A FRAMEWORK FOR THE REGISTRATION OF COLOR IMAGES WITH 3D MODELS

Borgeat, Louis; Poirier, Guillaume; Beraldin, Angelo; Godin, Guy; Massicotte, Philippe; Picard, Michel

Publisher’s version / Version de l'éditeur:

Image Processing (ICIP) Proceedings, pp. 69-72, 2009-11-10

A FRAMEWORK FOR THE REGISTRATION OF COLOR IMAGES WITH 3D MODELS

Louis Borgeat, Guillaume Poirier, Angelo Beraldin, Guy Godin, Philippe Massicotte and Michel Picard

National Research Council of Canada

ABSTRACT

This paper describes an environment to automatically or semi-automatically compute the precise mapping between a set of 2D images and a triangulated 3D model built from high-resolution 3D range data. This environment is part of our Atelier3D framework for the modeling, visualization and analysis of large sensor-based datasets. This work was done to initially support three cultural heritage application projects: the modeling of the Grotta dei Cervi in Italy, of the Erechtheion in Athens, Greece, and of Leonardo's Mona Lisa. The proposed method combines image-based registration, feature matching, robust estimation techniques and advanced multi-resolution rendering with a powerful user interface.

***Index Terms*—** Image registration, cultural heritage, large datasets, 3D sensors, texture mapping

1. INTRODUCTION

As sensor technology improves, it is now feasible to acquire vast amounts of 2D and 3D data from cultural sites in a very short time. Modern sensors can acquire billions of 3D samples and terabytes of pixel data in a matter of hours. However, the process of transforming all the raw data into one accurate model can still be very time and resource consuming and can quickly become a major project bottleneck. In this paper we address the issue of having to precisely map a large number of high-resolution digital photographs onto a 3D model in order to produce a seamless integrated color model for visualisation and analysis. The described processing is part of the steps that prepare the data for visualization within the Atelier3D framework [1]. This frameworks aims at providing a complete set of tools for the efficient and acurate modeling, visualisation and analysis of large sensor-based 3D datasets. Atelier3D already permits visualisation and analysis of large models on desktop hardware. We are now in the process of integrating modeling components within this environement.

Numerous techniques have been proposed to address the problem of automatic registration between 3D and 2D images. Hantak[2] provides a good litterature review and classification of automatic registration techniques and compares different image-based similarity metrics for 3D/2D intensity registration. Typically, proposed methods iteratively use image-based registration followed by a new pose estimation[3][4] until convergence. Variations involve feature or edge detection in the images and direct intensity comparison using different metrics and optimizers. Little work presents results for larger image databases, which are in practice still often processed using highly interactive software relying on manual selection of feature points between images.

What we propose in this paper is a set of tools that enable an application user to process rapidly large amounts of texture data for a broad range of application contexts. Indeed, the problem of 2D/3D registration can vary significantly between types of practical applications. In some cases the intensity image obtained from the 3D data will be very similar to the corresponding 2D image and in other cases quite different and therefore harder to register; many older range sensor do not even provide such intensity information. If the 2D and 3D sensors are physically attached, we get a good initial pose and camera calibration, but in other cases one must work only from randomly taken photographs. No single algorithmic solution can currently cover all acquisition setups and physical environments. What we propose is a framework to tackle most situations with minimal user intervention. We will first describe an interactive graphical user environment to rapidly produce a good initial alignment/calibration and to validate results. We then describe a set of automatic tools to iteratively refine an initial estimate until sufficient accuracy is achieved. We finally present results for a few application projects. Here, our main original contribution lies more in the proposed system for large dataset processing than in a new specific individual algorithm. As such, we have tested it for different practical applications in the heritage field characterized by 3D datasets composed of hundreds of millions of triangles and large texture databases.

2. IMAGE REGISTRATION

The interactive registration process that leads to a first pose and camera calibration goes as follows (some steps can be skipped or repeated depending on the quality of the initial alignment):

First, the user aligns a semi-transparent version of the 3D model with the image using mouse-based navigation. The

Special thanks to Prof. V. Valzano and A. Bandiera, Coordinamento SIBA, Universit del Salento, Lecce, Italia.

Fig. 1. A typical image of the Erechtheion: Distinct background and foreground causing big differences under pose rotation, strong cast shadows, significant differences in intensity. *Top*: 2D image (triangle SIFT points). *Middle*: 3D rendering (round SIFT points). *Bottom*: typical image difference for a small section after mouse (left), quick SIFT (middle), and image based (right) phases.

rendering can be done using a typical pinhole OpenGL model built based on the camera pose and field of view. If the camera is already partially calibrated, the interactive rendering can also include a GPU shader that renders the 3D model by taking into account intrinsic lens parameters, including distortion, thus allowing for a more realistic manual alignment. The actual focal length is available from the EXIF tags in the image file and the camera CCD information is extracted from a database of known cameras. The performance of all phases of the processing is not affected by the size of the 3D model since this is integrated in the Atelier3D framework and that the model used in all steps is multi-resolution[5].

A GPU implementation of the SIFT algorithm[6] can then produce on demand a set of candidate matches between the 3D model and the image rendered as positioned by the user (Fig. 1). The user can then manually select pairs of corresponding points in the SIFT set or select new pairs using the interface to improve the alignment. At any time, a new calibration based on the selected matches can be recomputed on request. This is done by running a Levenberg-Marquardt least-squares optimizer on the point pairs to produce a full intrinsic and extrinsic camera calibration or just a new pose, depending on user selection. Projection errors for the points are provided in conjunction with the new rendering for rapid evaluation. The user can finally run the automatic alignment procedure at any point to attempt to refine the current alignment or iterate in the interactive process.

All the optimization steps that are part of the interactive process run in negligible time from the user's perspective, hence maintaining full interactivity. The automatic processing can take up to 30 minutes for a very misaligned high resolution image and a very large dataset in order to perform multiple image-based/recalibration iterations. Running it for a few iterations will allow to correct for larger deformation that the image based algorithm cannot model in a single pass. This process is obviously easily parallelizable for each image and can be done in the background in a practical context.

Fig. 2. Intensity correction for two merged 3D sensor viewpoints of the Erechtheion.

We normally use the intensity of the reflected laser beam directly as the color for the rendered surface, even if the user can choose to also add some lighting or filtering to highlight the shape. In order to make it easier to align the two images, this intensity value is corrected based on the distance to the surface and the angle at which the surface was measured, or using a more advanced technique if more is known on the inner physics and processing of the sensor, or on the reflectivity of the materials in the image. Figure 2 shows such an intensity image with the corresponding digital photograph for a part of the Erechtheion dataset.

For the automatic phase of the alignment, we use a multistep hybrid 2D/3D procedure. The 2D part builds on the image processing tools provided in the ITK[7] Toolkit. For the 3D, we re-use the Levenberg-Marquardt least-squares optimizer used in the interactive phase to extract a calibration from a given image alignment transformation. In the first step of the process, the 3D model is rendered based on default (laser intensity only) or adapted lighting specifications from the user. More complex GPU shaders can be used to perform segmentation of feature-based matching instead of working on the original rendering and photograph. The automated pipeline first determines a rigid alignment between the 2D image and the rendering. A large sample set of correspondence points is taken at random between the images, and an new calibration is computed from those points. The image is re-rendered based on that calibration, and two or or more passes of that process are applied, but this time with an affine 2D transform followed by a deformable transform. Image comparison is done using an expectation maximization variation[8] that combines good adaptability for comparison of images of different modalities and a relatively linear search space to avoid falling into local minima, an issue for this kind of application when using this metric[2]. By re-rendering after each step, we insure that the rendered part of the 3D model really corresponds to the part visible in the image,but also compensates for limitations in the deformation model used by the registration algorithm. Masks are computed from the 3D rendering and used by the image-based algorithms to align only the part of the rendering where data is visible and ignore the background color.

The 3D model will also contain errors which cannot be compensated only by the camera pose and calibration. Errors will vary depending on the model and principle for the sensor, but also due to registration and integration error between different sensor viewpoints. Therefore, we add a final deformation to the photograph after the last calibration phase to account for spatial errors in the 3D model itself and for residual image deformations not included in the calibration model. Image data can then be texture mapped onto the model or sampled to produce color-per-vertex data[5].

3. RESULTS

We have experimented with the proposed algorithms within three different application projects, all characterized by different sensors and acquisition setups. The first one is the Grotta dei Cervi in Italy[9](Fig. 3). In that case, the cave walls where scanned using a single wavelength laser triangulation prototype system coupled with a rigidly attached high resolution digital camera. Therefore, the initial camera calibration and pose estimation were relatively good. The intensity returned by the sensor was also very similar to the images. In that case the automatic processing directly provided pixel-level accuracy for all 231 images of size 3504x2336, and allowed for seamless connections between textures. SIFT also provided us directly with an adequate set of matches, even if we voluntarily did not use the initial pose information but opted for a trackball manual alignment instead. In the case of the Erechtheion[10], a monument in Athens, the sensor was a commercial time-of-flight phase-based system and the photographs were taken separately by a professional photographer, a more challenging case with no initial pose estimation, noisier and less complete 3D data, more complex structures and cast shadows. The 3D model on which the data was applied is composed of 350 million polygons. For the first 200 images, alignment followed the manual procedure, with mouse alignment being sufficient to seed the automatic process in some cases, but requiring selecting a few points to improve the initial pose in others. A frequent case for those images was that the SIFT algorithm would select pairs that were concentrated in too small an area of the image, and the user would need to select 1 or 2 extra pairs to get a good calibration to start the automated process. Even in this more difficult context, the set of available tools did allow an expert operator to accelerate the processing of data by at least an order of magnitude. Figure 1 shows typical results for an image. The project is still incomplete since we are currently working

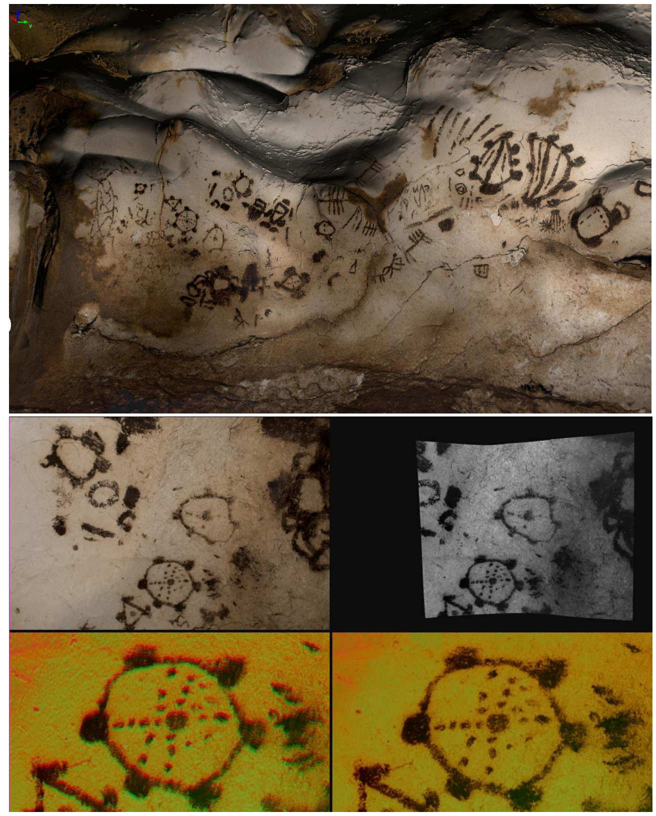

Fig. 3. *Top*: view of a texture mapped model of the Grotta dei Cervi, 231 high-res textures are merged in the complete model. *Middle*: photo and corresponding section of the 3D model. *Bottom*: alignment error before and after automatic processing.

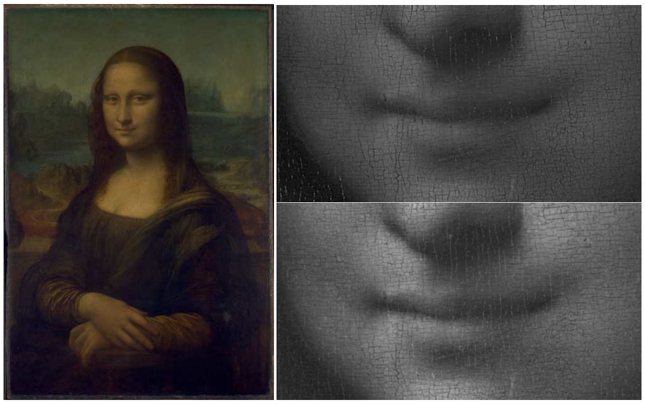

Fig. 4. *Left*: color composed from multi-spectral data. *Right*: comparable luminance images made from laser(bottom) and multi-spectral(top) data after registration.

on on other aspects of the image processing, such as shadow removal and other image correction issues outside the scope of this paper. Finally for the case of the Mona Lisa in Fig.4, the range sensor for the project [1] uses 3 laser wavelengths to help produce a color reflectance estimate for the measured surface. As part of the same project[11], a 13-band multi-spectral camera was also deployed to image the Mona Lisa, providing a more precise color measurement. We have used the same image-based algorithms to align the color from the 3D sensor with the multi-spectral image, by using the channels from the multi-spectral image that corresponded to the laser wavelengths, we produced two comparable luminance images for alignment. However, we needed to render the 3D model in multiple tiled sections, since the size of the multi-spectral image(7854x11498) is beyond the maximum size of a rendering buffer even of the most recent GPU. This process produced two very similar images and allowed to easily compute the registration between them, which will allow us to improve the resulting dataset by removing specularities and ambiguities present in each individual image.

4. CONCLUSION

We have presented a framework that allows a user to rapidly compute the precise alignment between a large set of high resolution images and a large high resolution 3D model by combining minimal user interaction with a set of state-of-the-art registration and optimization algorithms. We have shown that the system can be used efficiently for a wide variety of application contexts. We are now investigating appropriate metrics to evaluate quantitatively the result of such a process, which is a challenge due in part to the multi-modal nature and wide variety of data sources to be covered by the framework, to the presence of user interaction, and to the lack of available ground truth for datasets of that size. We are also working on improving complementary components such as color correction and image fusion and shadow removal.

Part of this work is part of the project Development of GIS at the Acropolis of Athens, financed by the EU and the Government of Greece, and supervised by the Acropolis Restoration Service, Hellenic Ministry of Culture. The partners in this project are Elpho, Athens, Geotech O.E., Athens, ETH, Zurich, NRC, Canada, Institute for Mediterranean Studies, FORTH, Rethymno, Crete, with external cooperators, Leica Geosystems, Switzerland and Basis Software Inc., USA.

5. REFERENCES

[1] L. Borgeat, G. Godin, P. Massicotte, G. Poirier, F. Blais, and J-A. Beraldin, "Visualizing and analyzing the Mona Lisa," *IEEE Computer Graphics and Applications*, vol. 27, no. 6, pp. 60–68, 2007.

[2] C. Hantak and A. Lastra, "Metrics and optimization techniques for registration of color to laser range scans," in *3DPVT '06*, Washington, DC, USA, 2006, pp. 551–558, IEEE Computer Society.

[3] K. Umeda, G. Godin, and M. Rioux, "Registration of range and color images using gradient constraints and range intensity images," in *ICPR '04*, Washington, DC, USA, 2004, pp. 12–15, IEEE Computer Society.

[4] P. Dias, V Sequeira, J.G.M. Gonçalves, and F. Vaz, "Automatic registration of laser reflectance and colour intensity images for 3D reconstruction," *Robotics and Autonomous Systems*, vol. 39, no. 3, pp. 157–168, June 2002.

[5] L. Borgeat, G. Godin, F. Blais, and C. Lahanier, "GoLD: Interactive display of huge colored and textured models," in *Proc. of SIGGRAPH 2005*, Los Angeles, California, Jul. 31 - Aug 4 2005.

[6] C. Wu, "SiftGPU: A GPU implementation of scale invariant feature transform (SIFT)," December 2008, http://www.cs.unc.edu/ ccwu/siftgpu.

[7] "The Insight Toolkit," http://www.itk.org.

[8] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, "Nonrigid multimodality image registration," in *SPIE Conference Series*, M. Sonka and K. M. Hanson, Eds., July 2001, vol. 4322, pp. 1609–1620.

[9] J-A. Beraldin, F. Blais, L. Cournoyer, M. Picard, D. Gamache, V. Valzano, A. Bandiera, and M. Gorgoglione, "Multi-resolution digital 3D imaging system applied to the recording of grotto sites: the case of the Grotta dei Cervi," in *VAST*, October 30 - November 4 2006.

[10] S. El-Hakim, J-A. Beraldin amd M. Picard, and L. Cournoyer, "Surface reconstruction of large complex structures from mixed range data - the Erechtheion experience," in *ISPRS*, July 2008.

[11] Jean-Pierre Mohen, Michel Menu, and Bruno Mottin, *Mona Lisa: Inside the Painting*, Harry N. Abrams, 2006.