumption*. Following it, for each range map we test the arc $(i, i - 1)$ and $(i, i + 1)$ of S before the others. We found that this simple variant permits to reduce the number of pair alignments performed by the system, and hence the total processing time. The last column of Table 1 shows the total processing time for this implementation. In some cases, the performance are greatly improved. Instead, there was no significant performance gain for certain dataset, for example for the Column one, since almost all the pairs are required to be tested to produce the final result.

7.1.5. Overall performance

The overall processing time reported demonstrates that the system is able to work on big datasets in reasonable time, taking into account that the system does not use prior knowledge about the dataset, and that it works in a completely automatic and unsupervised way.

The histograms of Fig. 7 provide more details about the performance of the system. Such histograms report the total time required to align a certain range map. Note that some range maps are aligned very quickly, in less than 250–500 ms. Instead, especially when the prioritization stage does not provide a reliable order, some range maps required many trials before to be aligned.

Table 2 Amount of overlap vs similarity measure.

Dataset	Low O (total)	Low O and Low s	Avg. s (w/ overlap)	Avg. s (w/o overlap)
Urn	96	70(73%)	0.37	0.16
Gargoyle2	1667	1142(68%)	0.38	0.29
Bas-relief1	4210	3545(84%)	0.23	0.16

Another interesting issue to point out is that a dense acquisition, in terms of viewpoints, allows the prioritization stage to find a very efficient order and, at the same time, the proposed normals-based descriptor to work with high robustness, as demonstrated by the results obtained with the gargoyle datasets. While the system failed to align the Gargoyle1 dataset due to a poor coverage of the surface of the statuine (these range maps contains poor geometric information since they are acquired positioning the device a bit far from the object), the dense version of this dataset (Gargoyle2) not only is aligned perfectly but in a very short time as demonstrated by the time reported in Fig. 10 and in the table.

Since our system is designed to work with high quality scans (the ones usually coming out from a scanning campaign) we expect that it fails to align noisy scans, like the ones obtained with a Kinect device for example. Anyway, the results obtained on the Toy Car dataset show that our simple descriptor is able to deal also with a moderate amount of noise (see Fig. 9).

7.2. Comparison and discussion

To our knowledge, the most rapid method for the coarse registration of range maps is the one of Bonarrigo et al. [3], which has been improved and implemented in GPU during the development of this work [7]. So, we compare our system with the latter one (referred with BPS2013 in the following). BPS2013 is based on the creation of a features database. The features adopted can be thought as an extension of the Lowe’s SIFT [20] to 3D range data. This database is used to find the range maps to align together. When a range map arrives in input, the most similar range maps are retrieved from the database and then a multi-view (coarse) registration algorithm is launched on these range maps. The multi-view registration is based on matching descriptors associated to the keypoints previously extracted. This rotation-invariant descriptor is based on normal and saliency, encoded in a sector-based grid. Rotation invariancy is achieved by aligning the maximum values of this grid. The comparison results are summarized in Table 3.

From the results reported we can see that our system is able to perform faster in some cases, while in another case, the computation time is considerably higher than BSP2013. This is because the database of features is extremely efficient in finding the range maps to match, while, in some cases, our prioritization stage is not able to obtain similar performance. For example, for the Capital we

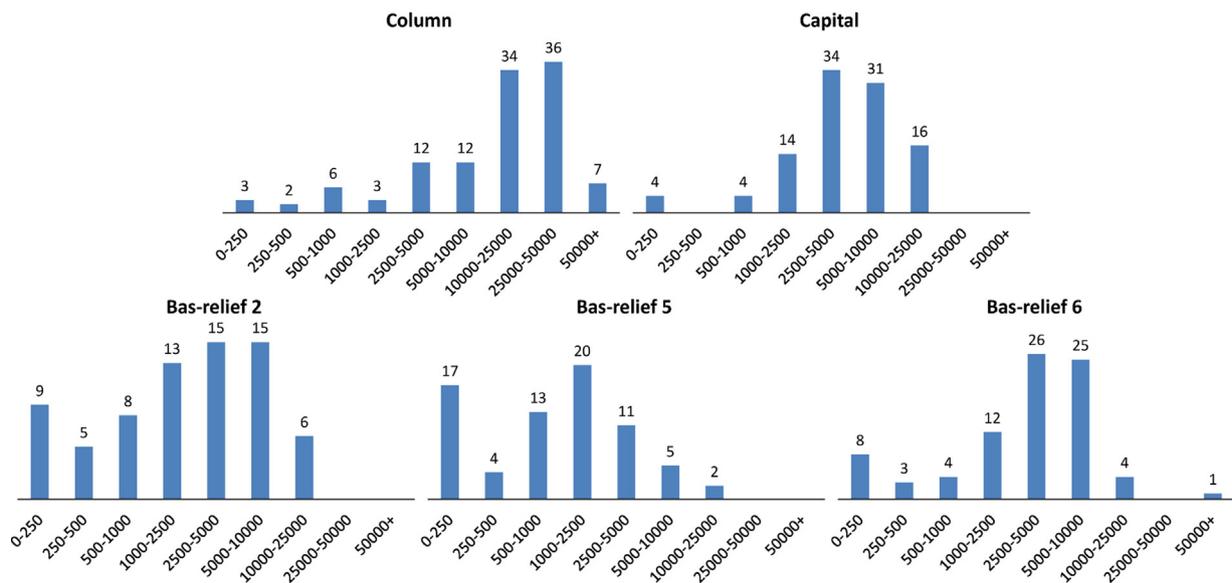


Fig. 7. Processing time details. On the abscissa is time intervals and on the y axis the number of range maps processed in that interval.

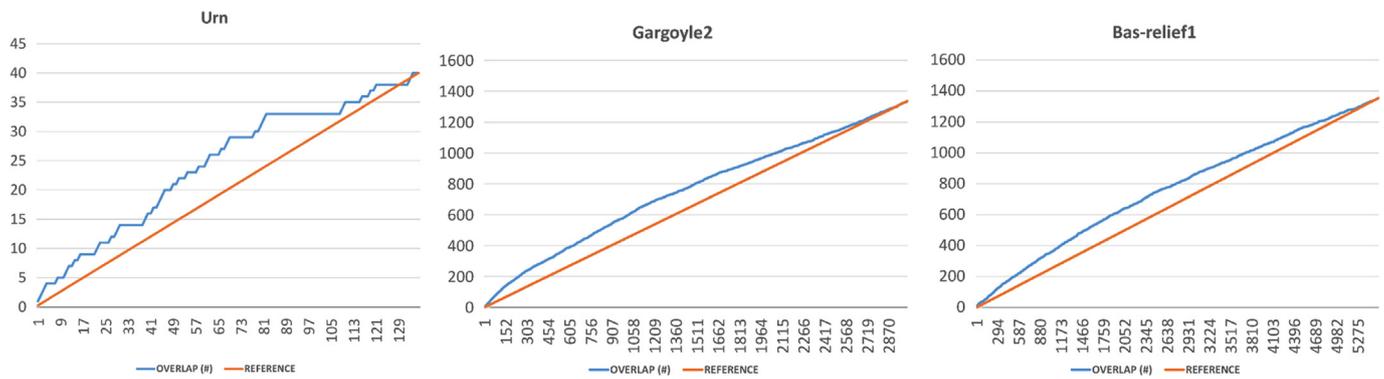


Fig. 8. Evaluation of the capability of the prioritization stage in finding overlapping range maps. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

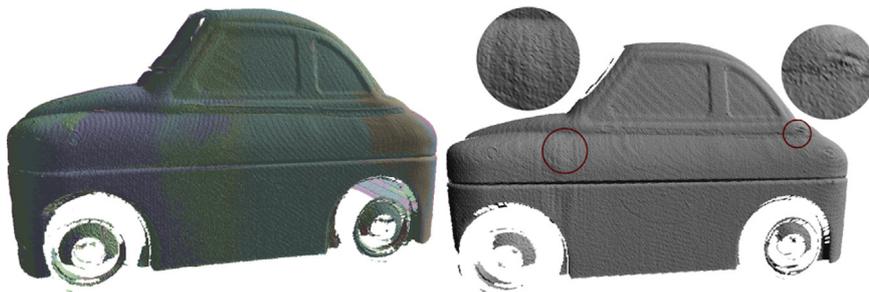


Fig. 9. Toy car. (Left) Result. (Right) A single range maps is shown to highlight the noise pattern of the scans.

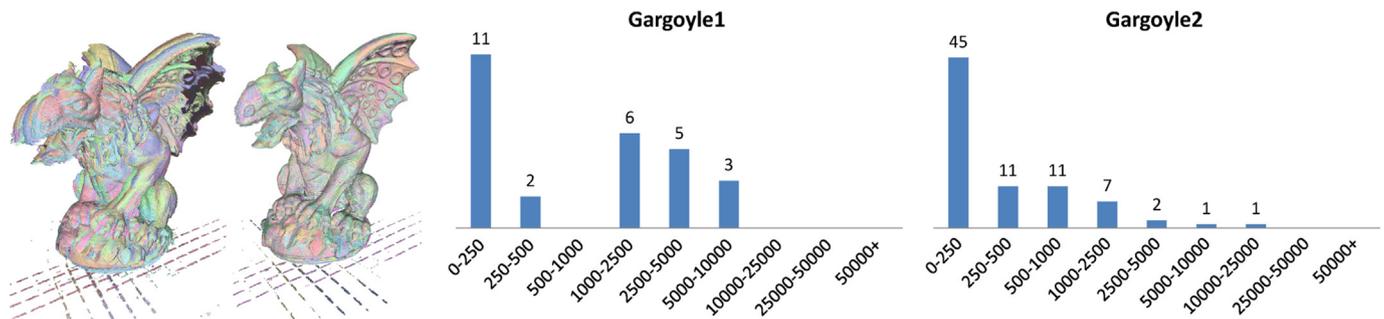


Fig. 10. Comparison between Gargoyl1 and Gargoyl2 datasets. (From-Right-To-Left) Output of our system. Alignment after ICP. Time statistics of the Gargoyl1 dataset. Time statistics of the Gargoyl 2 (dense dataset).

Table 3
Comparison against the BSP2013 system.

Dataset	Range maps	Aligned [BSP2013]	Total time (s) [BSP2013]	Total time (s) [PROPOSED]
Urn	17	17	6.9	6.0
Gargoyl2	78	78	32.4	34.7
Capital	103	101	1 m 13	6 m 13
Toy Car	23	23	11.6	7.5

did about 1300 alignments to align the dataset against the about 200 of the BSP2013. This result suggests that SIFT-related features are very suitable for finding similar range maps, while correlation suffers too much of viewpoints differences.

Although that the sequence finding performs better than our approach, the proposed pairwise registration algorithm performs with similar times, or better in some cases (183 ms is the average alignment time for the Gargoyl dataset, 229 ms for the Toy Car, while BSP2013 needs 250–300 ms on the average).

Concerning robustness, we obtain basically the same results. This is an interesting result because the descriptors of BSP2013 is very robust in terms of repeatability. Even if our descriptor is very

simple, using it for the matching with the validation stage ensure very good results in terms of robustness (we do not report any false positive). This suggests that a powerful validation stage plays an important role in a registration system of this type, more than to have a highly repeatable matching descriptor.

However, we can state that our system is very straightforward to implement, since it is based on the computation of an FFT (a lot of available libraries and code exists) and a shader code to compute variance and match the simple normal-based descriptor. The BSP2013 instead requires the management of the database, the clusterization of the features extracted (this is necessary in order to keep the retrieve time near-constant), the update of the database of features, the calculation of the mesh saliency, and the matching of the rotation invariant descriptors.

7.3. Quality

In order to evaluate the quality of the results obtained, we run an Iterative-Closest-Point (ICP) algorithm [1] to the output of our system. If the alignment computed is sufficiently close to a good one, we expect that the range maps converge to the real surface. In our cases this convergence is very fast (1–2 iterations on the

implementation used [21]). As shown in Fig. 11, the final aligned range maps have a very good quality, as can be noted by inspecting the details of the range maps which are perfectly aligned after the ICP. The false color used help to distinguish the different range maps, and so to better evaluate visually their respective alignment. Note also that, thanks to the validation stage, the output of the system does not present wrongly aligned range maps.

7.4. Limitations

The proposed system suffers from two limitations. The first one concerns the pairwise registration algorithm; the acquisition devices is assumed to have only a small rotation about the z axis because the proposed descriptor is not invariant against this type of rotation. The second limitation regards the prioritization stage and it is related to the fact that, if the devices movement are so big that the point of view differs a lot, the correlation could fail to provide a good matching order. This does not influence the final outcome of the alignment but may negatively influence the processing time. Despite these two limitations we would like to underline that, in a real scanning pipeline, these conditions are

quite rare. In fact, often the surface acquisition of small-medium object is dense and the scanner is rotated around its z axis only to acquire very specific undercuts.

8. Conclusion

In this work we have proposed a simple-to-implement GPU-friendly system capable to align many range maps in short time. We have demonstrated that even if simple and with some limitations, the proposed system is capable to produce very high quality results on real data coming from real acquisition campaigns. Additionally, the proposed validation stage has been demonstrated to be very robust avoiding to produce wrong alignments. In the near future, this system will be made publicly available as a Meshlab plugin [21]. A future challenge regards the extension of the system with a fast rotation-invariant descriptor. Designing and testing different metrics, for example other ones used in image stitching, to improve the prioritization stage is also another interesting direction of work.

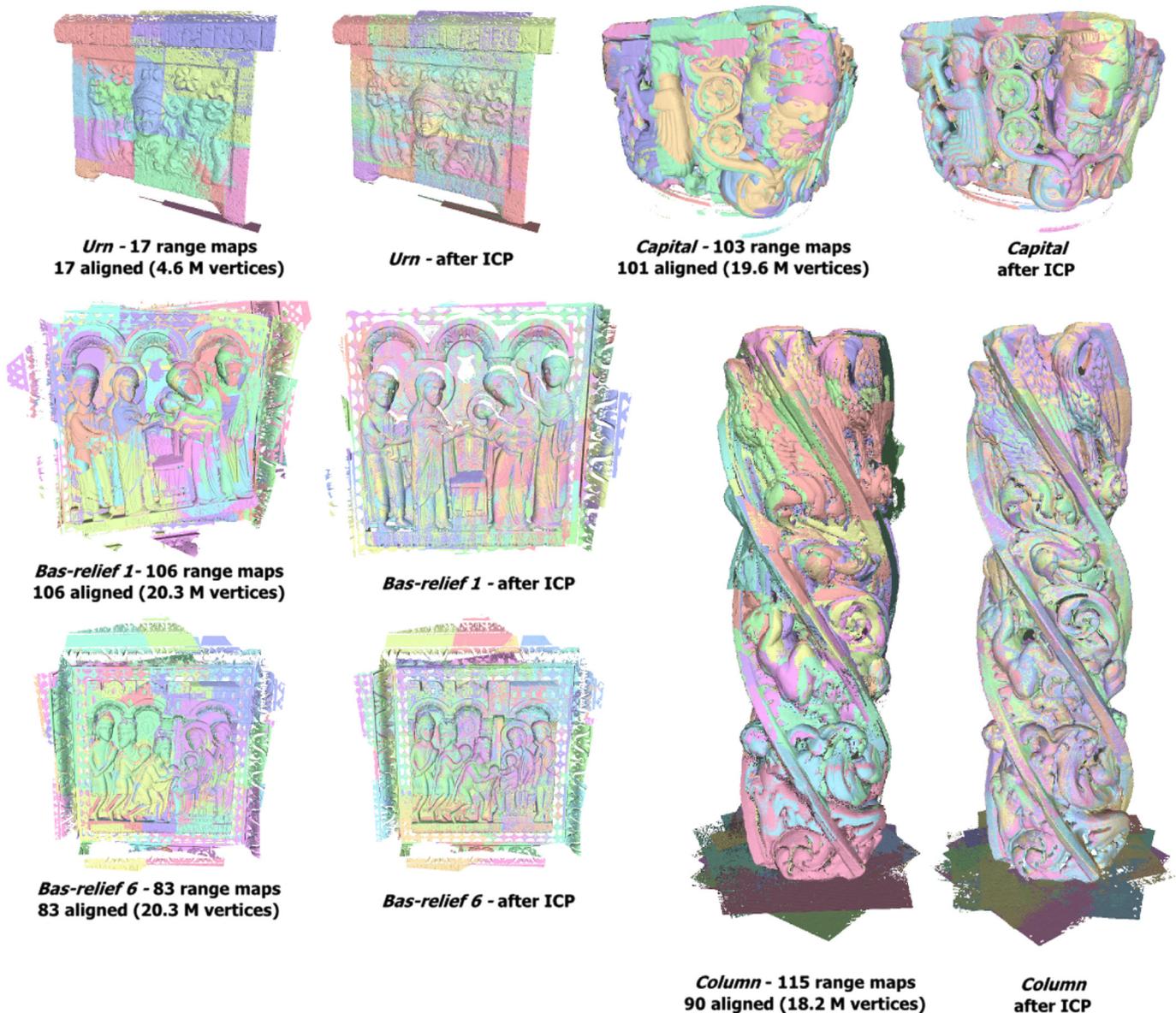


Fig. 11. Some of the results obtained from real dataset.

Acknowledgment

This work was partially funded by EU FP7 Project ICT FET “Harvest4D” (Grant Agreement G.A. no. 323567). We would like to thank also Alberto Signoroni and his colleagues to help us to compare our system with his one.

References

- [1] Besl PJ, McKay ND. A method for registration of 3-D shapes. *IEEE Trans Pattern Anal Mach Intell* 1992;14(2):239–58.
- [2] Pulli K. Multiview registration for large datasets. In: Proceedings of second international conference on 3D digital imaging and modeling. IEEE; 1999. p. 160–8.
- [3] Bonarrigo F, Signoroni A, Leonardi R. A robust pipeline for rapid feature-based pre-alignment of dense range scans. In: ICCV2011. 2011. p. 2260–7.
- [4] Fantoni S, Castellani U, Fusiello A. Accurate and automatic alignment of range surfaces. In: Proceedings of 3DIMPVT2012. 2012. p. 73–80.
- [5] Johnson A, Hebert M. Surface registration by matching oriented points. In: 3DIM97. 1997. p. 121–9.
- [6] Li X, Guskov I. Multi-scale features for approximate alignment of point-based surfaces. In: Proceedings of SGP 2005. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association; 2005.
- [7] Bonarrigo F, Pezzotti N, Signoroni A. On-the-fly automatic alignment and global registration of free-path collected 3d scans. In: Digital heritage international congress (DigitalHeritage), 2013; vol. 1. 2013. p. 181–4.
- [8] Pingi P, Fasano A, Cignoni P, Montani C, Scopigno R. Exploiting the scanning sequence for automatic registration of large sets of range maps. *Comput Graph Forum (Proceedings of Eurographics '05)* 2005;24(3):517–26.
- [9] Chen CC, Stamos I. Range image registration based on circular features. In: 3DPVT. IEEE Computer Society; 2006. p. 543–50. ISBN 978-0-7695-2825-0. <http://dx.doi.org/10.1109/3DPVT.2006.111>.
- [10] King BJ, Malisiewicz T, Stewart CV, Radke RJ. Registration of multiple range scans as a location recognition problem: hypothesis generation, refinement and verification. In: 3DIM. IEEE Computer Society; 2005. p. 180–7. ISBN 0-7695-2327-7. <http://dx.doi.org/10.1109/3DIM.2005.68>.
- [11] Chen CS, Hung YP, Cheng JB. RANSAC-based DARCES: a new approach to fast automatic registration of partially overlapping range images. *IEEE Trans Pattern Anal Mach Intell* 1999;PAMI-21(11):1229–34.
- [12] Fischler M, Bolles R. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun ACM* 1981;24(6):381–95.
- [13] Chua CS, Jarvis R. 3d free-form surface registration and object recognition. *Int J Comput Vis* 1996;17(1):77–99.
- [14] Makadia A, Patterson AI, Daniilidis K. Fully automatic registration of 3d point clouds. In: Proceedings of the 2006 IEEE computer society conference on computer vision and pattern recognition, vol. 1. CVPR '06; Washington, DC, USA: IEEE Computer Society; 2006. p. 1297–304. ISBN 0-7695-2597-0. <http://dx.doi.org/10.1109/CVPR.2006.122>.
- [15] Aiger D, Mitra NJ, Cohen-Or D. 4-points congruent sets for robust pairwise surface registration. *ACM Trans Graph* 2008;27(3):85:1–10.
- [16] Frigo M, Johnson SG. The design and implementation of FFTW3. *Proc IEEE* 2005;93(2):216–31 (Special issue on “Program Generation, Optimization, and Platform Adaptation”).
- [17] Casasent D, Psaltis D. Position, rotation, and scale invariant optical correlation. *Appl Opt* 1976;15(7):1795–9.
- [18] Reddy B, Chatterji BN. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans Image Process* 1996;5(8):1266–71.
- [19] Corsini M, Dellepiane M, Ganovelli F, Gherardi R, Fusiello A, Scopigno R. Fully automatic registration of image sets on approximate geometry. *Int J Comput Vis* 2013:1–21.
- [20] Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vision* 2004;60(2):91–110.
- [21] Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G. Meshlab: an open-source mesh processing tool. In: Sixth eurographics italian chapter conference. 2008. p. 129–36.