Multimodal 3D Shape Recovery from Texture, Silhouette and Shadow Information

Luca Ballan

Department of Information Engineering
University of Padova
Padova, Italy 35131

Guido Maria Cortelazzo

Department of Information Engineering
University of Padova
Padova, Italy 35131

Abstract

Recent efforts attempt to combine together information of different passive methods. Critical issues in this research are the choice of data and how to combine such data in order to increase the overall information. The combination of stereo matching and silhouette information has recently received considerable attention both for obtaining high quality 3D models and for modelling 3D dynamic scenes. In this paper we present a 3D shape recovery system which fuse together silhouette, texture and shadow information. More precisely, we formulate the fusion problem of these three types of information. Experimental verification shows that the new method is capable to reconstruct a wider range of objects.

1. Introduction

The methods for recovering the 3D geometry of objects can be classified in various ways. A typical classification distinguishes passive from active methods. 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 scanned. Active sensing is not always feasible, e.g., for modeling distant or fast-moving objects. On the other hand, passive techniques essentially require standard image capture devices such as photo-cameras or video-cameras. For these reasons, 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 silhouette, shape from shading, space carving, shape from defocus and shadow carving.

Recent efforts attempt to combine information from different passive methods. Those methods, called multimodal, have two main properties: the robustness to measurement errors and the capability of reconstructing a wide range of objects. Critical issues in this research are what type of data to use and how to combine them, in order to actually increase the overall amount of information. The combination of stereo matching and silhouette information has recently received considerable attention both for obtaining high quality 3D models [3], [17] and for modelling 3D dynamic scenes [10], an application often referred to as 3D video.

Our purpose is to use shadow information as well as silhouette and stereo, in the 3D recovery process. Shadow information is the same used by shadow carving method described in [16]. In other words, we detect all the regions of an object that are not illuminated by a light source and then we use such information in the surface estimation process.

In this paper we present a 3D shape recovery system which fuses together silhouette, texture and shadow information. More precisely, we formulate the fusion problem of these three types of information starting from the results obtained in our previous work [1]. Finally, we present some experimental results and advantages of the use of shadow information in the reconstruction process.

This paper has four sections. Sections 2 formulates the problem within classical deformable models framework, defines a new functional related to shadow information and proves its theoretical advantages. Section 3 presents some experimental results. Section 4 draws the conclusions.

2. The proposed method

The proposed 3D passive shape recovery procedure combines silhouette, stereo and shadow information as schematically shown in Fig.1. The first step of this pipeline has the purpose of extracting information from images and analyze them using known monomodal methods like shape-from-silhouette (SFS), stereo-matching and shadow detection algorithms. Afterwards, the second part fuses together all

Proceedings of the Third International Symposium on
3D Data Processing, Visualization, and Transmission (3DPVT'06)
0-7695-2825-2/06 $20.00 © 2006 IEEE

these kinds of information obtaining the final reconstructed surface. In order to carry out this last step, we need to formulate a fusion problem for these three types of information.

2.1. Information extraction

Given a set of calibrated images of the object taken from different positions, one can extract its silhouettes by using a segmentation algorithm [6].

Shadow information extraction, on the contrary, is not a trivial task as we cannot easily discriminate whether a surface point is a shadow or not. This is hard to determine since it is difficult to distinguish between a low reflectance point, i.e. a dark color point, and a point in a shadow region. Indeed, absence of light from a particular direction can be due to low reflectance as well as to insufficient illumination. Besides, insufficient illumination may be due to light sources too far from the object as well as to an actual shadow region. In the latter case one must ensure that the shadow is generated by the object itself and not by other objects in the scene.

In [16], Savarese et al. propose a conservative shadow detection method, i.e., a technique which classifies a point as shadow only when it is certain that it is a shadow. The inverse condition is not required so that there can be shadow points classified as non-shadow points. Obviously, the more shadow points are detected the more accurate is the reconstruction result. One must first set a threshold which separates light points from dark points. Afterwards, a point P of the object surface is classified as *"shadow detectable"* if and only if there exists at least one picture where it appears lighter than the threshold, otherwise it is classified as *"shadow undetectable"*. This provision ensures that the point is not a low reflectance point but it excludes all the points which are never lighted by the actual light sources. For every image, a point is a shadow point if and only if it is *"shadow detectable"* and it is darker than the threshold.

Silhouettes are first used by a SFS method [9] [5] [13] in order to obtain a coarse estimate of the surface. The main advantages of these methods are that 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.

Textures are used by stereo matching methods [8], [18], [20] 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. The latter problem can be partially avoided using a pyramidal approach [14] or other methods which use a-priori information about surface continuity (see [11]). Silhouette information can then be used to correct possible stereo-matching errors.

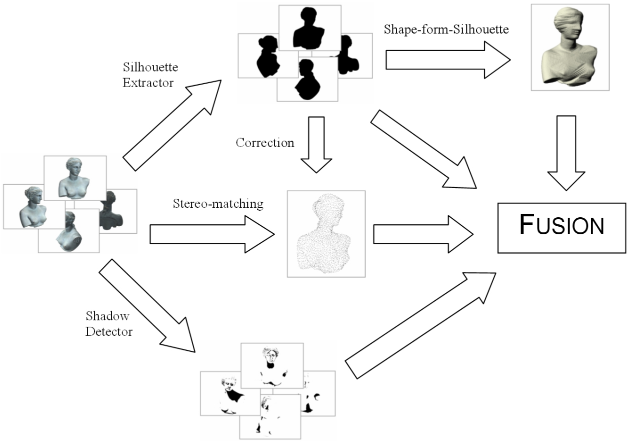

Figure 1. The proposed passive 3D modeling pipeline.

3D stereo 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.

2.2. Problem formulation

Stereo matching algorithm provides us with a set, call it Σ, of n points lying on the real surface Λ. Besides, from calibration we know a set of m views V_j, i.e., functions mapping $\Re^3$ in $\Re^2$ through a projective transformation associated to each camera that took photo of the object. Furthermore, from SFS method we have the projection $P_j = V_j(\Lambda)$ of the original surface Λ in each view V_j, i.e., the set of points representing the silhouettes of Λ viewed from each point of view V_j. Finally, we have a set of triads $(O, L, V_j)_k$ where $O \subset \Re^2$ is the set of image points representing the shadow regions detected in the picture taken from the point of view V_j with only one omnidirectional light source placed in $L \in \Re^3$.

The problem of fusing silhouettes, stereo and shadow information concerns the estimate $\overline{\Lambda}$ of the real surface Λ from extracted information Σ, $\{P_j\}$ and $\{(O, L, V_j)_k\}$. More precisely, we formulate our problem as an optimization problem within the domain of two-manifolds in $\Re^3$. In other words, the estimated model $\overline{\Lambda}$ will be the minimum point of a functional ξ defined as:

$$\xi(s) = \int_s k_{int} \cdot \xi_{int} + k_{tex} \cdot d_\Sigma + k_{sil} \cdot \xi_{sil} + k_{shad} \cdot \xi_{shad} ds \tag{1}$$

where s is a two-manifold and k_{int}, k_{tex}, k_{sil}, k_{shad} are constants a-priori fixed. The basic idea under this formulation is to penalize surfaces which have points, silhouettes and shadows inconsistent with the ones extracted form the real surface Λ in the previous step. Thus, minimizing Eq.(1) is equivalent to finding the manifold that interpolates stereo, silhouette and shadow data keeping a sort of smoothness over the entire surface in the way specified by ξ_{int}. Moreover d_Σ, ξ_{sil} and ξ_{shad} are the functionals that penalize discrepancies with stereo, silhouette and shadow information, respectively.

As in [12], ξ_{int} is set equal to the mean curvature $\overline{\kappa}$ of s, while, $d_\Sigma \colon \Re^3 \rightarrow \Re$ is defined as the distance between a point P on s and the point cloud Σ, as follows:

$$d_\Sigma\left(P\right) = \min\left\{d\left(P,x\right) \mid \forall x \in \Sigma\right\} \tag{2}$$

Moreover, as in our previous work [1], we define ξ_{sil}, i.e., the silhouette functional, as follows:

$$\xi_{sil} = S_c\left(V_c\left(P\right)\right) \tag{3}$$

where $V_c\left(P\right)$ is the projection of P onto the image plane of view V_c, $c = \arg\min_j d\left(V_j\left(P\right), \partial P_j\right)$ and $d\left(v, \partial P_j\right)$ is the signed distance between v and ∂P_j, which is positive if $v \in P_j$ or else negative; for additional details refer to [1]. We can prove that ξ_{sil} is zero all over the surface s iff all the silhouettes of s are equal to the respective silhouettes of Λ.

Finally, $\xi_{shad} \colon \Re^3 \rightarrow \Re$ is defined as follows

$$\xi_{shad}\left(P\right) = d_\Gamma\left(P\right) \tag{4}$$

where $d_\Gamma\left(P\right)$ is the distance between a point P on s and the surface Γ, obtained by applying shadow carving algorithm [15] to the surface s. Hence, Eq.(4) is zero all over the surface s iff s is equal to its shadow-carved surface, in other words, when it is consistent with shadow information.

Unfortunately, functional $\xi\left(s\right)$ presents several local minima points which will become a problem when we try to solve Eq.(1) numerically. In [19], Xu and Prince analyze local minima of d_Σ that arise when the surface s reaches a boundary concavity of Σ.

Differently from SFS and shadow-carving approaches, this problem formulation is not conservative, i.e., it doesn't solve for an upper bound of Λ. Therefore, concavities of $\overline{\Lambda}$ can be deeper than the shadow-carving estimation and than the real surface Λ. This is due to the functional ξ_{int} which, in order to maintain smoothness, eliminates surface corners.

A limitation of formulation (1) is due to the nature of self-shadow information. Indeed, there is a strong relationship between a shadowed region Φ and the portion of the surface Ψ that obscures it from the light source. In other words, if the estimation of the latter is incorrect then we obtain a bad estimate of the former. A typical error prone situation is depicted in Fig.2. Fortunately, Ψ, i.e. the portion of the surface that generates the shadowed region, is in general not shadowed except in case of double concavities. Therefore, the estimation of Φ can be done precisely using stereo and silhouette information.

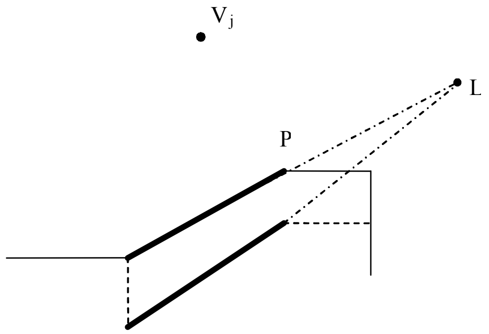

Figure 2. Typical error prone situation: L is the light source; the bold line indicates the shadow region; P is the portion of the surface that generates the shadow. If the estimation of P is incorrect (dotted line) then we obtain a bad estimation of the concavity.

A solution of Eq.(1) can be found by first computing the Euler-Lagrange equations for $\xi\left(s\right)$ and then by solving them through a gradient descent method. Obviously, the opposite of the gradient represents the forces deforming the surface (as described in [12] and [7]). Namely, a surface is made to evolve subject to four types of forces, an internal and tree external ones. The first one, F_{int}, keeps the surface as smooth as possible, while the others, F_{tex}, F_{sil} and F_{shad}, make it to converge to Λ. Formally, the evolution of the model at point P can be described as:

$$s\left(0\right) = s_0 \tag{5}$$

$$\frac{\partial s}{\partial t}\left(t\right) = F_{int} + F_{tex} + F_{sil} + F_{shad} \tag{6}$$

where $s\left(t\right)$ is the estimate of Λ at iteration t and s_0 is the estimate obtained through the shape-from-silhouette method.

F_{tex} deforms the model in order to minimize its distance from the point cloud Σ. 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. Finally, F_{shad} does the same for shadow information.

The force derived from ξ_{int} [2] is independent from the chosen parameterization of the surface s. In a numerical framework, s is treated as a 3d-mesh so that, we work on a specific parameterization of s. Consequently, we need an internal force that also regularizes the parameterization of the surface. For this reason, [7] proposes an internal force defined as follows

$$F_{int}\left(P,s\right) = \nabla^2 s\left(P\right) - \nabla^4 s\left(P\right) \tag{7}$$

Unfortunately, as described in [1], these forces produce low-quality final models. Therefore, as described and justified in the same article, we will use a mixture of (7) and [2] for F_{int}.

As mentioned above and as suggested by H. Esteban in [3], we use Gradient Vector Flow (GVF) [19] as F_{tex} in order to avoid local minima. Details about silhouette force were presented in our previous work [1].

As concerns F_{shad}, we take a similar approach as the use used for carving method. We know that a point p on s is inconsistent with shadow information intrinsic in the triad (O, L, V_j) iff p is visible from the projection center of V_j, $p \in O$ and p is visible from L. In other words, p is inconsistent when it belongs to a shadowed region of the picture taken from the view V_j but it is illuminated by L. Therefore, we define F_{shad} as follows

$$F_{shad}\left(P\right) = -i\left(P\right) \cdot n\left(P\right) \tag{8}$$

where $n\left(P\right)$ is the outer normal to the surface in P. $i\left(P\right)$ is a scalar function which is equal to 1 iff there exists a triad (O, L, V_j) such that P is inconsistent with it, 0 otherwise. Therefore, a shadow force is applied to all those points which are inconsistent with shadow information, pushing them inside the surface with intensity equal to k_{shad}. Thus, F_{shad} tends to eliminate shadow inconsistencies and, consequently, to minimize ξ_{shad}.

3. Experimental results

3.1. Reconstruction error and mesh quality

Reconstruction error can be evaluated if the measures of the original surface Λ are available. In this case, one of the possible metrics could be the measure of the volume difference between Λ and $\overline{\Lambda}$. More precisely, let S and $\overline{S}$ be three-manifolds, i.e., solids, such that their boundary ∂S and $\partial\overline{S}$ are equal to Λ and $\overline{\Lambda}$ respectively, then

$$\varepsilon = \frac{Volume\left(\left(S\backslash\overline{S}\right) \cup \left(\overline{S}\backslash S\right)\right)}{Volume\left(S\right)} \tag{9}$$

could be use as an evaluation of the reconstruction error.

Moreover, one can evaluate the distances between the two surfaces as follows

$$d\left(p,\Lambda\right) \qquad ,\forall p \in \overline{\Lambda} \tag{10}$$

where $d\left(p,\Lambda\right)$ is, obviously, the distance between p and Λ. It is convenient to make a statistics of these distances evaluating mean, standard deviation and maximum value ($d_{average}$, σ_d and d_{max} respectively).

Besides, in order to evaluate the quality of the obtained 3D mesh the quality parameters introduced in [4] can be very useful. Q_{equ} is the index of parametric regularity of a mesh face f

$$Q_{equ}\left(f\right) = \frac{6}{\sqrt{3}}\frac{A}{s \cdot h} \in [0,1] \tag{11}$$

where A is the area of the face, s the semi perimeter and h the length of the longest edge. Q_{equ} is a value between 0 and 1, where 1 corresponds to a equilateral triangle and thus to maximal regularity. Another quality index is Q_{plan}, which refers to the geometrical description of the mesh. Q_{plan} is defined as:

$$Q_{plan} = \frac{n \cdot n_1 + n \cdot n_2 + n \cdot n_3}{3} \in [0,1] \tag{12}$$

where n is the normal to the face and n_1, n_2, n_3 are the normals of the triangles which are adjacent to the three face edges. A good mesh must describe high curvature regions with a high sampling rate. Q_{plan} could be increased by sampling the mesh at high rate, but this would generate huge size models without enhancing the level of detail of low curvature regions. A sampling rate proportional to local curvature is therefore advisable.

3.2. Tests

The algorithm behavior, when only silhouette and stereo information is available, was tested and results were presented in our previous work [1]. Instead, in order to evaluate the capability to use shadow information in the reconstruction process, we performed tests on synthetic models. This is the only way of knowing exactly the original surface Λ, thus being able to evaluate reconstruction error. Pictures of the synthetic models were generated by a rendering software using n 43mm target cameras arranged in a circle centered on the object center, like in Fig.3. Lights in the scene are all omnidirectional and their shadow are generated using area shadow method which is more similar to reality than a simple ray-tracing method. In this way, we introduce an error in the determination of the shadow area which will be detected using the threshold mentioned above. The rendering was supervisioned by a script which also calculated the relative projection matrices.

Although the first model in Fig.4 seems very simple, it represents the typical object reconstructible only fusing together shadow, stereo and silhouette information. Indeed, its outer sides present a good texture quality which cover the incapability of the silhouette-based methods to precisely describe its geometry without using a large number of images of Λ. These methods indeed, does not behave correctly

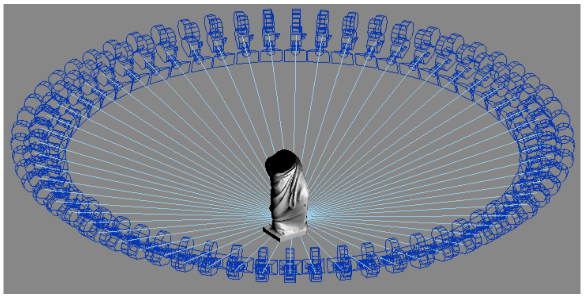

Figure 3. Arrangement of the cameras all around the synthetic model.

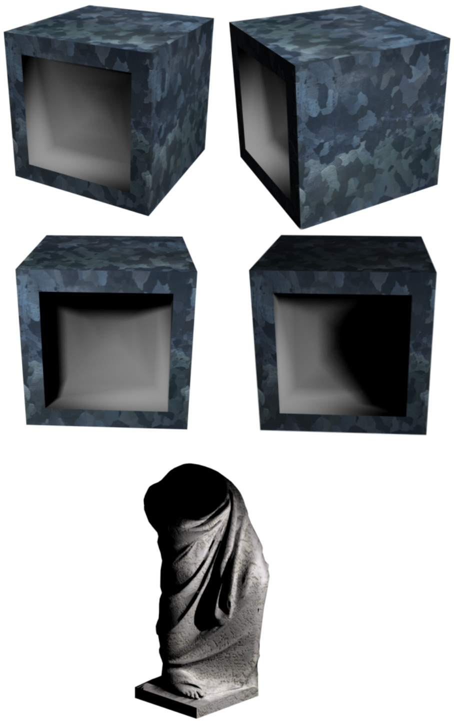

Figure 4. Synthetic models used to evaluate our system. Above: the cube. Below: a marble tunic

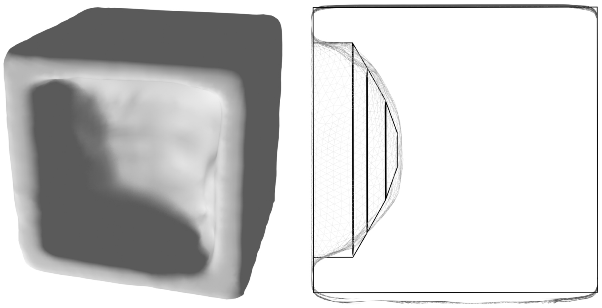

Figure 5. Left: Reconstruction of the synthetic model of Fig.4above. Right: Comparison with the original one ($\overline{\Lambda}$ is in grey).

in case of sharp cornered objects because their silhouettes produce a lot of redundant information. Using stereo information instead, we need only eight images to reconstruct the outer sides.

Inner sides of the concavity does not have any textures so neither stereo nor silhouette can describe this area. However, we have eleven photos of the concavity taken from the same point of view with different illuminations. More precisely, we have disposed eleven light sources which are switched on one at a time. Triads (O, L, V_j) are then extracted using shadow detection algorithm and finally passed to the fusion algorithm obtaining the result in Fig.5.

Cube size is 40x40x40 units and the cameras are 200 units away from the center of the object. Image resolution is 1024x768 pixels. Excluding the bottom region of the cube which is not visible by the camera, reconstruction error is equal to $\varepsilon = 0.57\%$ which means that we have a discrepancy of 0.57% of the original volume. d_{max} is equal to 2.14 units (3% of the diameter of the object) while $d_{average}$ is equal to 0.41 units (1% of the diameter) with standard deviation 0.30.

As we can see in Fig.6, compared with shadow carving algorithm, our method produces a model that presents a better parametric quality and a better geometric quality. More precisely, $\overline{Q_{equ}}$ and $\overline{Q_{plan}}$ are equal to 0.84 and 0.995 respectively.

4. Conclusions

This paper presents a new 3D passive multimodal digitization scheme which fuses together silhouette, texture and shadow information. As proved by tests, the proposed system presents both properties that mark a multimodal technique. Indeed, it can be proved to be resilient to measurement errors and capable of reconstructing a wide range of objects such as those featuring:

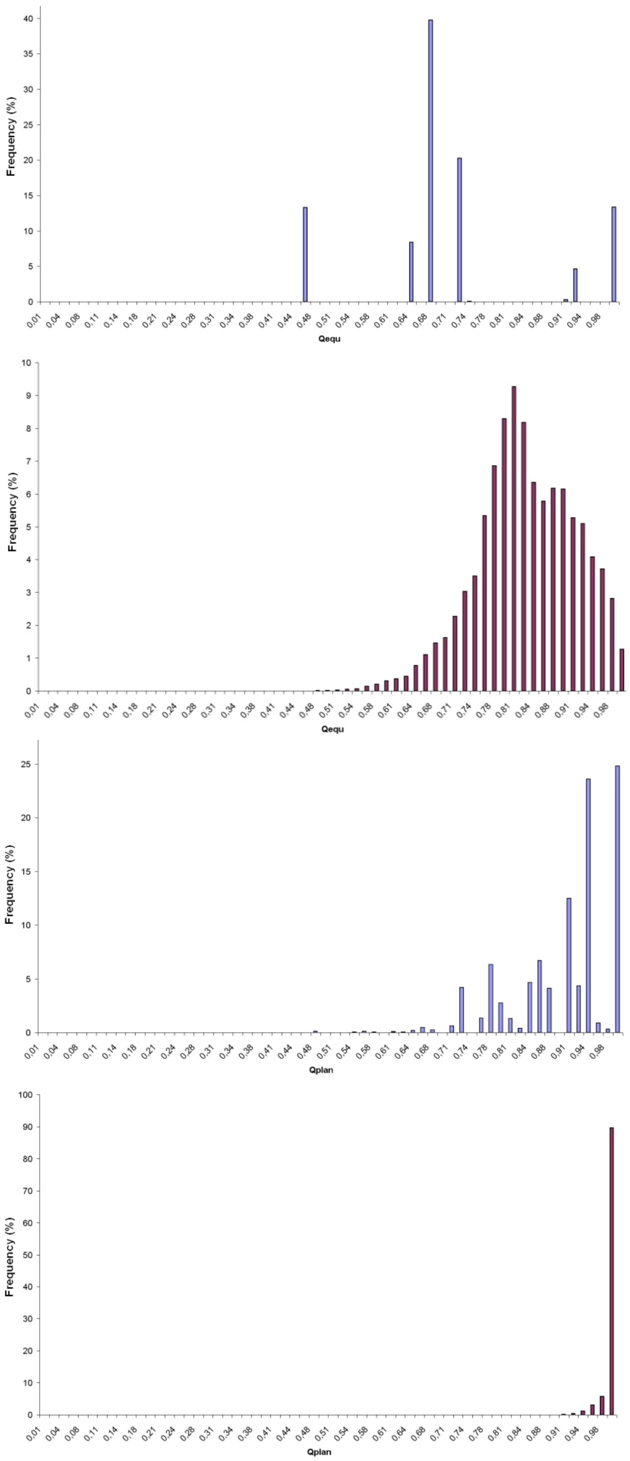

Figure 6. Model quality histogram comparison between the one obtained using shadow carving method S and the one obtained with our method R. From top to bottom: Q_{equ} of S, Q_{equ} of R, Q_{plan} of S and Q_{plan} of R.

- Surfaces characterized by good quality texture, sufficient lighting and not too high a specular reflectance;
- Specular surfaces or surfaces without texture or with a periodical texture, provided that some pictures image the profile of such surfaces (information in this case comes from the silhouettes);
- Concavities characterized by good texture and sufficient lighting.
- Concavities characterized by high reflectance and absence of lighting (information comes from the shadow regions).

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.

Furthermore, reconstruction error is rather satisfactory. For instance, using centimeters we have that the cube reconstructed from pictures 1024x768 taken from 2 meter distance with 43mm cameras is affected by an average error of 4 millimeters. Such an error can be remarkably reduced using digital cameras of higher resolution.

Moreover, the use of shadow information enlarges the range of reconstructible objects. Indeed, shadow information compensates for the lack of texture and silhouette information in the case of dark regions. Besides, differently from shadow carving algorithms, our method gives better results with respect to geometric and parametric quality. Other benefits of our system are inherited from our previous work [1].

Further research will concern the combination of shape from shading method together with silhouettes, stereo and shadow.

References

[1] L. Ballan, N. Brusco, and G. M. Cortelazzo. 3d passive shape recovery from texture and silhouette information. *CVMP 2005, London*, 2005.

[2] M. Desbrun, M. Meyer, P. Schroder, and A. H. Barr. Implicit fairing of irregular meshes using diffusion and curvature flow. *International Conference on Computer Graphics and Interactive Techniques*, pages 317–324, 1999.

[3] C. H. Esteban and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004.

[4] P. J. Frey and H. Borouchaki. Surface mesh quality evaluation. *International journal for numerical methods in engineering*, 45:101–118, 1999.

[5] A. Laurentini. The visual hull concept for silhouette based image understanding. *IEEE PAMI*, 16(2):150–162, 1994.

[6] L. Lucchese and S. K. Mitra. Color image segmentation: A state of the art approach. *Proc. of the Indian National Science Academy*, 67(2):207–221, march 2001.

[7] A. W. M. Kass and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1987.

[8] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proc. Royal Society of London*, 204:301328, 1979.

[9] Y. Matsumoto, K. Fujimura, and T. Kitamura. Shape-from-silhouette/stereo and its application to 3-d digitizer. *Proceedings of Discrete Geometry for Computing Imagery*, pages 177–190, 1999.

[10] T. Matsuyama, X. Wu, T. Takai, and S. Nobuhara. Real-time 3d shape reconstruction, dynamic 3d mesh deformation, and high fidelity visualiziation for 3d video. *Computer Vision and Image Understanding*, 96(3):393–434, 2004.

[11] G. V. Meerbergen, M. Vergauwen, M. Pollefeys, and L. V. Gool. A hierarchical symmetric stereo algorithm using dynamic programming. *International Journal of Computer Vision*, 47:275–285, 2002.

[12] S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*, volume 153 of *Applied Mathematical Sciences*. Springer, 2003.

[13] M. Potmesil. Generating octree models of 3d objects from their silhouettes in a sequence of images. *Computer Vision, Graphics, and Image Processing*, 40:1–29, 1987.

[14] N. Roma, J. Santos-Victor, and J. Tom. A comparative analysis of cross-correlation matching algorithms using a pyramidal resolution approach. *2nd Workshop on Empirical Evaluation Methods in Computer Vision, Dublin*, 2000.

[15] F. B. S. Savarese, H. E. Rushmeier and P. Perona. Implementation of a shadow carving system for shape capture. *3DPVT*, pages 12–23, 2002.

[16] S. Savarese, H. E. Rushmeier, F. Bernardini, and P. Perona. Shadow carving. *ICCV*, pages 190–197, 2001.

[17] S. N. Sinha and M. Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. *ICCV, Beijing, China*, 2005.

[18] R. Szeliski and D. Scharstein. Symmetric sub-pixel stereo matching. *Seventh European Conference on Computer Vision*, 2:525–540, 2002.

[19] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, pages 359–369, 1998.

[20] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. *ICCV, Nice, France*, 2003.

Proceedings of the Third International Symposium on
3D Data Processing, Visualization, and Transmission (3DPVT'06)
0-7695-2825-2/06 $20.00 © 2006 IEEE