The 12th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2011)
M. Dellepiane, F. Niccolucci, S. Pena Serna, H. Rushmeier, and L. Van Gool (Editors)

Fast and Robust Semi-Automatic Registration of Photographs to 3D Geometry

Ruggero Pintus, Enrico Gobbetti, and Roberto Combet

Visual Computing Group - CRS4, Italy – http://www.crs4.it/vic/

Abstract

We present a simple, fast and robust technique for semi-automatic 2D-3D registration capable to align a large set of unordered images to a massive point cloud with minimal human effort. Our method converts the hard to solve image-to-geometry registration problem in a Structure-from-Motion (SfM) plus a 3D-3D registration problem. We exploit a SfM framework that, starting just from the unordered image collection, computes an estimate of camera parameters and a sparse 3D geometry deriving from matched image features. We then coarsely register this model to the given 3D geometry by estimating a global scale and absolute orientation using minimal manual intervention. A specialized sparse bundle adjustment (SBA) step, exploiting the correspondence between the model deriving from image features and the fine input 3D geometry, is then used to refine intrinsic and extrinsic parameters of each camera. Output data is suitable for photo blending frameworks to produce seamless colored models. The effectiveness of the method is demonstrated on a series of real-world 3D/2D Cultural Heritage datasets.

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics [I.3.3]: Picture and Image Generation—; Computer Graphics [I.3.7]: Three-Dimensional Graphics and Realism—.

1. Introduction

Modern 3D acquisition systems are able to rapidly digitize an object geometry with high accuracy and resolution, producing massive digital models with billions of samples. Such highly detailed models are extremely well suited for Cultural Heritage (CH), where both dense and extensive sampling is required. Pure geometry acquisition is the first step to digitally preserve CH items to prevent them from loss and deterioration; it is useful for renovation planning, research, and digital simulation. However, just a dense geometry is not enough for all CH needs: additional color information plays a key role in reconstructing a high-quality digital model.

Many approaches exist to obtain object color. Some range scanners acquire both color and range data at the same time, but their color resolution and quality are often insufficient for CH purposes; moreover, some of them lack this capability at all. One possible and automatic solution is a calibrated camera rigidly mounted on the scanner. Unfortunately, the different position of the laser beam and the image sensor results in possible occlusions, so that the color in some portions of the geometry will probably be missed. Even when these alignment problems can be solved, this simultaneous acquisition approach has too many limitations. For example, lighting conditions should often be different between 3D scanning and photographic campaigns. Moreover, often the photographic dataset is required to be modified in a second moment, e.g., to evaluate the effects of restoration. Recent powerful sensors allow us to measure color at a high resolution by simply using off-the-shelf cameras. Color acquisition with a free-handheld camera is thus highly desirable, but gives rise to 2D/3D registration problems.

Several approaches have been proposed that cope with the image to geometry registration problem, ranging from manual to semi or totally automatic pipelines. The more they are automatic, the more these pipelines rely on features in the geometry and in the image set, or on mutual information such as reflectance, color or normal attributes. Although manual methods are quite reliable, they are time-consuming and require a lot of user effort; on the other hand, completely automatic approaches are relatively fast, but their success depends deeply on the object type; hence, they are not robust enough to work with a generic dataset.

Our objective is the design of a practically useful robust method for simultaneously registering the input photos to the

DOI: 10.2312/VAST/VAST11/009-016

geometry with minimal user intervention and with no prior knowledge about the image set or the object geometry. The underlying idea of our approach is to exploit the state-of-the-art robust structure-from-motion (SfM) algorithms from the image-based 3D reconstruction domain. These methods cannot fully replace active acquisition methods for complete 3D reconstruction in the general case, but have proved to produce good results in self-calibrating sets of images. In this paper, we exploit a SfM method for unordered image collections to coarsely estimate relative camera poses, intrinsic camera parameters, as well as a sparse 3D model deriving from matched image features. The user has to manually select few 3D/2D matches in one image to coarsely map the SfM model to the dense input geometry in an affine manner. Then, a specialized sparse bundle adjustment (SBA), which additionally exploits the correspondence between geometries, solves for the final intrinsic and extrinsic camera parameters of all cameras. Our approach combines and extends state-of-the-art solutions, leading to an integrated system with unprecedented capabilities. In particular, our novel semi-automatic 2D/3D registration pipeline is capable to rapidly and precisely align a set of photos to the geometry of an acquired real object without prior knowledge on input data. We have evaluated our technique on a series of real-world 3D/2D CH datasets. Results show that the obtained 2D/3D calibration is suitable for photo blending frameworks to produce globally coherent colored models.

2. Related work

Image-to-geometry is a well-known topic in Computer Graphics and Computer Vision. Numerous techniques exist that try to solve this wide and extensively studied problem in different ways. These methods can be divided in three major classes, whether they depend on matches, features or statistics. For an overview of registration results in terms of different information-theoretic metrics please refer to the survey of Hantak and Lastra [HL06]. The output of these methods (intrinsic and extrinsic parameters aligning each camera with a 3D model), together with a geometry and a set of photos, are used in a texture blending framework to obtain colored models [PGC11]. Here, we discuss the approaches most closely related to our contribution.

Manual 2D-3D correspondence selection. These methods deeply depend on the user intervention, which has to manually select a number of correspondences between each image and the 3D geometry. Since this straightforward approach is tiring and time-consuming, some works try to reduce or ease the operations performed by the user. Borgeat et al. [BPB*09] exploit a GPU implementation of the SIFT algorithm to extract features in an interactive way in the current image and in the rendered model, to assist the user, showing him possible correspondences. Another approach [FDG*05] is based on a graph representation where the nodes are the 2D photos and the 3D model, while arcs encode both image-to-geometry and image-to-image correspondences. A graph-based framework exploits dataset redundancy and decreases the number of manually selected matches; further, it is easier to find image-to-image than image-to-geometry point pairs. Apart from the robustness of these manual approaches, they easily become unfeasible with large image sets. Our method asks the user to align a very small subset of images and it automatically computes the rest of image-to-image correspondences. Thus, it minimizes the amount of input given by the user, remaining suitable even if the image set size grows (hundred of photos).

Automatic 2D-3D feature detection and matching. Feature-based techniques find features that are present on both photos and the 3D model, and try to find consistent correspondences to solve the image-to-geometry problem in a completely automatic framework. This problem is in general very complex, since photographs and geometric models have a very different appearance, and it is hard to automatically find and match similar features. For these reasons, methods in this area are limited to some specific models. For instance, some works are employed to align building and urban environment datasets, based on the assumption that architectural models result in sharp edges in 3D and high contrast features in 2D. Some of them [SA01, KSSS09] rely on line features in 3D LIDAR datasets and in maps or floor plan images of outdoor and indoor scenes. Stamos et al. [SLC*08] developed a registration pipeline which relies on the presence of linear or circular 3D features in the range images. Other approaches rely on orthogonality constraints [LS05], clusters of vertical and horizontal lines [LYWZ06], or contours and silhouettes [LHS00]. Although we require a small user intervention, our method is applicable to a more general environment, since we rely only on finding similarities among images, which is a much simpler problem.

Semi-automatic 2D-3D statistical correlation. Statistics-based approaches are mostly based on the Mutual Information statistical measure, proposed for the first time by Viola and Wells [VI97], and compute camera poses by correlating information from images and the rendering of the geometry. Such information is the color or gray-scale intensity from photos and, exploiting range scanner capability of measuring intensity of the reflected laser beam, one or more 3D attributes, such as infra-red intensity [WLH*04, HL06], reflectance [IOT*07], LIDAR elevation and probability of detection [MKI09]. Corsini et al. [CDPS09] proposed an illumination-based registration that renders the geometry using ambient occlusion, specularity and normal cues. These approaches have two main drawbacks: first, they require a camera pose initialization to converge to the right solution, thus they are not completely automatic; then, attributes used for correlation purposes are not always provided. Again, our method is suitable for a more generic dataset, because it does not depend on any additional attribute.

Geometric multi-view reconstruction and matching. Some state-of-the-art works convert the operation of aligning an image set with a 3D model into a 3D-3D registration task. With a robust multi-view techniques they both find a global coarse camera pose estimation and derive a sparse point cloud from images. The alignment of the input geometry with the computed point cloud implicitly solves the original 2D-3D registration problem. Zhao et al [ZNH04] recover relative camera positions and a point cloud from a video sequence using motion stereo. The user has to manually register only two frames with the 3D model to obtain the absolute orientation and global scale. Then, an Iterative Closest Point (ICP) approach refines the registration. Instead of being limited to dense and ordered frame sequences, our method deals with unordered set of sufficiently overlapping photos, with a comparable amount of user intervention. Moreover, this method performs a rigid ICP refinement, while we adopt a deformable registration that uses a SBA to refine parameters of each single camera independently. Similarly to our, a recent work [LL09] addresses the 2D/3D alignment with a similar SfM framework and with an unordered set of images; however, the refinement step in that method depends on the presence of planes in the geometry. Thus, special light patterns are projected on planes to artificially produce features in uniform surfaces. As mentioned before, our approach is totally independent from these kinds of assumptions. Further, the cost function in that work depends on measures in both world (point-to-plane distances) and pixel (re-projection errors) coordinates; it weights the contribution of these terms with a heuristic parameter, that heavily depends on the object geometry/extent, and requires manual tuning. Conversely, our energy function contains only squared error measurements in image space and does not require any additional parameter.

3. Technique overview

Our technique is outlined in Fig. 1. We take as input a dense 3D model and a set of n photographs. The photographic dataset can cover the complete surface of the 3D object, only a part of it, or a larger area. No constraints are placed on the nature of the input dense geometry; it could be either a triangle mesh or a point cloud, and we don't need particular geometric attribute (e.g., normals or influence radii) or the presence or known geometric features (e.g., lines).

Our 2D/3D calibration is performed in three stages: SfM, coarse alignment, and fine registration. In the first fully automatic stage, we apply a SfM algorithm for unordered image collection to self-calibrate images and obtain an initial sparse 3D reconstruction of the part of the model covered by the photographic campaign (Sec. 5). This provides us a sparse 3D model derived from matched image features, all camera poses in a common reference frame, and the intrinsic parameters of each camera. In the second stage, the SfM model, reconstructed up to an unknown scale-factor, is coarsely aligned to the dense input 3D model with minimal user intervention. The user manually selects correspondences between a small subset of photos (typically just one) and the detailed model (Sec. 6). These matches, together with camera parameters, are used to solve for the affine transformation mapping the SfM world to the dense model. In the final stage, a SBA, which constrains the features detected in the images to lie on the fine 3D model, calculates the final registration in a non-rigid deformable manner (Sec. 7). The output data (camera parameters) can then be used, together with the n photos and the dense model, to blend the texture data on the geometry to produce a globally coherent colored model.

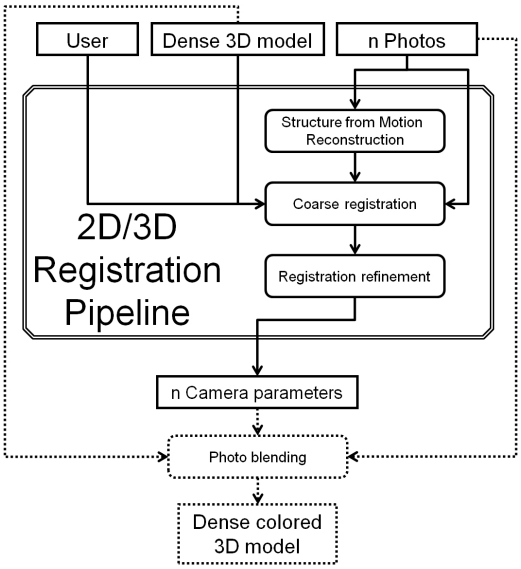

Figure 1: ***Pipeline.*** *Given the image set, a SfM algorithm computes a sparse point cloud and related camera poses. The user manually aligns one or more photographs to the dense model and our method uses this cue to coarsely register the SfM and the input model. The final registration, refined with a specialized SBA, can be used to obtain a globally coherent colored model, blending all registered photos together on the input point cloud.*

4. Photo capture

Besides avoiding to take images with excessive blur or noise, and under- or over-exposured regions, our pipeline does not impose particular constraints on the image set, since SfM algorithms exist to cope with challenging data, such as images that exhibit large variations in illumination, viewpoint, zoom, resolution, and contain outliers and clutters. For a description of typical SfM capabilities and limitations see the work of Snavely et al. [SSS08]. Further, techniques exist which perform texture blending for producing seamless colored models with such non-ideal color information

[PGC11]. For both methods, we only need sufficient overlap among images. A good practice is to have the same feature being visible in, at least, three or four photos.

5. Structure from Motion reconstruction

The first step of our pipeline is the self-calibration of the image collection, independently from the dense 3D geometry. This task is performed using a robust SfM algorithm suitable for aligning unordered large image collections [SSS06]. For each image, the method computes several thousand SIFT keypoints [Low04] and uses approximate nearest neighbors [AMN*98] and RANSAC [FB81] approaches to estimate right matches between them across multiple images. Then, a SfM algorithm recovers camera poses and sparse geometry by minimizing a non-linear energy function proportional to the re-projection error of 3D points into original image features. Given N_C photos, the output is a list of N_C estimations of intrinsic (i.e., focal length, principal point and distortion coefficients) and extrinsic (i.e., rotation and translation) camera parameters $C = [c_1, c_2, ..., c_{N_C}]$, a list of N_P 3D points $P = [p_1, p_2, ..., p_{N_p}]$ (i.e., sparse points), and pixel coordinates $s_{i,j}$ of the projection of a sparse point p_j in the i_{th} input image (i.e., keypoint location for that 3D point).

6. Coarse alignment

After the SfM stage, we have two geometric representations with different scales, reference frames and resolutions: one dense, provided as input, and the other sparse, deriving from SfM. To position cameras into the reference frame of the detailed input model, we need to find the affine transformation that determines the scale, rotation and translation, which better aligns the coarse and fine geometric models. Unfortunately, an automatic alignment approach is not reliable and would often be infeasible, because we are dealing with a sparse-to-dense registration of models at different scales, and we do not want to impose prior knowledge on geometry. For instance, a feature-based method is not appropriate for such subset-matching problem, i.e., where one point-set is matched to a point-set of greater cardinality. 4PCS [AMCO08], quadratic assignment [BCPP98], or generic non-linear optimization techniques could be applicable, but they are time-consuming and their convergence is not always guaranteed. We thus prefer to require the intervention of the user, which can be minimal since very few parameters need to be estimated. Thus, the user has to align one image to the fine model by graphically selecting few matches (i.e., typically from 7 to 12) between 3D points in the fine model and image pixels. Using the intrinsic parameters computed by SfM, we can estimate the pose of the selected camera in the reference frame of the fine model by minimizing re-projection error, i.e. the sum of squared distances between the picked image points and the projection of the selected object points. Optionally, the process can be repeated independently for two or three images, chosen so that the mutual distances between camera pairs are as large as possible, minimizing error drift. It should be noted that this procedure assumes that the SfM pipeline is capable to produce a model which is approximately correct and does not contain major geometric errors, especially systematic ones. If this is not the case, e.g., in the presence of large drifts possibly generated by sequential SfM approaches, coarse alignment may fail. In our experience, such failure case occurs very rarely in practice. Moreover, drift-related problems can be mitigated by using more robust multi-stage SfM pipelines [GFF10, SSS10] or by manually splitting the input image dataset, applying our technique to each obtained subset, and merging the results before refinement.

Using the intrinsic and extrinsic parameters estimated for that small set of cameras, we build a set of correspondences between points in the dense and sparse SfM models. For each feature in the chosen image subset, which already corresponds with a point in the sparse 3D from SfM, we cast a ray to find the corresponding point in the detailed model. At the end of this process, we obtain two sparse point clouds that are subsets of the two 3D geometries with known correspondences. We then find the global scale factor and a rigid alignment of these point-sets (i.e., rotation and translation) by applying a well-known absolute orientation algorithm [Hor87]. We implemented it in a robust RANSAC-based framework to remove possible outliers due to the non-complete overlapping among datasets [CMK03]. This affine transformation is then applied to the SfM model to approximately register the sparse geometry and all the cameras in the same reference frame of the dense model.

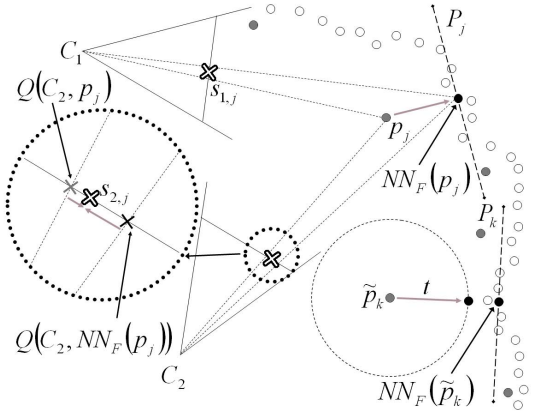

Figure 2: ***Fine registration.*** *A coarse registration between the original (white dots) and the SfM geometry (gray dots) is given. The fine registration jointly tunes camera parameters and sparse point positions to make the SfM geometry fit as much as possible the fine model; it minimizes the error between keypoints s and the re-projections of truncated nearest neighbor points $NN_F(p)$ in the dense model.*

7. Fine alignment

After the coarse alignment stage, we have a good initial configuration of the camera poses and their intrinsic parameters. However, these stages do not fully exploit the amount of accurate geometrical data present in the dense model. In fact, SfM reconstruction is completely independent from the fine 3D and it produces a list of camera parameters and 3D points consistent only in the domain of images. To improve the current registration, we should link this result with the dense geometry, putting some constraints on the sparse 3D points; more precisely, it is desirable that the SfM geometry fits as much as possible the fine model. To obtain a fully consistent model, we should formulate our fine alignment in a non-rigid manner, jointly moving the sparse points towards the dense 3D and accordingly tuning the parameters of each camera independently. We refine a camera model consisting in intrinsic parameters (i.e., focal length, principal point and the first two radial distortion coefficients) and extrinsic parameters (i.e., the rotation-translation map).

In our approach, as shown in Fig. 2, for each sparse 3D point p_j (gray dots), we compute the nearest neighbor $NN_F(p_j)$ (black dots) in the fine model F (white dots) and we find optimal camera parameters C and 3D points P, that minimize the following cost function:

$$E(C,P) = \sum_{j=1}^{N_P} \sum_{i=1}^{N_C} v_{ij} \left\| Q\left(C_i, NN_F\left(p_j\right)\right) - s_{i,j} \right\|^2 \quad (1)$$

where the term v_{ij} is a visibility factor, that is equal to 1 if the point p_j is visible in the image i, otherwise is 0, $s_{i,j}$ is the keypoint image coordinate (see Sec. 5) and $Q(C,p)$ is a function that projects a 3D point p into an image with parameters C. With our modified $Q(C,NN_F(p))$ that contains nearest points on the original model, we estimate a reprojection error in the image domain that strictly depends on the fine geometry, forcing the configuration of the cameras to be consistent with it. The big circle in Fig. 2 highlights how, by moving the point p_j and the camera C_2, we reduce the distance between $Q(C,NN_F(p))$ and $s_{2,j}$, as shown by the arrows. This is done jointly on all sparse points and all cameras. The refinement is formulated as a SBA, a non-linear least squares minimization problem on the 3D structure and viewing parameters.

To compute our cost function, we perform a point-to-surface distance computation, finding the point on the fine model/-surface nearest to each sparse point. If our input model is a triangle mesh, we find the projection of a point on the nearest triangle. In case of a point cloud input model, since a point-to-point metrics is not recommended due to errors caused by the original sampling, we need to find the surface that best approximate the point cloud. We use a fitting plane approximation, because the initial coarse alignment gives us enough confidence that the sparse point is close to the surface. To quickly perform this operation, we build a kd-tree populated with the points in the dense model. For each sparse point p_j we extract a fixed number of its neighbors (e.g., eight) and we find their interpolating plane P_j (Fig. 2). $NN_F(p_j)$ is the point on the plane nearest to p_j. If all sparse points are valid (i.e., inliers) the algorithm will produces a reliable registration. On the other hand, outliers (e.g., point $\tilde{p}_k$ in Fig. 2) are likely to result in fluctuating nearest point estimations on the surface or, in general, large re-projection errors. To reduce the contribution from outliers, we adopt a robust estimation function by truncating the searching distance to a maximum allowed value. In practice, if the distance between $\tilde{p}_k$ and the computed point $NN_F(\tilde{p}_k)$ is higher than a tolerance t, we force the nearest neighbor to the point on a sphere of radius t and center $\tilde{p}_k$ nearest to the plane P_k (Fig. 2).

Since the coarse alignment had produced a good initial estimate of the camera parameters, we can perform a local search over these parameters to find the final solution. To solve the minimization we chose the Levenberg-Marquardt algorithm [NW06]; thanks to its effective damping strategy, converges quickly from a wide range of initial guesses.

8. Results

Our technique was implemented on Linux using C++. The SfM software used for our tests is *Bundler* [SSS06, Sna]. For the minimization problem in the refinement step, we employ a C/C++ package for generic SBA based on the Levenberg-Marquardt algorithm, developed by Lourakis and Argyros [LA09]. The user interface to manually calibrate input photos is built using OpenGL and Qt tools. Our benchmarks were executed on a PC with a Dual-Core AMD Athlon II X2 3.1GHz CPU Processor, 4GB RAM, a 500GB 7200RPM Hard Disk and a NVidia GeForce GTX 460. To evaluate the effectiveness of our approach, in terms of computational time and quality, we present results on three CH datasets mapped with a large number of images: a *Grave* from a prehistoric necropolis and two parts of a *Church*, a roman basilica. The dense geometries are acquired with a time-of-flight laser scanner Leica ScanStation2 and the photos are captured with a Nikon D200 camera. Details on dataset sizes and processing times are listed in Table 1. In this paper, we focus on evaluating our method on CH test cases, comparing it with results obtained through manual alignment. As a future work, we plan to also numerically evaluate the results with respect to ground truth (e.g., using calibrated cameras and markers [SvHVG*08]).

Our method requires setting only one parameter: the global tolerance t for outlier removal. In this work, we automatically initialize this value to ten times the average sampling distance of the fine model. This works well for approximately uniformly sampled models. We are currently working on making the method more robust in the presence of variable sampling rate by locally adapting the tolerance to a multiple of the local sampling rate of the fine model.

Model	3D (# Points)	Images (wxh)	SfM (# Points)	Manual 2D/3D	Coarse	Fine	Total
Grave	8.3M	21 (1936x1296)	17m36s(19K)	3m/1 photo	4s	18m20s	40m
Church's Apse	14M	40 (1936x1296)	10m50s(7.6K)	7m/2 photos	6s	13m40s	32m
Church's Detail	4.7M	49 (1936x1296)	28m14s(17K)	4m/1 photo	8s	22m50s	55m

Table 1: ***Datasets and computational times.*** *We show the sizes of input geometry and image dataset, and the statistics about the time spent to compute the SfM, to manually align a small subset of photos, and to perform coarse and fine registrations.*

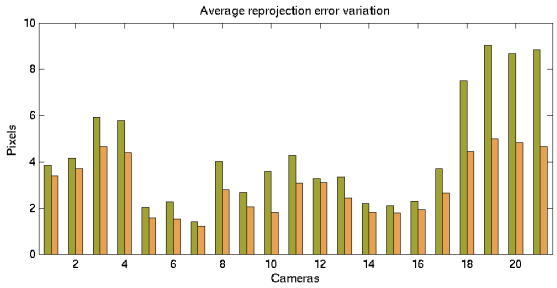

Figure 3: ***Re-projection error variation*** **(Grave).** *Each two bar set shows the per-camera average re-projection error variation before (left) and after (right) the refinement. Our algorithm improves the average error up to* 4 *pixels. Global average error over all images is decreased by about* 2 *pixels.*

Figure 4: ***Visual re-projection error variation*** **(Grave model).** *One image from the input dataset is presented. In the inset, corresponding with the white rectangle in the larger image, we show in red the detected SIFT points and in green the re-projection of 3D nearest neighbor points after the coarse alignment. The refinement step reduces the re-projection error, moving green pixels towards the SIFTs. Cyan pixels are the re-projections after the minimization.*

The *Grave* model consists in a dense geometry of 8*M* points and 21 photos. The fine geometry was acquired with 2*mm* sampling space. Using only information from images, the SfM routine takes less than 18 minutes to produce a sparse point cloud of 19*K* points (0.2% of the dense model). The user spent 3 minutes to manually register one single image (12 2D/3D correspondences selected); this time is comparable to per-image manual alignment times presented in Franken et al. [FDG*05]. We save 95% of the user intervention, drastically reducing the manual operation. Coarse and fine registrations respectively take 4 seconds and about 18 minutes. The total time, including SfM computation, is 39 minutes. Before and after the fine registration, we estimate the re-projection error both for each single camera and globally for the entire image dataset. Figure 3 shows for each photo the variation of the average re-projection error in pixel units; our refinement strategy reduces the error by up to 4 pixels. The global average re-projection error over all the images goes from 5.8 to 3.6 pixels. Figure 4 shows one of the acquired photos and, in a close view, how the refinement step moves re-projection of 3D nearest neighbor points, obtained after the coarse alignment (green dots), towards the detected SIFT points (red dots). Cyan pixels are the re-projections after the minimization. Figure 5 shows how the output of the proposed method is used in a texture blending framework [PGC11] to obtain globally coherent colored models.

The fine geometry of the *Church's Apse* model, acquired at sub-centimetric resolution, has 14*M* points, and 40 associated photographs. The SfM data is computed in 11 minutes and contains 7.6*K* points (i.e., 0.05% of the dense model). Generally, the user aligns the first image and, since the coarse alignment takes only few seconds, it can quickly check whether the result is reasonable as a good initial guess for the local refinement or he has to register another photo. In this case, he decided to select 2D/3D correspondences of another image before launching the minimization routine. The manual operation takes 7 minutes, while the algorithm computes fine registration in 14 minutes. The entire registration, included SfM step, takes about 32 minutes. With our technique, we produce a robust result saving the 90% of the estimated time of the manual pipeline. The final global re-projection error is 1.7 pixels. The colored model after applying blending algorithms is shown in Figure. 6.

The last test is performed on another detail of the *Church* model. The fine geometry contains 4.7*M* points, while the SfM point set has 17*K* points; we have acquired 49 photos. Here, the motivation to show this dataset is that its SfM geometry contains a lot of outliers (Figure 7), due to the complexity of the input image set (e.g., reflections on mirrors and specular highlights), or belonging to object parts

Figure 5: ***Grave.*** *Colored model obtained using the output of our method (camera parameters) in a photo blending framework.*

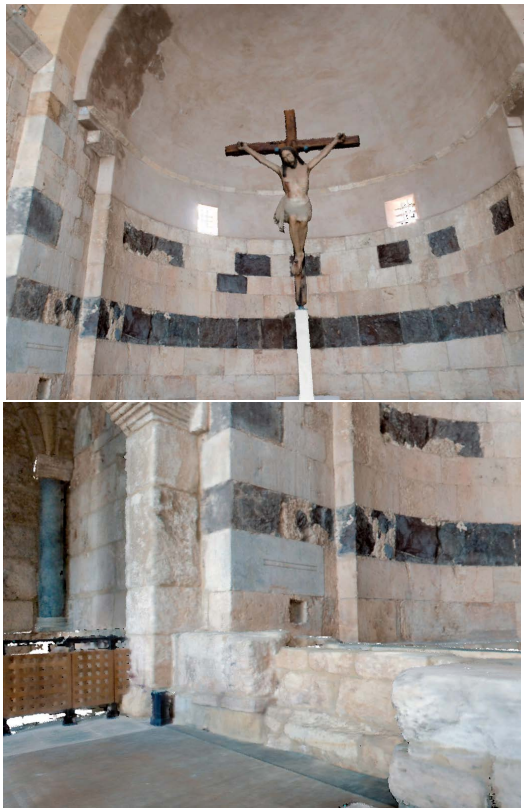

Figure 6: ***Church's Apse.*** *Colored model obtained using the output of our method in a photo blending framework.*

not acquired with the time-of-light scanner. In other words, there are a big number of SfM points that do not belong to the dense model and/or to the real geometry. The user aligns only one single image and after the fine registration and photo blending we obtain the colored model in Figure 8. The global average re-projection error is 1.2 pixels. Thus, in these non-ideal cases our method is very robust to outliers, without increasing the user intervention too.

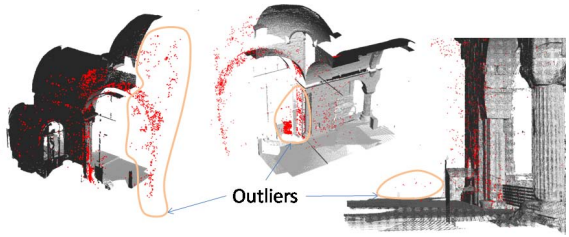

Figure 7: ***Outliers.*** *Registration of sparse (red) and dense (white) geometries of a* Church's Detail. *SfM geometry contains a lot of outliers. Our method proves to be robust in this non-ideal case, without requiring more user intervention.*

9. Conclusions and future work

We have presented an efficient, fast and robust technique for registering a set of images with a 3D geometry. Our semi-automatic approach minimizes the user intervention and is generally applicable to different kinds of 3D models. The quality and reliability of the method is demonstrated on a series of real-world Cultural Heritage 3D/2D datasets. It proves to be robust in the presence of input data with big amount of outliers, and produces a good input data for photo blending framework. The natural next challenging step in future works should be the development of a reliable way to solve the sparse-to-fine geometry alignment in an automatic manner, to completely avoid manual intervention.

Acknowledgments. This research is partially supported by EU FP7 grants 242341 (INDIGO) and by Sardegna DISTRICT (P.O.R. Sardegna 2000-2006 Misura 3.13).

References

[AMCO08] AIGER D., MITRA N. J., COHEN-OR D.: 4-points congruent sets for robust surface registration. *ACM ToG 27*, 3 (2008).

[AMN*98] ARYA S., MOUNT D. M., NETANYAHU N. S., SILVERMAN R., WU A. Y.: An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J. ACM 45*, 6 (1998), 891–923.

Figure 8: ***A Church's Detail.*** *Colored model obtained using the output of our method in a photo blending framework.*

[BCPP98] BURKARD R. E., CELA E., PARDALOS P. M., PITSOULIS L. S.: The quadratic assignment problem, 1998.

[BPB*09] BORGEAT L., POIRIER G., BERALDIN J.-A., GODIN G., MASSICOTTE P., PICARD M.: A framework for the registration of color images with 3d models. In *ICIP* (2009), pp. 69–72.

[CDPS09] CORSINI M., DELLEPIANE M., PONCHIO F., SCOPIGNO R.: Image-to-geometry registration: a mutual information method exploiting illumination-related geometric properties. *Comput. Graph. Forum 28*, 7 (2009), 1755–1764.

[CMK03] CHUM O., MATAS J., KITTLER J.: Locally optimized ransac. In *DAGM-Symposium* (2003), pp. 236–243.

[FB81] FISCHLER M. A., BOLLES R. C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24*, 6 (1981), 381–395.

[FDG*05] FRANKEN T., DELLEPIANE M., GANOVELLI F., CIGNONI P., MONTANI C., SCOPIGNO R.: Minimizing user intervention in registering 2d images to 3d models. *The Visual Computer 21*, 8-10 (2005), 619–628.

[GFF10] GHERARDI R., FARENZENA M., FUSIELLO A.: Improving the efficiency of hierarchical structure-and-motion. In *CVPR* (2010), pp. 1594–1600.

[HL06] HANTAK C., LASTRA A.: Metrics and optimization techniques for registration of color to laser range scans. In *3DPVT* (2006), pp. 551–558.

[Hor87] HORN B. K. P.: Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A 4*, 4 (1987).

[IOT*07] IKEUCHI K., OISHI T., TAKAMATSU J., SAGAWA R., NAKAZAWA A., KURAZUME R., NISHINO K., KAMAKURA M., OKAMOTO Y.: The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects. *IJCV 75*, 1 (2007), 189–208.

[KSSS09] KAMINSKY R., SNAVELY N., SEITZ S., SZELISKI R.: Alignment of 3d point clouds to overhead images. pp. 63 –70.

[LA09] LOURAKIS M. A., ARGYROS A.: SBA: A Software Package for Generic Sparse Bundle Adjustment. *ACM Trans. Math. Software 36*, 1 (2009), 1–30.

[LHS00] LENSCH H. P. A., HEIDRICH W., SEIDEL H.-P.: Automated texture registration and stitching for real world models. In *Pacific Graphics* (2000), pp. 317–.

[LL09] LI Y., LOW K.-L.: Automatic registration of color images to 3d geometry. In *CGI* (2009), pp. 21–28.

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *IJCV 60*, 2 (2004), 91–110.

[LS05] LIU L., STAMOS I.: Automatic 3d to 2d registration for the photorealistic rendering of urban scenes. In *CVPR (2)* (2005).

[LYWZ06] LIU L., YU G., WOLBERG G., ZOKAI S.: Multiview geometry for texture mapping 2d images onto 3d range data. In *CVPR (2)* (2006), pp. 2293–2300.

[MKI09] MASTIN A., KEPNER J., III J. W. F.: Automatic registration of lidar and optical images of urban scenes. In *CVPR* (2009), pp. 2639–2646.

[NW06] NOCEDAL J., WRIGHT S.: *Numerical Optimization*, 2nd ed. Springer, July 2006.

[PGC11] PINTUS R., GOBBETTI E., CALLIERI M.: A streaming framework for seamless detailed photo blending on massive point clouds. In *Eurographics Areas Papers* (2011), pp. 25–32.

[SA01] STAMOS I., ALLEN P. K.: Automatic registration of 2-d with 3-d imagery in urban environments. In *ICCV* (2001).

[SLC*08] STAMOS I., LIU L., CHEN C., WOLBERG G., YU G., ZOKAI S.: Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes. *International Journal of Computer Vision 78*, 2-3 (2008).

[Sna] SNAVELY N.: http://phototour.cs.washington.edu/bundler/.

[SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: exploring photo collections in 3d. *ACM Trans. Graph. 25*, 3 (2006), 835–846.

[SSS08] SNAVELY N., SEITZ S. M., SZELISKI R.: Modeling the world from internet photo collections. *IJCV 80*, 2 (2008).

[SSS10] SINHA S. N., STEEDLY D., SZELISKI R.: A multi-stage linear approach to structure from motion. In *ECCV 2010 Workshop on Reconstruction and Modeling of Large-Scale 3D Virtual Environments* (2010).

[SvHVG*08] STRECHA C., VON HANSEN W., VAN GOOL L., FUA P., THOENNESSEN U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Proc. CVPR* (2008), pp. 1–8.

[VI97] VIOLA P. A., III W. M. W.: Alignment by maximization of mutual information. *IJCV 24*, 2 (1997), 137–154.

[WLH*04] WILLIAMS N., LOW K.-L., HANTAK C., POLLEFEYS M., LASTRA A.: Automatic image alignment for 3d environment modeling. In *SIBGRAPI* (2004), pp. 388–395.

[ZNH04] ZHAO W.-Y., NISTÉR D., HSU S. C.: Alignment of continuous video onto 3d point clouds. In *CVPR (2)* (2004).