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)

Improving High-Speed Scanning Systems by Photometric Stereo

Frédéric Larue, Matteo Dellepiane and Roberto Scopigno

ISTI-CNR, Pisa, Italy

Abstract

High-speed scanning systems can be extremely valuable for Cultural Heritage applications, especially when large collections of small objects have to be acquired. However, fine details may not be acquired using this technology. Nevertheless, it is possible to try to recover them by taking advantage of the additional data provided by these systems: the calibrated video sequence of the acquisition, and the position of the projector light for each frame. In this paper, we propose a workflow that processes the video sequence with a photometric stereo approach, in order to refine the coarse geometry provided by the scanner. A normal map is first extracted by a method that accounts for the unevenly distributed sampling that generally results from the particular trajectory followed by this kind of scanners during the acquisition. This normal map is then integrated in order to recover missing geometric features. Good performances are achieved, since the whole workflow is particularly suited to GPU programming.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Digitizing and scanning

1. Introduction

3D scanning systems present a major interest for Cultural Heritage (CH), from the digital archiving point of view as well as for study and restoration purposes. Recent technology advances have given rise to new high speed systems, like in-hand scanning devices, that present big advantages for CH. First of all, they are able to produce digital copies in a few minutes, which is really important when it is necessary to digitize huge collections in reasonable times. Moreover, they provide interactive feedback. This avoids the need of heavy post-processing to present intermediate results, as well as manual intervention in the processing chain, so that the acquisition can be performed by a non-expert user. Unfortunately, rapid acquisition could lead to a loss in accuracy with respect to traditional scanning devices.

Most of the high speed digitization systems are based on structured light, using a setup made of a video camera and a projector. This often makes it possible to produce, in addition to the object geometry, a fully calibrated video flow where camera parameters as well as lighting conditions (projector location) are perfectly known for each frame. We propose to exploit this additional information, which is generally ignored, in order to recover the finest geometric details missed by such kind of scanners. We begin by extracting a normal map from the video flow. Due to the fact that it generally follows a particular trajectory, the distribution of the recovered sampling is rarely good enough to correctly constrain the normal fitting problem. We then analyze the sampling distribution so as to avoid incoherent normal orientations where the sampling is too sparse. Once the normal map is computed, an integration approach is used to refine the initial geometry by displacement mapping.

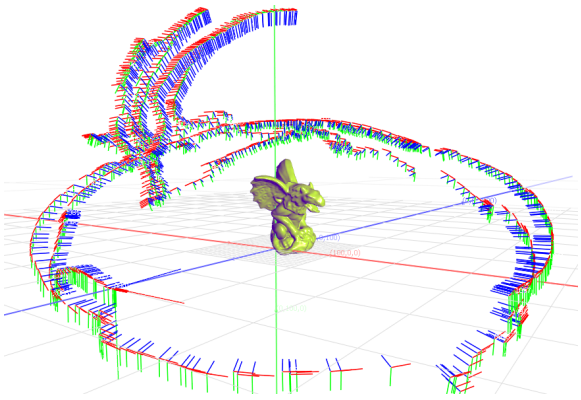

Figure 1: *Example of a 3D model with the associated acquisition frames, which shows a typical path for a small object.*

2. Related work

Developing high speed scanning systems means facing two main issues. The first one is the availability of technology for the 3D acquisition. Concerning this point, fast structured-light solutions, where a high speed camera and a projector are used to recover range maps in real-time, proved to be the most robust approach [HHR01]. The second issue is the problem of aligning data in a very fast way, which is usually solved with smart implementations of the ICP algorithm [RHHL02, WWLG09]. In the last few years, some real-time scanning solutions have been proposed [RHHL02, ZCS03, WLG07] that essentially differ on the way projection patterns are handled, and on the ICP implementation. But they are then potentially able to produce a fully calibrated video sequence in addition to the geometry.

Photometric stereo consists in computing high quality normal/range maps by taking several photographs from the same viewpoint, but with different illumination directions [RTG97, Noz10, MWGA06]. The need of moving the light source generally makes this kind of methods impractical for the design of convenient hardware setup. A solution is to fix the light to the camera, and to compute normal/depth from the illumination changes observed while moving the object [LHYK05, HMJI09]. But then, new non trivial problems occur, like camera localization or self occlusions management from one image to another, that introduce big inaccuracies. Even if some real-time photometric stereo methods exist [Noz10, MWGA06], they can hardly compete in terms of convenience with structured light scanners for application fields like CH.

Similarly to our approach, [NRDR05] propose a hybrid scanner that uses normals computed by photometric stereo to improve the geometry obtained by a metrology scanner. Like in our case, occlusion from one viewpoint to another can be immediately and efficiently detected. But their method relies on a specific hardware configuration that is absolutely not intended for performing real-time digitization. To our knowledge, the use of the video flow produced by real-time scanners as input for a photometric stereo approach has never been done. This could be explained by the fact that this flow is generally used on-the-fly, for range map computation, and not conserved afterwards for post-processing.

3. Photometric stereo based geometry improvement

The video flow of real-time scanning systems has already been used to produce 3D models with quality color textures [LDHS11]. We extend here this protocol in order to achieve high quality normal fitting and geometry refinement.

3.1. Basic estimation of normal vectors

In the following, vectors are assumed to be column vectors. Let $\{F_i\}$ be the set of frames corresponding to the acquisition sequence. To each frame F_i is associated a light position ℓ_i, corresponding to the scanner projector's location. Assuming that the object surface is Lambertian, the color c_i observed at a given surface point p in F_i is given by equation 1:

$$c_i = \frac{\rho_d}{d_i^2}\left(n_p^T\, l_i\right), \quad \text{with } d_i = \|\ell_i - p\|,\ l_i = \frac{\ell_i - p}{\|\ell_i - p\|} \tag{1}$$

where n_p is the normal at p and ρ_d a constant depending on the light intensity, the surface diffuse albedo and the camera transfer function. If p is visible in several frames $\{F_i\}_{1\le i\le N}$, the normal n_p that fits the best the N measurements can be found by solving the following equation:

$$\rho_d\, n_p = \arg\min\ \xi(X) \tag{2}$$

where ξ is a quadratic form defined as:

$$\xi(X) = (LX - C)^2 \tag{3}$$

with:

$$L = \begin{bmatrix} w_1\, l_1^T \\ \vdots \\ w_N\, l_N^T \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} w_1\, c_1\, d_1^2 \\ \vdots \\ w_N\, c_N\, d_N^2 \end{bmatrix} \tag{4}$$

In equation 4, w_i represents a confidence value for the i^{th} measurement. In our case, it comes from the weighting framework proposed by Larue *et al.* [LDHS11]. Solving equation 2 is equivalent to finding the value of X for which the first derivative of ξ is null, which leads to equation 5:

$$\xi'(X) = 0 \iff X = (L^T L)^{-1}\,(L^T C) \tag{5}$$

The normal n_p is finally obtained by normalizing X.

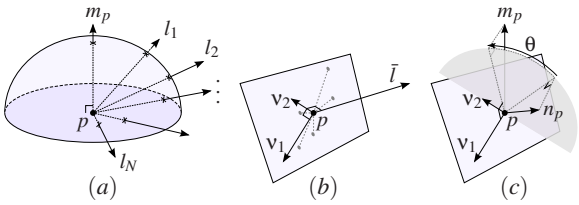

Figure 2: *Outline of our normal fitting correction.*

3.2. Light sampling analysis for improved normal fitting

The previous fitting method works well when light directions are correctly distributed on the hemisphere over p. However, in the case of high-speed scanning systems, digitization is generally performed by following a simple trajectory (sometimes using a turntable) which often leads to a direction sampling almost constrained to a plane (see Figure 1). This leads to an estimated normal which is reliable along this plane but really uncertain along the orthogonal direction.

To alleviate this problem, we propose to analyze the sampling distribution by performing at each point p a PCA on the set of light directions. The idea is then to rotate the fitted normal n_p so as to move it closer to the initial mesh normal m_p, but only along the direction of smallest sampling dispersion. To achieve this, we first compute the mean light

direction $\bar{l}$ (Figure 2-*a*). Vectors l_i are then projected onto the plane passing through p and orthogonal to $\bar{l}$, and the PCA is performed on the resulting set of 2D points. This leads to two unit eigenvectors ν_1 and ν_2, as well as their respective eigenvalues λ_1 and λ_2, with $\lambda_1 > \lambda_2 > 0$ (Figure 2-*b*).
By definition, ν_2 is the direction along which the sampling is the poorest. The correction is then applied by rotating n_p around ν_1 of an angle $\left(1 - \frac{\lambda_2}{\lambda_1}\right)\theta$, where θ is the angle between the projections of n_p and m_p onto the orthogonal plane to ν_1 (Figure 2-*c*). When the sampling is well distributed over the hemisphere of light directions, the eigenvalue ratio is close to 1 and then the fitted normal is almost not modified. On the contrary, when this ratio decreases, it means that the sampling distribution is not regular. In this case, n_p is corrected but only along the less reliable direction.

3.3. Geometry refinement from recovered normals

From our rectified normal field, a displacement map can be computed by integration in order to improve the initial geometry. We perform this integration in an iterative manner, by moving each point p along its initial normal m_p in the mesh of an amount computed by considering the set $\mathcal{N}_p$ of its nearest neighbors:

$$p_{new} = p + \frac{m_p}{|\mathcal{N}_p|} \sum_{q \in \mathcal{N}_p} m_p^T \frac{n_q^T (q-p) n_q + (q-p)}{2} \qquad (6)$$

The left part of the sum term tends to "push" p to follow the curvature defined by its neighbors' estimated normals, while the right part tends to preserve the position of p on the surface, so as to avoid excessive displacement. Equation 6 is then applied to all surface points corresponding to parameterization texels, and repeated until convergence.

4. Results

We tested the presented method on datasets obtained from an in-hand scanner. The comparison showed in Figure 3 justifies by itself our approach: as it can be seen, both 3D models directly produced by the scanner clearly lack fine geometric features, where our refinement based on the analysis of the scanning video sequence enables to recover most of them. This is particularly visible on the hat of the Dwarf model and on the tail of the Gargoyle. The last row shows the amplitude of the geometry deformation (in terms of Hausdorff distance) induced by the normal integration. As expected, regions close to features are clearly the most affected.
Figure 4 illustrates the efficiency of our normal correction procedure by showing the normal field computed for the same object with and without correction. The top row shows a region of the object where the sampling is well distributed (namely, that mostly covers the hemisphere of the possible lighting directions). There is almost no difference between both versions of the normal field, which is obvious since the ratio of the eigenvalues given by our PCA approach is, in this case, really close to 1. Both fittings behave similarly.

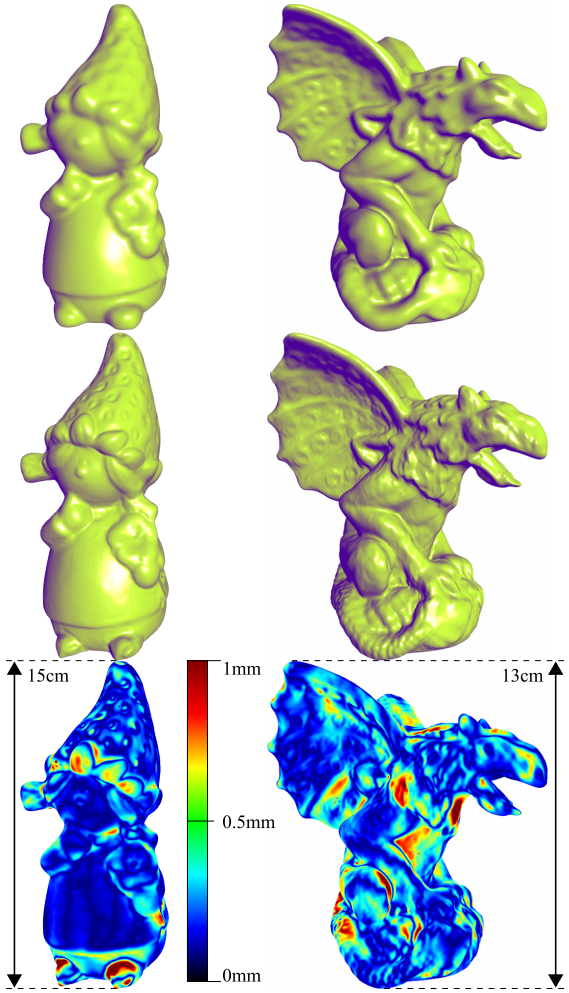

Figure 3: Top: *geometry acquired by an in-hand scanner.* Middle: *geometry improved by our method.* Bottom: *Hausdorff distance between both models.*

Conversely, the bottom row shows a surface part sampled from scanner locations that are almost collinear. As it can be seen, the standard fitting approach produces normals that present an exaggerated curvature as well as some cracks in the surface orientation. By introducing our correction, this erroneous behavior of the normal field completely disappears. In this case, since the eigenvalue ratio decreases, the estimated normal is forced to get closer to the original mesh normal along the direction of highest uncertainty. However, as explained before, the estimated normal is not modified along the direction where the sampling has its best distribution. This is why feature enhancing can still be achieved.
It is important to point out that every step of the proposed approach is suited to GPU programming. For equation 5, matrices $(L^T L)$ and $(L^T C)$ can be constructed incrementally by processing and accumulating input pictures one by one, ac-

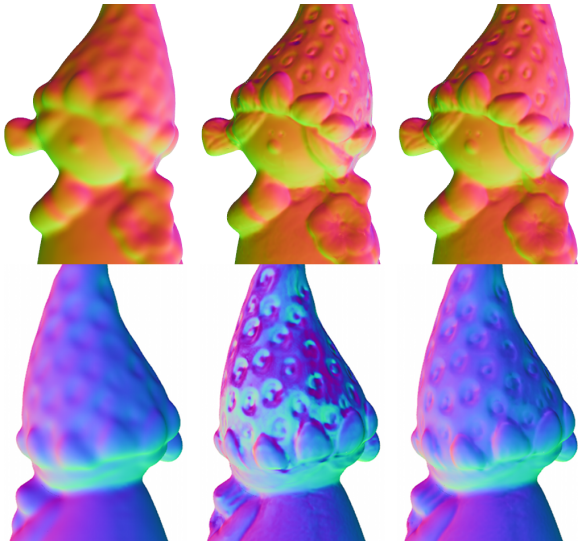

Figure 4: *Differences induced by our constrained fitting on the normal computation for a region with a good sampling (*top row*) and a region with a poor sampling (*bottom row*).* From left to right: *initial mesh normals; normals from unconstrained fitting; normals from our constrained fitting.*

Dataset	*Video length (# frames)*	*Processing time (in seconds)*		
		Nor. fitting	*Integr.*	***Total***
Gargoyle	1070	198	5	**203**
Dwarf	2110	277	6	**283**

Table 1: *Processing statistics for our test datasets.*

counting for shadowing and occlusions thanks to the initial geometry. For section 3.2, the PCA can be efficiently computed using shaders, inasmuch as eigenvectors and eigenvalues have a simple analytical formulation in the 2D case. Finally, integration of section 3.3 being performed by an iterative process, GPU implementation is also straightforward. Table 1 gives some processing times. The fitting task is clearly the most time consuming, since each video frame must be loaded and processed one by one. In spite of this, the whole improvement procedure only requires a few minutes.

5. Conclusion

In this paper we presented a technique that improves the geometric details of 3D models obtained from fast in-hand scanners. The video sequence obtained from the acquisition, together with the information about the light projector position, are used to refine the initial geometry. The method takes into account the possibility of a limited coverage in terms of direction of views, preventing possible incorrect normal estimations. The proposed method can be perfectly integrated with the typical acquisition using fast scanners: the rough geometry can give feedback about the coverage, while a short automatic post processing phase retrieves the fine details in minutes. Two of the possible future extensions of this work are: an accurate testing of the accuracy of the system, based on a comparison of objects acquired with different scanning technologies, and a weighting scheme accounting for the deviation of the measured lighting sampling with respect to the perfect Lambertian model currently used.

Acknowledgment

The research leading to these results has received funding from the EU 7^{th} Framework Program (FP7/2007-2013), through the 3D-COFORM project, under grant agreement n. 231809. We thank the ETH of Zürich, Switzerland, for having provided the datasets from their in-hand scanner.

References

[HHR01] HALL-HOLT O., RUSINKIEWICZ S.: Stripe boundary codes for real-time structured-light range scanning of moving objects. In *ICCV 2001* (2001), pp. 359–366 vol.2. 2

[HMJI09] HIGO T., MATSUSHITA Y., JOSHI N., IKEUCHI K.: A hand-held photometric stereo camera for 3-d modeling. In *ICCV* (2009), IEEE, pp. 1234–1241. 2

[LDHS11] LARUE F., DELLEPIANE M., HAMER H., SCOPIGNO R.: Automatic texturing without illumination artifacts from in-hand scanning data flow. In *Int. Workshop on Multimedia for Cultural Heritage (MM4CH)* (April 2011), Springer. 2

[LHYK05] LIM J., HO J., YANG M.-H., KRIEGMAN D.: Passive photometric stereo from motion. In *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2* (Washington, DC, USA, 2005), ICCV '05, IEEE Computer Society, pp. 1635–1642. 2

[MWGA06] MALZBENDER T., WILBURN B., GELB D., AMBRISCO B.: Surface enhancement using real-time photometric stereo and reflectance transformation. In *Eurographics Symposium on Rendering/Eurographics Workshop on Rendering Techniques* (2006), pp. 245–250. 2

[Noz10] NOZICK V.: Pyramidal normal map integration for real-time photometric stereo. In *EAM Mechatronics 2010* (Japon, 2010), pp. 128–132. 2

[NRDR05] NEHAB D., RUSINKIEWICZ S., DAVIS J., RAMAMOORTHI R.: Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH 2005) 24*, 3 (Aug. 2005). 2

[RHHL02] RUSINKIEWICZ S., HALL-HOLT O., LEVOY M.: Real-time 3d model acquisition. *ACM Trans. Graph. 21* (July 2002), 438–446. 2

[RTG97] RUSHMEIER H. E., TAUBIN G., GUÉZIEC A.: Appying shape from lighting variation to bump map capture. In *Proc. of the Eurographics Workshop on Rendering Techniques '97* (London, UK, 1997), pp. 35–44. 2

[WLG07] WEISE T., LEIBE B., GOOL L. V.: Fast 3d scanning with automatic motion compensation. In *IEEE CVPR'07* (June 2007). 2

[WWLG09] WEISE T., WISMER T., LEIBE B., GOOL L. V.: In-hand scanning with online loop closure. In *3DIM09* (October 2009). 2

[ZCS03] ZHANG L., CURLESS B., SEITZ S.: Spacetime stereo: shape recovery for dynamic scenes. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (june 2003), vol. 2, pp. II – 367–74 vol.2. 2