

Passive reconstruction of high quality textured 3D models of works of art

N. Brusco, L. Ballan and G. M. Cortelazzo

Università degli Studi di Padova, Dipartimento di Ingegneria dell'Informazione

Abstract

A wide-spread use of 3D models in cultural heritage application requires low cost equipment and simple modeling procedures. In this context, passive 3D reconstruction methods allow to build 3D models from a set of calibrated cameras, without the need of expensive machinery. Unfortunately the surfaces characteristics often lead to bad quality reconstructions. Recent efforts attempt to combine together information from different passive methods in order to improve the overall quality of the result. The combination of stereo matching and silhouette information has recently received considerable attention. Typically the major contribution to the appearance of the model comes from texture, rather than from geometry. The straightforward application of the photographs over the model can lead to artifacts, due to errors in 3D reconstructions, which must be minimized. This work, building on recent results, proposes a variation of an algorithm for 3D geometry recovery from stereo and silhouette information within a classical deformable model framework, which improves the quality of the shape. In order to avoid visible texture artifacts, it also proposes a new algorithm for texture synthesis based on wavelet decomposition. Experimental verification shows the effectiveness of the proposed solution with respect to robustness, computational speed and quality of the final result.

1. Introduction

In the scientific community the interest for 3D modeling of cultural heritage's objects such as vases and statues is gaining interest. The documentation of objects should not just concern their geometrical characteristics but also their color, since usually texture gives the main contribution to the visual appearance of an object [CB04].

Methods for recovering the 3D geometry of objects can be divided into passive and active ones. Passive sensing refers to the measurement of visible radiation which is already present in the scene; active sensing refers instead to the projection of structured light patterns onto the object or scene to be modeled.

Active sensing facilitates the computation of 3D structure by intrinsically solving the correspondence problem, a major issue with passive techniques. In general, active techniques such as those based on laser range scanning or light pattern projection tend to be more accurate but more expensive and slower than

their passive counterparts. Furthermore, when it is more important to realize photorealistic 3D models than metrologically accurate surfaces, passive 3D reconstruction methods, coupled with good quality textures, can lead to good results.

For these reasons and since passive techniques essentially require standard image capture devices such as photo-cameras or video-cameras, the interest towards passive 3D reconstruction techniques is bound to remain rather high. Historical passive sensing methods are stereo vision, structure from motion, shape from shading, shape from silhouette and space carving.

Recent efforts attempt to combine together informations of different passive methods. Critical issues in this research are what type of data to use and how to combine them, in order to actually increase the overall information. The combination of stereo matching and silhouette information has recently received considerable attention both for obtaining high quality 3D models [ES04] and for modelling 3D dynamic scenes

[MWTN04], an application often referred to as 3D video.

This work addresses this approach within the classical deformable models framework and proposes a solution which has some advantages in order to obtain a high quality 3D model.

After obtaining a 3D model, it is necessary to map textures over it in order to obtain a photorealistic result. Textures are available as a collection of images, which often are the same images used for 3D reconstruction. Mapping of such images over the model could lead to artifacts, which are more evident when there are errors in the 3D reconstruction, as it is the case of low precision techniques as typical passive methods. Such artifacts usually correspond to blurring of high frequencies spatial features or in visible discontinuities at the boundaries between adjacent triangles, which are taken from different original source images.

This work presents an original texture processing method based on wavelet analysis for obtaining high quality textured 3D models.

This paper has six sections. The second section recalls the 3D passive recovery from stereo and silhouette information and points out the most delicate issues. Sections 3 reformulates the problem within classical deformable models framework, defines a new force related to silhouette information, proves some theoretical advantages of the proposed reformulation and shows how to solve it. Section 4 addresses a fine grain improvement leading to a re-sampling more respectful of the geometrical quality of the mesh. Section 5 describes the method for the creation of a photorealistic texture for the 3D model. Section 6 presents some experimental results. Section 7 draws the conclusions.

2. Shape recovery from stereo and silhouette information

The proposed 3D passive shape recovery procedure combines silhouette and stereo-matching information as schematically shown in Fig.1. The silhouettes are obtained by a segmentation algorithm [Can86] [LM01] from a sequence of photographs of the object taken from different positions depending on the characteristics of the scene. Stereo matching is also applied to the picture pairs of the sequence of photographs if the object is textured. Not all the pictures used for the silhouettes methods are used for stereo matching.

Silhouettes are first used by a shape-from-silhouette method [MFK99] [Lau94] [Pot87] in order to obtain a coarse estimate of the surface; they are then used in order to correct for stereo-matching errors. The main advantages of shape-from-silhouette methods are that

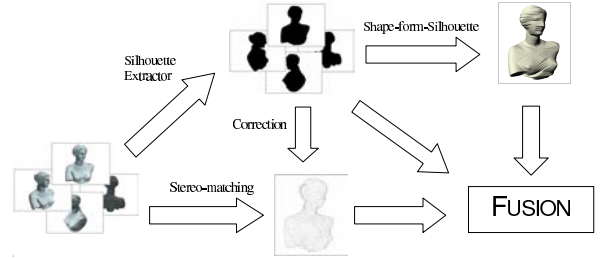


Figure 1: *The proposed passive 3D modeling pipeline.*

the obtained objects are well shaped and there are no problems with reflecting objects or objects without texture (if the segmentation algorithm is robust). The major drawback is that concavities cannot be modelled.

Texture is used by stereo matching methods [Kos93] [KSK98] which, differently from silhouette based techniques, can model concavities. Stereo-matching does not work in regions without significant texture or where the available texture exhibits some periodicity. 3D data near the silhouette edge are usually missing, since in these regions the object points can be easily mismatched with the background. Luckily, shape-from-silhouette methods can model these regions rather well.

Assuming the above 3D reconstruction process provides us with the coordinates of n points lying on a real surface Λ , the points could be expressed as $x_i = y_i + \varepsilon_i$, where y_i are the true values and ε_i are the measurement errors. We also have m views V_j of Λ , which can be considered as functions mapping \mathcal{R}^3 in \mathcal{R}^2 through projective transformations. For each view V_j we know the projection $P_j = V_j(\Lambda)$ of the original surface Λ , i.e. the set of points representing the silhouettes of Λ . The fusion problem of silhouettes and stereo matching information concerns the estimate of Λ from x_i and P_j .

Such a problem can be solved within a classical deformable model framework [ES04] [MWTN04]. Namely, a surface is made to evolve subject to three types of forces, an internal and two external ones. The first one, F_{int} , keeps the surface as smooth as possible, while the others, F_{tex} and F_{sil} , make it to converge to Λ . Formally, the evolution of the model at point P can be described as:

$$s(0) = s_0 \quad (1)$$

$$\frac{\partial s}{\partial t}(t)(P) = F_{int}(P, s) + F_{tex}(P) + F_{sil}(P, s) \quad (2)$$

where $s(t)$ is the estimate of Λ at iteration t , s_0 is the estimate obtained through the shape-from-silhouette

method, and

$$F_{int}(P, s) = \nabla^2 s(P) - \nabla^4 s(P) \quad (3)$$

F_{tex} deforms the model in order to minimize its distance from cloud x_i ; F_{sil} deforms the model in order to make it consistent with silhouette information i.e., F_{sil} tends to make the model silhouettes as similar as possible to the acquired ones.

In [ES04] F_{tex} is expressed as the Gradient Vector Flow (GVF) [XP98], obtained from point cloud x_i with the aim of eliminating the local minima arising when the surface reaches a boundary concavity.

As for F_{sil} , in [MWTN04] it is defined as follows:

$$F_{sil}(P) = \sum_{j=1}^m f^{V_j}(P) \quad (4)$$

where $f^{V_j}(P)$ is nonzero iff $V_j(P)$ is external to P_j or is internal to P_j and on the boundary of $V_j(s)$ (that is, $V_j(s)$ is the silhouette of s viewed from V_j). In this case $f^{V_j}(P)$ is the back-projection of the 2D vector joining $V_j(P)$ with the nearest point on $V_j(s)$ boundary. Hence the force is nonzero only along the curves obtained by sectioning s with the image planes relative to V_j and passing through P . The force field is therefore strongly discontinuous and can't have variational origin, i.e., it can't be derived from the Euler-Lagrange equations of a minimum problem. As a consequence, convergence to a model consistent with silhouette information is not guaranteed. Moreover, the force is calculated as the sum of terms separately computed on each silhouette.

In [ES04] F_{sil} is defined as

$$F_{sil}(P) = \alpha(P) \cdot d_{vh}(P) \cdot n(P) \quad (5)$$

where $n(P)$ is the normal to the surface in P and $d_{vh}(P)$ is the signed distance from the visual hull defined as $d_{vh}(P) = \min_j d(V_j(P), \partial P_j)$, where ∂P_j is the boundary of P_j and $d(V_j(P), \partial P_j)$ is the signed distance between the projection v of P viewed from V_j and ∂P_j , positive when $v \in P_j$ and otherwise negative. $\alpha(P)$ can be expressed as

$$\alpha(P) = \begin{cases} 1 & \iff d_{vh}(P) \leq 0 \\ \frac{1}{(1+d(V_c(P), \partial V_c(s)))^k} & \iff d_{vh}(P) > 0 \end{cases} \quad (6)$$

where $c = \arg \min_j d(V_j(P), \partial P_j)$ and $\partial V_c(s)$ is the boundary of the silhouette of s viewed from V_c . The force field is continuous, defined over the entire model with the same direction and versus as the surface normal. Namely, the boundary points are subject to a force equal (both in magnitude and versus) to the distance vector from the visual hull. For the points internal to $V_j(s)$ the force intensity is modulated by $\alpha(P)$ so as to be inversely proportional to their distance



Figure 2: *The Buddha.*

from the boundary. In this way, the vertices can leave the visual hull and enter the concavity opposing only a force decreasing by $\alpha(P)$ with respect to the distance from the boundary. Eq.(5) avoids the difficulties arising from the sum of Eq.(4).

For the above reasons we have chosen to use F_{sil} introduced in [ES04].

3. Refinement

The results obtained by the method proposed in section 2 are rather good with respect to parametric and geometric quality, as shown by Fig.3.

However, with this method, the final model is bound to have a parametrization similar to an isometry, i.e., to be uniformly sampled (see Fig.4). Namely, this is due to the use of F_{sil} . However, a good quality mesh should be sampled proportionally to its local curvature while in deformable models the sampling rate is fixed. In this case some regions could be undersampled and others oversampled, with consequent poor mesh quality or too large model sizes, respectively.

In order to solve this problem we defined a second phase of evolution, to be started after the model reaches adequate convergence to the original object. A selective subdivision of the model based on local curvature is first needed. Then, intuitively, we could think of evolving the model by Eq.(2). Unfortunately, as shown in Fig.5, this would not result in a correct evolution because F_{sil} is not correctly related to the curvature of a multiresolution mesh, which is our case.

Therefore, we choose to use another type of silhou-



Figure 3: 3D Model of Buddha reconstructed using snakes.

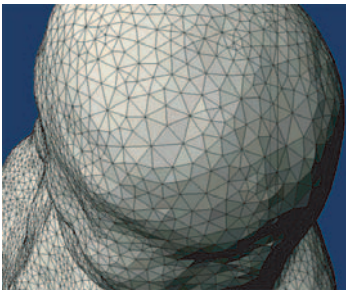


Figure 4: Detail of Buddha wireframe.

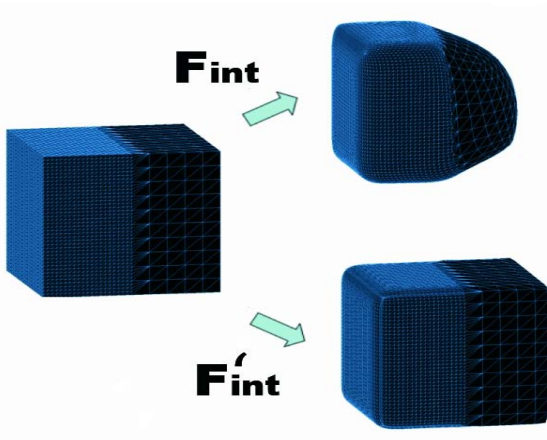


Figure 5: Comparison between F_{int} and F'_{int} .

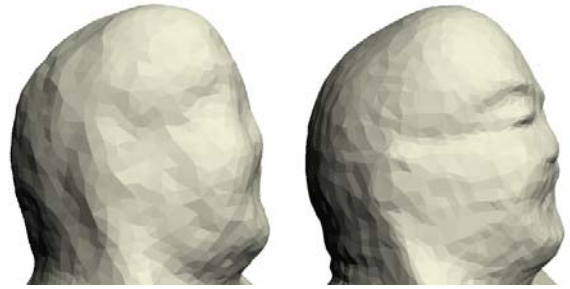


Figure 6: The resolution enhancement allows us to capture details of the original model in high curvature regions. Left: model obtained using classical deformable models. Right: model obtained using variable sampling rate.

ette force

$$F'_{int}(P, s) = -\bar{\kappa}n \quad (7)$$

where $\bar{\kappa}$ represents the mean curvature of the surface and n is the surface normal. As shown in Fig.5, F'_{int} is independent of the chosen parametrization. Unfortunately F'_{int} is hard to compute and we refer to [DMSB99] for a numerical implementation.

In this way only the surface geometrical properties are changed, while the parametrization chosen through subdivision is untouched. We thus obtain a sampling which respects compactness and geometrical quality of the mesh. The resolution enhancement allows us to capture details of the original model in high curvature regions (see Fig.6). In the formulation of [ES04] this would have been impossible, as it would imply a prohibitive model size.

4. Texture

We assume that images V^i , $i = 0 \dots N - 1$ of the object are available, taken by a pre-calibrated camera. A new texture V^* for the full 3D model may be obtained through projection of textures V^i and a weighted average [ABC04] [ABC03]. Unfortunately, any imperfection in the surface geometry description produces misalignment the separate projections, so that averaging tends to blur high frequency spatial features.

Another solution consists in keeping the original images V^i and selecting a single most appropriate original view image for each triangle. Although stitching avoids the blurring problem, it tends to produce visible discontinuities at the boundaries between adjacent triangles which are mapped on different original source views.

To hide these artifacts we could form each possi-

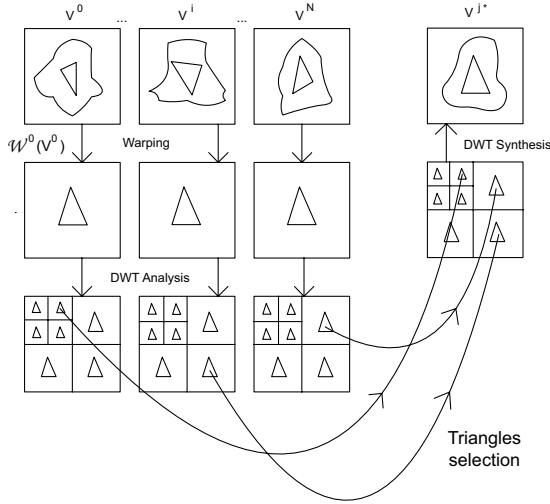


Figure 7: Generation of V^{j*} .

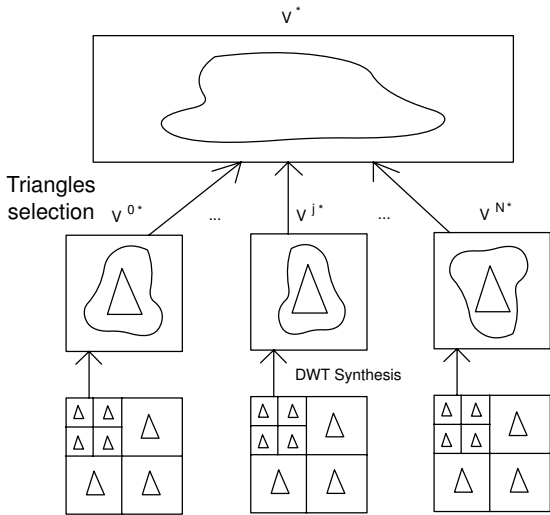


Figure 8: Generation of the global texture V^* .

ble rendering V^* in real time in the discrete wavelet transform or DWT domain as described in [ZBTC05]. Since rendering engines do not cope with the distortion framework of [ZBTC05], here we prefer to produce a texture statically connected with the geometry, which can be easily rendered with any commercial software.

Our problem is that of obtaining a global multispectral texture V^* from N images $V^i, i = 0, 1, \dots, N - 1$. The registration of each texture on the mesh of triangles induces on each image a subdivision in planar triangular patches. Consider surface triangle Δ_n and assume that the geometry portion including Δ_n is visible

from the viewpoint of both image V^i and V^j . In this case both images have a planar triangular patch imaging Δ_n , which we will denote as $V^i(\Delta_n)$ and $V^j(\Delta_n)$ respectively. The relationship between such two planar patches is an affine warping. Call \mathcal{W}_{ij} the warping operator which transforms image V^i in the image $\mathcal{W}_{ij}(V^i)$. This operator warps every $V^i(\Delta_n)$ into the planar triangular patch corresponding to the triangle of geometry Δ_n projected on V^j .

The texture generation procedure, which we call DWT stitching, has the following steps:

1. For each image $V^j, j = 0 \dots N - 1$, warp each image $V^i, i = 0 \dots N - 1$ to obtain the set of N^2 warped image $\mathcal{W}_{ij}(V^i)$, as pictorially shown in the left upper part of Fig. 7.
2. apply DWT analysis on the set of N^2 images $\mathcal{W}_{ij}(V^i), i, j = 0 \dots N - 1$ over the set of subbands, LL_d^i and LH_d^i, HL_d^i, HH_d^i for $d = 1, 2, \dots, D$, where D is the number of DWT decomposition levels, as pictorially shown in the left part of Fig. 7.
3. for each $V^{j*}, j = 0 \dots N - 1$, build the DWT coefficients from those of the images $\mathcal{W}_{ij}(V^i), i = 0 \dots N - 1$ by assigning to each triangle $V^{j*}(\Delta_n)$ of every subband the DWT coefficients of that subband associated to only one source image $\mathcal{W}_{ij}(V^i(\Delta_n)), i = 0 \dots N - 1$. This source image of index i is selected as the one with normal more parallel to that of the surface triangle Δ_n . This step is pictorially shown by the lower part of Fig. 7.
4. recover images $V^{j*}, j = 0, \dots, N - 1$ from their coefficients by DWT synthesis, as pictorially shown by the right part of Fig. 7.
5. generate the global texture by projecting on each triangle Δ_n of the mesh the $V^{j*}(\Delta_n)$ with the same criteria used before for the choice of triangles, as pictorially shown in Fig. 8.

Note that images V^{j*} are modified versions of V^j consistent with geometry. Blurring is avoided, since DWT synthesis maintains high frequency details. Furthermore discontinuities are smoothed. In the final model, each triangle Δ_n of the geometry is taken from the most parallel view V^{j*} , and the wavelet coefficients of $V^{j*}(\Delta_n)$ come from the ones of $V^j(\Delta_n)$, since the choice of triangles is the same in points 3 and 5. However, the portion of image $V^{j*}(\Delta_n)$ is different from $V^j(\Delta_n)$ because the contribution of adjacent triangles on the boundaries smooths the transition between different V^{j*} . An example of V^j and V^{j*} is shown in Fig.9. Details of Fig.9c) and Fig.9d) show the little but important texture variations.

The final texture is made of $V^{j*}, j = 0 \dots N - 1$, and of a mapping from each triangle Δ_n and a corresponding V^{j*} . The final textured model of buddha is shown in Fig.10. Benefits of DWT are shown in Fig.11:



Figure 9: The Buddha: a) V^j and b) V^{j*} c) detail of V^j d) detail of V^{j*} .



Figure 10: The textured Buddha

in Fig.11a) stitching without DWT is applied, and a discontinuity is evident. In Fig.11b), thanks to DWT, the transition is very smooth.

5. Experimental results

Tests for the shape reconstruction were first performed on synthetic models. The pictures of the synthetic models were generated by a rendering software. The model was framed by n 43mm target cameras. The

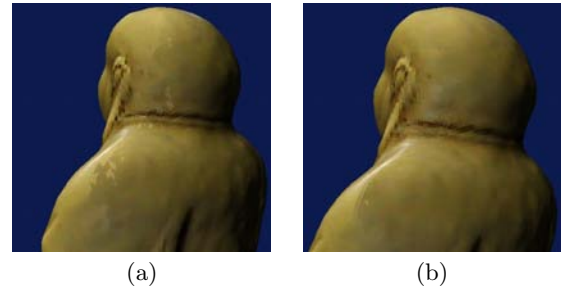


Figure 11: back of Buddha a) without DWT stitching b) with DWT stitching.

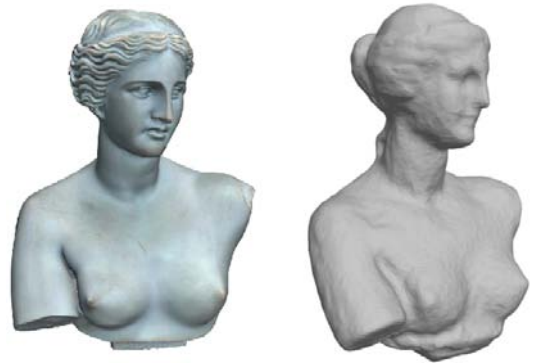


Figure 12: Synthetic model used to evaluate the reconstruction error of our system.

rendering was supervised by a script which also calculated the relative projection matrices.

Differently from stereo-based procedures, our method produces closed surfaces which are manifolds and which can be shown to be both geometrically and parametrically regular. Differently from shape-from-silhouette procedures, our method can accurately model concavities and produce a model closer to the real object shape. In Fig.3 the model is shown, obtained through snakes.

Moreover, we observed that silhouette information compensates for the lack of texture information and enhances the level of details in regions where texture is present. This property comes from the fact that intrinsic error of silhouette information is remarkably smaller than texture error. We also found that silhouette information corrects matching errors.

Reconstruction error was evaluated on a synthetic $120 \times 200 \times 260 \text{ mm}^3$ model shown in Fig.12. The model was acquired at 50cm distance with a $1024 \times 768 \text{ pixel}$ spatial resolution and a field of view of 35° .

Having both the original and the reconstructed

model, we finally estimated the average and maximum distance between the two surfaces. In our case we obtained $d_{average} = 0,82\text{ mm}$ (0,2 on the diameter) with a $0,62\text{ mm}^2$ variance and $d_{max} = 9,6\text{ mm}$.

Test was performed over the Buddha of Fig.2. The acquisition system consisted in a rotating table and a fixed camera. This choice is due to the simplicity and the automation of the acquiring set-up, although an hand-held camera with a calibration system can be employed. The object was positioned on the table while the software controlled table rotation and picture shooting. Sixty images were taken, spaced 6 degrees apart, for the construction of the 3D model. Texture has been obtained from six images, spaced 60 degrees apart, and applied to the model, which can be seen in Fig.10. 3D reconstruction, with the two-stages approach described above, is nice, even in presence of some reflections over the model surface. Because of an intrinsic error in 3D reconstruction, due to concavities not well modelled because of a lack of texture, some artifacts inevitably appear, but the overall result is definitely good. High frequency details are maintained and some artifacts due to illumination or error in geometry acquisition disappear, as shown in Fig.11.

6. Conclusions

The main contributions of this paper can be summarized as follows: the definition of a second stage of evolution in the silhouette-stereo fusion framework of [ES04] in order to obtain good quality meshes; an algorithm for the construction of a photorealistic texture which minimizes artifacts, based on wavelet decomposition. More specifically this paper reformulates a 3D passive multimodal reconstruction digitization scheme using both texture and silhouette information which synergically combines the typical properties of both passive techniques. Indeed, it can be proved to be resilient to measurement errors and capable of reconstructing a remarkably wide range of objects which includes:

- Surfaces characterized by good quality texture, sufficient lighting and not too high a specular reflectance (we recall that stereo-based methods completely fail to acquire objects even with minimal specular reflectance);
- Specular surfaces without texture or with a periodical texture, provided that the pictures take the profile of such surfaces (information in this case comes from the silhouettes);
- Concavities characterized by good texture and sufficient lighting.

The proposed method still doesn't allow the reconstruction of reflecting or transparent regions, nor the

modelling of objects not exhibiting the above mentioned features.

Indeed, the proposed approach gives a *closed regular manifold* with a *regular parametrization*, unlike stereo-based methods where neither the manifold nor the closure hypothesis generally hold. Therefore the proposed approach leads to a feasible minimum problem with respect to mesh quality.

Furthermore, the reconstruction error is rather satisfactory. For instance, the surface reconstructed from 1024x768 pictures taken from 50cm distance is affected by an average error of 0,8mm. Such an error can be remarkably reduced using digital cameras of higher resolution.

We also presented an algorithm for texturing the 3D model, in order to obtain a photorealistic result, coherent with the kind of errors typical of passive methods. The proposed method avoids most of the artifacts usually present in textured 3D models.

Further research will concern the combination of other passive methods together with silhouettes and stereo and the reformulation of the silhouette-stereo fusion framework of [ES04]. Current work attempts to incorporate in the method the *shadow-carving* [SRBP01].

References

- [ABC03] ANDREETTO M., BRUSCO N., CORTELAZZO G. M.: Color equalization of 3d textured surfaces. In *Proceedings of Eurographics 2003* (2003), pp. 168–173.
- [ABC04] ANDREETTO M., BRUSCO N., CORTELAZZO G. M.: Automatic 3d modeling of textured cultural heritage objects. *IEEE Trans. on Image Processing* 13, 3 (March 2004).
- [Can86] CANNY J.: Computational approach to edge detection. *PAMI* 8 (1986), 679–698.
- [CB04] CHENG I., BASU A.: In *Reliability and Judging Fatigue Reduction in 3D Perceptual Quality* (September 2004).
- [DMSB99] DESBRUN M., MEYER M., SCHRODER P., BARR A. H.: Implicit fairing of irregular meshes using diffusion and curvature flow. *International Conference on Computer Graphics and Interactive Techniques* (1999), 317–324.
- [ES04] ESTEBAN C. H., SCHMITT F.: Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding* 96, 3 (2004), 367–392.
- [Kos93] KOSCHAN A.: Dense stereo correspondence using polychromatic block matching. In *Proc. of the*

- 5th Int. Conf. on Computer Analysis of Images and Patterns, Budapest, Hungary* (1993), pp. 538–542.
- [KSK98] KLETTE R., SCHLNS K., KOSCHAN A.: *Three-Dimensional Data from Images*. Springer, Singapore, 1998.
- [Lau94] LAURENTINI A.: The visual hull concept for silhouette based image understanding. *IEEE PAMI* 16, 2 (1994), 150–162.
- [LM01] LUCCHESI L., MITRA S. K.: Color image segmentation: A state of the art approach. *Proc. of the Indian National Science Academy* 67, 2 (march 2001), 207–221.
- [MFK99] MATSUMOTO Y., FUJIMURA K., KITAMURA T.: Shape-from-silhouette/stereo and its application to 3-d digitizer. *Proceedings of Discrete Geometry for Computing Imagery* (1999), 177–190.
- [MWTN04] MATSUYAMA T., WU X., TAKAI T., NOBUHARA S.: Real-time 3d shape reconstruction, dynamic 3d mesh deformation, and high fidelity visualization for 3d video. *Computer Vision and Image Understanding* 96, 3 (2004), 393–434.
- [Pot87] POTMESIL M.: Generating octree models of 3d objects from their silhouettes in a sequence of images. *Computer Vision, Graphics, and Image Processing* 40 (1987), 1–29.
- [SRBP01] SAVARESE S., RUSHMEIER H. E., BERNARDINI F., PERONA P.: Shadow carving. *ICCV* (2001), 190–197.
- [XP98] XU C., PRINCE J. L.: Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing* (1998), 359–369.
- [ZBTC05] ZANUTTIGH P., BRUSCO N., TAUBMAN D., CORTELAZZO G.: Greedy non-linear optimization of the plenoptic function for interactive transmission of 3d scenes. *International Conference of Image Processing ICIP2005, Genova* (September 2005).