EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2014)
R. Klein and P. Santos (Editors)

Site-specific art and 3D: an example of spatial analysis and reconstruction

Matteo Dellepiane[1] and Mara De Matteis[1]

[1]Visual Computing Laboratory - ISTI-CNR, Italy

Abstract

Site-specific art is a concept that goes back to the beginning of human race: the works of art were often created by artists taking into account not only their shape and appearance, but also the context in which they would be put. For this reason, moving the artifacts from its original placement (or the changes which happen around it) tend to decrease its impact, and possibly weaken its potentials. Site-specific art is a very powerful concept also for contemporary artists. This paper focuses on the analysis of L.O.V.E., a sculpture from the controversial artist Maurizio Cattelan. Cattelan donated the sculpture to Milano, under the condition that it should not be moved from its original place (in front of Milano Stock Exchange). The aim of the paper is to use 3D reconstruction techniques to show and analyse the monument, stressing its relation with the context around it. A multi-view stereo matching campaign was perfomed to have an accurate reconstruction of the context, then the photos provided by the community were integrated in the reconstruction to show the "point of view" of the people. These data provide interesting indications about the aims of the authors, and they provide additional material for the interpretation of the work of art.

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

1. Introduction

The advent of multi-view stereo matching techniques opened new, very interesting perspectives in the context of 3D acquisition. The main advantages coming from this new set of methods come from their low cost (even the camera of a smartphone can be used) and from the availability of software tools to easily process and visualize data.
While these techniques are mainly used starting from an ad-hoc set of images, acquired following dedicated paths and strategies, the seminal works in the literature had the goal to be able to handle images taken from the community collections (i.e. Flickr, Google). This posed stronger challenges from a technical point of view (limited information about extrinsics of the camera, no assumptions on the presence of the object of interest in the image), but it also opened other interesting questions like: can the community really contribute in the 3D reconstruction of our Heritage? Which other indications can we find in the behavior of the community when taking images (the so-called "wisdom of the crowd")?
This paper builds on the above questions by presenting a project where multi-view stereo matching was used to reconstruct and analyze a famous example of site-specific art [Kwo04]: the statue called "L.O.V.E." (acronym for the Italian words for Love, Hate, Revenge, Eternity), by Alessandro Cattelan (Figure 1).
This controversial work of art represents a hand with all the fingers cut, except for the middle one, hence recalling a common insulting gesture in Italy. The statue was placed in front of Palazzo Mezzanotte. The building hosts the Italian Stock Exchange, the symbol of the economic power.
The artwork was first shown in September 2010, with the aim of being displayed for a period of only two weeks. The author decided to donate the artwork to the city, with the only requirement that it should not be moved from its place. This originated an articulated discussion about the opportunity of keeping the statue, until 2012, when it was decided to leave it in its original place.
The project involved two phases: in the first one a dedicated photographic campaign was performed to acquire the geometric structure not only of the statue, but also of its context. This ended up in the creation of an accurate 3D model of the statue, and of a point cloud of the architectural space around it.

DOI: 10.2312/gch.20141301

Figure 1: *L.O.V.E. by Maurizio Cattelan, Piazza degli Affari, Milano.*

In a second stage, a group of images related to the statue were downloaded from a community photo collection. They were mapped on the 3D reconstruction of the square, in order to provide a visual representation of the "point of view of people". This could help in the interpretation of the meaning of the statue, and more importantly to the analysis of its site-specific role.

2. Related work

Multi-view stereo reconstruction methods provide means to reconstruct the spatial structure of real scenes starting from a set of un-calibrated images. The proposed methods essentially build on the idea of stereo reconstruction, by relaxing strong assumptions like the continuity of data flow (needed, for example, in the Simultaneous Localization And Mapping [LDWov, DRMSne, KM09] approaches). These methods essentially rely on a combination of Structure from Motion [SSS06] and Multi View Stereo Reconstruction. The reconstruction procedure is usually divided in three main steps: Features recognition and Matching [Low04, SmFPG06], Camera Calibration and Bundle Adjustment [TMHF00, LA09, WACS11], and Multi-view dense reconstruction [GSC*07, FP10, FCS*10]. The previous pipeline has become a standard for Multi-view Stereo Reconstruction, and it was implemented in several versions, structured as both webservices and closed systems, where the user must provide all the images which were acquired, and the interaction is limited to the possibility to tune the parameters of the various steps of the reconstruction.
All the above methods were implemented to be extremely robust to variability in the input data: the images were usually obtained directly from the community photo collections. This has a very interesting potential, and it's certainly applicable for important sites [AFS*11, FHZ*13]. Unfortunately, not all the sites could provide enough coverage, while it could be possible to complete missing parts [LM11]. For this reason, multi-view stereo matching has been adapted to be used also in more acquisition-oriented photographic campaigns, for example by implementing a competitive game [TSH*11], or guiding the user in the acquisition process [DCCS13].
Recent developments [TKM*13] are moving in the direction of a more on-site acquisition specific applications, where the user obtains a feedback about the coverage and reconstruction in real time.
Regarding the Cultural Heritage field, multi-view stereo matching has become a very interesting alternative to 3D scanning, especially due to its much lower costs. While the quality of data cannot be compared with the one of active scanning devices, multi-view stereo matching is already largely used in contexts where mainly an interpretation of the data is needed [DDUC*13]. Another interesting option is the use of this technique for the reconstruction of partially or completely destroyed objects using historical data [GRZ03].
The presented paper explores a different perspective of multi-view stereo matching, which is mainly related to the study of the spaces around a work of art, and to its impact in the perception of the public.

3. Methodology

The original goal of the project was to be able to reconstruct the statue starting only from the data downloaded from the community photo collections: for this purpose several tens of images were retrieved from the web, and an initial reconstruction was tried. Nevertheless, it was clear from the beginning that a proper reconstruction of the statue would have been impossible. This was not due to the number of images, but to the fact that most of them were taken from the same point of view, so that a complete coverage of the statue was not possible. This is quite common when an heritage site has to be reconstructed from images: except for the super-famous monuments (like the ones shown in the state-of-the-art literature), it could be hard to obtain a full coverage.
This could be related to a couple of factors: the spatial environment, that sometimes prevents from framing the object from some points of view, and the tendency of the public to conform to the stereotypical representation of the monument (i.e. the photo of the tourist holding the Tower of Pisa). Actually, this already indicated that there was a preferred point of view from the public, but in order to have a more useful visualization, it was decided to perform an on-site acquisition to obtain accurate three-dimensional information.

Hence, a total of 70 images were acquired for that purpose: a subset of them is shown in Figure 2. Images were processed using Photosynth Toolkit (http://www.visual-experiments.com/2010/08/19/my-photosynth-toolkit/). This is a set of scripts that makes use of the Photosynth [SSS06] webservice for the matching and calibration of images, and of PMVS [FP10] to locally perform dense recon-

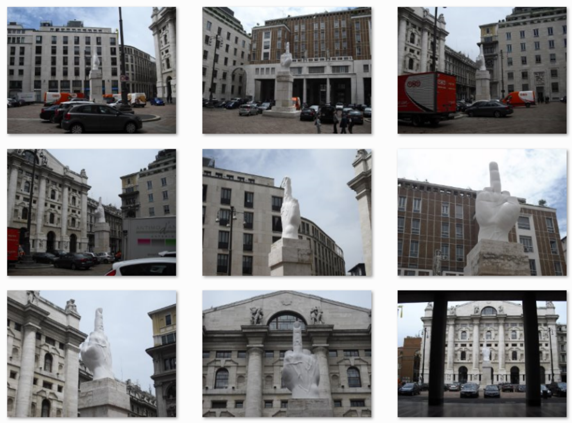

Figure 2: *Some of the images acquired for the site reconstruction*

struction. Alternative (both free and commercial) tools could have been used in this case, but the chosen one was considered robust enough to provide accurate results. Moreover, the project was developed in the context of a University course [DS12] where freeware and non hardware-demanding solutions are proposed to students.
The result of the pipeline was a dense description not only of the statue, but also of the surrounding environment, for a total of nearly 1.6M reconstructed points: 400k of them represented the statue. Figure 3 shows the point cloud generated by dense reconstruction. The scene was scaled at a real unity of measure (meters) using the size of the basement of the statue as a reference.

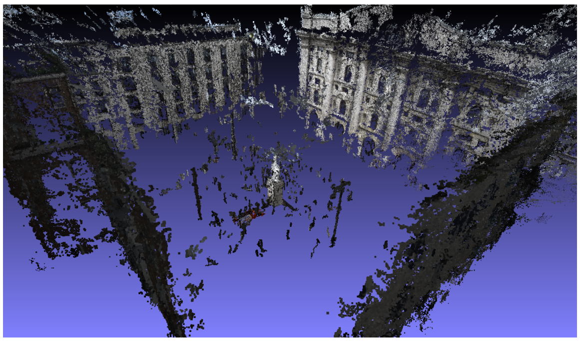

Figure 3: *A snapshot of the data obtained after multi-view stereo reconstruction*

The statue data were extracted and processed using the MeshLab [CCC*08] tool. The processing pipeline was the usual one with this type of data: an initial cleaning of the point cloud, the generation of a triangulated surface through Poisson surface reconstruction [KBH06], a final cleaning of the model. The surface reconstruction filled the missing upper part of the statue in a plausible way (see Figure 4), generating an almost complete and accurate reconstruction, mad of nearly 500k triangles.

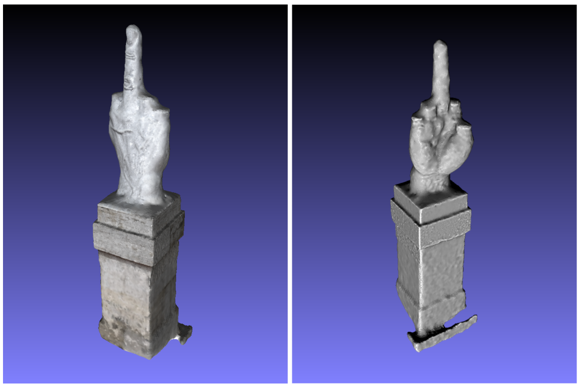

Figure 4: *Two snapshots of the triangulated model of the statue. Left: with color. Right: pure geometry.*

Finally, the statue model was integrated in the cleaned reconstruction of the square, providing a representation of the environment (see Figure 5).

Figure 5: *A snapshot of the reconstructed environment.*

When the reference three-dimensional space was ready, a set of images were retrieved from the community collection. The retrieval was obtained using a set of keywords (*L.O.V.E., Piazza Affari, Cattelan, Borsa, Dito medio, stock exchange, Milano*) related to the work of art, and accessing to several community collections like Flickr and Google Images. As already mentioned, the set of images (52 in total, see Figure 6 for a few examples) that remained after removing the ones not depicting the statue mainly showed only a portion of the environment. Moreover, as usual, these images include nocturnal and heavily processed elements.

Nevertheless, the multi-view stereo reconstruction was applied on this dataset, obtaining a reconstruction of the scene which was limited to the front part of the statue

Figure 6: *Some of the images downloaded from community collections.*

(and the facade of the Stock Exchange). Part of the images (mainly the ones showing the rear part of the statue) taken from different points of view were not calibrated with the rest (see Figure 7).
The next step was to integrate the two reconstructions in a single environment. Using MeshLab, the reconstruction from the community photo collections was aligned with the description of Piazza Affari, using the Alignment Tool. All the images were aligned in the same reference system by applying the same roto-translation (and scaling) that brought the two point clouds in the same reference system. The remaining images, which were not included in the reconstruction, were added using MeshLab through mutual information [CDPS09] or by hand.

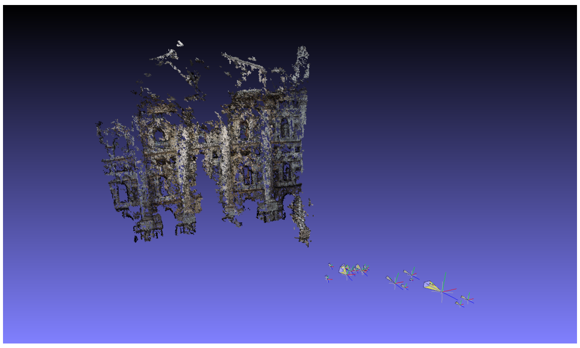

Figure 7: *The reconstructed area from community collections, with the corresponding images positions.*

4. Results

After the alignment of all the images on the reference scene, it was possible to analyze the position of each image. Figure 8 shows the position of the cameras associated to each image which was aligned. Forty-two images out of fifty-two (nearly 80%) were taken from a frontal position, showing that the aim of the author, the fact that the statue should have been always shown with the stock exchange building as a background, was obtained. Figure 9 shows two of the images aligned on the reconstruction: the top one was automatically calibrated with the others, the bottom one had to be aligned by hand.

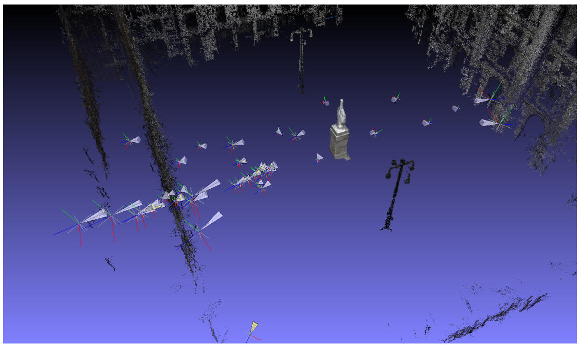

Figure 8: *A snapshot of the position of all the images aligned on the reference scene.*

This shows how this site-specific work of art is able to drive the perception of the public only if kept in its original position. Its position could be also seen as a constraint for everyone: the position of the statue prevents from taking a picture of the facade of the building without having it on the frame. The reconstructed environment was used to analyze the areas where it was possible to frame the facade of the building without framing the statue.
A simple algorithm found the points of view (given a camera with a defined field of view, and after defining a height from the ground compatible with a person size) where it was possible to frame all the points describing the facade (red rectangle in Figure 10) but not including the statue (green rectangle in Figure 10). The analysis showed that it was not possible to obtain this with normal lenses in any part of the square, while only in a small portion of it (yellow areas) it was possible to get the coverage using a wide-angle lens. The only way to frame the building by staying between it and the statue is with the use of fish-eye lenses.
This could suggest a different point of view with respect to this work of art: it's not only influencing the perception of a urban space, but it's changing it by introducing a new element which is almost impossible to remove from any point of view.

Figure 9: *Two images taken from community collections, aligned with the reference point cloud.*

Figure 10: *A view from top of the square: the yellow zone indicate where the facade of the stock exchange can be framed with wide-angle lenses without including the statue.*

5. Conclusions

This paper presented the study of the role of a site-specific work of art using 3D reconstruction techniques. The chosen artifact (L.O.V.E. by Maurizio Cattelan) represents a controversial example, that generated long discussions not only for its specific shape, but also for its specific placement. In this case, the reconstruction of the three-dimensional shape not only of the statue, but also of the surrounding environment, can be useful to better appreciate the role of modern art in the urban spaces.
An ad-hoc photographic campaign was performed, and multi-view stereo matching techniques were applied to obtain an accurate description of the square. Then, several images related to the work of art were downloaded from the community collections on the web. The images were then aligned on the 3D model, in order to show the "point of view of the public".
Low cost three-dimensional reconstruction methods can be a valuable tool for the analysis of the impact of modern art in the context of a urban environment. Moreover, they can be used also for the planning step: for example, the placement of a work of art (or of any other element, like an advertisement sign) could be "tested" in advance. An accurate 3D reconstruction could be also used for studying in a more scientific fashion the relation between the elements of a urban or indoor space, in order to better understand the purposes of the author, or to reproduce the appearance of the work of art in its original position.

Acknowledgements. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n. 270404 (NoE "V-Must").

References

[AFS*11] AGARWAL S., FURUKAWA Y., SNAVELY N., SIMON I., CURLESS B., SEITZ S. M., SZELISKI R.: Building rome in a day. *Commun. ACM 54*, 10 (Oct. 2011), 105–112. URL: http://doi.acm.org/10.1145/2001269.2001293, doi:10.1145/2001269.2001293. 2

[CCC*08] 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), Eurographics, pp. 129–136. 3

[CDPS09] CORSINI M., DELLEPIANE M., PONCHIO F., SCOPIGNO R.: Image-to-geometry registration: a mutual information method exploiting illumination-related geometric properties. *Computer Graphics Forum 28*, 7 (2009), 1755–1764. URL: http://vcg.isti.cnr.it/Publications/2009/CDPS09. 4

[DCCS13] DELLEPIANE M., CAVARRETTA E., CIGNONI P., SCOPIGNO R.: Assisted multi-view stereo reconstruction. In *3DTV-Conference, 2013 International Conference on* (July 2013), pp. 318 – 325. URL: http://vcg.isti.cnr.it/Publications/2013/DCCS13. 2

[DDUC*13] DELLEPIANE M., DELL UNTO N., CALLIERI M., LINDGREN S., SCOPIGNO R.: Archeological excavation monitoring using dense stereo matching techniques. *Journal of Cultural Heritage 14*, 3 (May-June 2013), 201–210. http://dx.doi.org/10.1016/j.culher.2012.01.011. URL: http://vcg.isti.cnr.it/Publications/2013/DDCLS13. 2

[DRMSne] DAVISON A., REID I., MOLTON N., STASSE O.: Monoslam: Real-time single camera slam. *Pattern Analysis and Machine Intelligence, IEEE Tr. on 29*, 6 (June), 1052–1067. doi:10.1109/TPAMI.2007.1049. 2

[DS12] DELLEPIANE M., SCOPIGNO R.: Teaching 3d acquisition for cultural heritage: a theory and practice approach. In *Proceedings of Eurographics Conference 2012 - Education Papers* (2012), Eurographics, Eurographics, pp. 25–32. URL: http://vcg.isti.cnr.it/Publications/2012/DS12. 3

[FCS*10] FURUKAWA Y., CURLESS B., SEITZ S. M., SZELISKI R., INC G.: R.: Towards internet-scale multiview stereo. In *In: Proceedings of IEEE CVPR* (2010). 2

[FHZ*13] FRAHM J.-M., HEINLY J., ZHENG E., DUNN E., FITE-GEORGEL P., POLLEFEYS M.: Geo-registered 3d models from crowdsourced image collections. *Geo-spatial Information Science 16* (2013). 2

[FP10] FURUKAWA Y., PONCE J.: Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*, 8 (2010), 1362–1376. doi:http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.161. 2

[GRZ03] GRUEN A., REMONDINO F., ZHANG L.: Image-based Automated Reconstruction of the Great Buddha of Bamiyan, Afghanistan. *Comp. Vision and Pat. Rec. W. 1* (2003), 13. doi:http://doi.ieeecomputersociety.org/10.1109/CVPRW.2003.10003. 2

[GSC*07] GOESELE M., SNAVELY N., CURLESS B., HOPPE H., SEITZ S. M.: Multi-view stereo for community photo collections. In *Proceedings of ICCV 2007* (Rio de Janeiro, Brazil, 2007), IEEE, pp. 265–270. 2

[KBH06] KAZHDAN M., BOLITHO M., HOPPE H.: Poisson surface reconstruction. In *Proc. of Eurographics symp. on Geometry processing* (Aire-la-Ville, Switzerland, Switzerland, 2006), Eurographics Association, pp. 61–70. URL: http://portal.acm.org/citation.cfm?id=1281957.1281965. 3

[KM09] KLEIN G., MURRAY D.: Parallel tracking and mapping on a camera phone. In *Proceedings of ISMAR '09* (Washington, DC, USA, 2009), IEEE Computer Society, pp. 83–86. URL: http://dx.doi.org/10.1109/ISMAR.2009.5336495, doi:10.1109/ISMAR.2009.5336495. 2

[Kwo04] KWON M.: *One Place After Another: Site-specific Art and Locational Identity*. MIT Press, 2004. URL: http://books.google.it/books?id=s8KviDnz1SwC. 1

[LA09] LOURAKIS M. I. A., ARGYROS A. A.: Sba: a software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software* (2009), 1–30. 2

[LDWov] LEONARD J., DURRANT-WHYTE H.: Simultaneous map building and localization for an autonomous mobile robot. In *Intelligent Robots and Systems '91* (Nov), pp. 1442–1447 vol.3. doi:10.1109/IROS.1991.174711. 2

[LM11] LAFARGE F., MALLET C.: Building large urban environments from unstructured point data. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (Nov 2011), pp. 1068–1075. doi:10.1109/ICCV.2011.6126353. 2

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *Int. J. C. Vision 60*, 2 (Nov. 2004), 91–110. URL: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94, doi:10.1023/B:VISI.0000029664.99615.94. 2

[SmFPG06] SINHA S. N., MICHAEL FRAHM J., POLLEFEYS M., GENC Y.: Gpu-based video feature tracking and matching. 2

[SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: exploring photo collections in 3d. In *ACM SIGGRAPH 2006* (2006), ACM, pp. 835–846. URL: http://doi.acm.org/10.1145/1179352.1141964, doi:http://doi.acm.org/10.1145/1179352.1141964. 2

[TKM*13] TANSKANEN P., KOLEV K., MEIER L., CAMPOSECO F., SAURER O., POLLEFEYS M.: Live metric 3D reconstruction on mobile phones. In *IEEE International Conference on Computer Vision (ICCV)* (Sydney, Australia, December 2013). 2

[TMHF00] TRIGGS B., MCLAUCHLAN P. F., HARTLEY R. I., FITZGIBBON A. W.: Bundle adjustment - a modern synthesis. In *Proc of ICCV '99* (London, UK, UK, 2000), Springer-Verlag, pp. 298–372. URL: http://dl.acm.org/citation.cfm?id=646271.685629. 2

[TSH*11] TUITE K., SNAVELY N., HSIAO D.-Y., TABING N., POPOVIC Z.: Photocity: Training experts at large-scale image acquisition through a competitive game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2011), CHI '11, ACM, pp. 1383–1392. URL: http://doi.acm.org/10.1145/1978942.1979146, doi:10.1145/1978942.1979146. 2

[WACS11] WU C., AGARWAL S., CURLESS B., SEITZ S. M.: Multicore bundle adjustment. *Work 10*, x (2011), 3057–3064. URL: http://grail.cs.washington.edu/projects/mcba/pba.pdf. 2