Study of Volumetric Methods for Face Reconstruction

Gang Zeng, Sylvain Paris, Maxime Lhuillier, Long Quan

Study of Volumetric Methods for Face Reconstruction

Gang ZENG
Dep. of Computer Science
HKUST
Clear Water Bay, Kowloon, Hong Kong

Sylvain PARIS*
ARTIS† / GRAVIR-IMAG
INRIA Rhône-Alpes
38334 Saint Ismier, France

Maxime LHUILLIER
LASMEA, UMR CNRS 6602
Université Blaise-Pascal
63177 Aubière, France

Long QUAN
Dep. of Computer Science
HKUST
Clear Water Bay, Kowloon, Hong Kong

***Abstract*— This paper presents an early study about a reconstruction method that extends previous space carving methods to handle all characteristics of human faces: complex non-Lambertian materials, untextured areas, highly detailed geometry, etc. This face study is seen as a preliminary step to a more general framework for non-Lambertian reconstruction. We therefore avoid specific techniques like parameterized face models. Since we expose our early studies, we mainly present the related existing work and discuss the pros and cons of each approach. We aim at discerning the strength and weaknesses of the classical tools in order to adapt and improve them to handle non-Lambertian materials while overcoming their limitations. We show our first results which are promising and validate our global approach. Throughout the discussion, we raise several questions that identify complex issues related to our goal. We provide some hints that pave the way for a better understanding of the global problem. We are confident in that further studies will lead to significant improvements over existing methods.**

I. INTRODUCTION

Human faces and heads are challenging the classical three-dimensional surface reconstruction algorithms relying on the Lambertian hypothesis. Actually, most parts of the head (skin, hair, eyes, etc) are non-Lambertian: their aspect strongly changes with the viewpoint because of various phenomena like highlights, transparency, subsurface scattering, etc. Therefore the Lambertian assumption is obviously not respected and faces need specific algorithms to be reconstructed efficiently. We consider this facial reconstruction problem as a first step toward a general reconstruction problem for non-Lambertian objects. Thus, we aim at avoiding techniques which are too specifically designed for faces like [1], [2], [3], [4] (*e.g.* using a face parametric model, or specific assumptions like face symmetry). We therefore cast our goal into the more general context of non-Lambertian surface acquisition.

Handling the whole range of the non-Lambertian effects in a unified way still appears unreachable. Nevertheless, some specific methods have been developed to go beyond the Lambertian assumption. For instance, Szelisky and Golland [5] describe a method that allows objects to be non-opaque without coping with general transparency (refraction, absorption, etc). Oren and Nayar [6] build a theory for reflective surfaces: they differentiate *real* features (*i.e.* directly seen by the viewer) from virtual features (*i.e.* seen after a reflection) and are able to reconstruct the geometry of these two types of features. However, this method is limited to features and cannot be extended to dense geometry recovery. These example works show that surface materials that duplicate other materials either by reflection or by transparency are very difficult to manipulate even with very restrictive assumptions. To avoid such strong restrictions, we focus in this paper on opaque and non-reflective surfaces which are sufficient to modelize faces. Moreover, since this assumption is not too restrictive, we can foresee in the near future an extension to a broad range of materials: plastic, unpolished metal, cloth, etc.

In this paper, we expose our first reflection on this topic: we mainly discuss the relevant existing results which inspire us, and we outline the issues that determine our approach. We then describe the way that we have decided to explore. We finally show some early results that pave the way for future improvements.

II. PREVIOUS WORK

We review some existing results that appear to be relevant to our goal. The following description is split into three parts:

- First, traditional space carving methods for Lambertian objects.
- Second, the methods that specifically study the non-Lambertian aspect of a image sequence.
- Third, the tools which are classically for 3D reconstruction.

Each technique is discussed to characterize its pros and cons, especially to analyze how it can fit our Non-Lambertian approach without being to restrictive.

* The visit of Sylvain PARIS at HKUST has been supported by the Eurodoc program from "Région Rhône-Alpes".

† ARTIS is a team of the GRAVIR/IMAG laboratory, a joint effort of CNRS, INRIA, INPG and UJF.

A. Space carving under Lambertian assumptions

Space carving methods are interesting for their simplicity which opens opportunities for enhancements.

For Lambertian objects, the radiance of a surface point is equal in all directions. Under this assumption, photo-consistency is widely used for volumetric reconstruction methods [7], [8], [9]. The photo-consistency of a point is usually defined as the standard deviation of the colors of its projections in the input images. Then this value is compared with a threshold to determine whether the surface point is consistent with the input images. If not, the point is removed from the final result.

In a real scene, the object surfaces always contain various non-Lambertian effects. To overcome this, one may change the threshold used to characterize the consistent surface points. This achieves satisfying results for the objects with limited specularities but fails as soon as strong non-Lambertian objects are present. It yields either a much larger over-estimation due to the high threshold necessary to accept the view-dependent variations, or yields a over-carved shape because the specular regions are removed. However, the definitions of both photo-consistency and the threshold are simple enough to be easily extended. For instance, Slabaugh et al. [10] use a simulated annealing instead of a threshold.

Most of the space-carving-like methods employ the plane sweep techniques and the voxel representations [7], [8] to handle visibility. While the inconsistent voxels are carved one by one in near-to-far order from the cameras, the visibility of the new surface voxels is updated. This gives an accurate computation since it takes occlusion into consideration. Another option [11] is to use depth map to compute visibility.

B. Methods based on non-Lambertian assumptions

A practical idea to handle non-Lambertian objects is to perform a preprocessing to remove highlights and transform a non-Lambertian object into a Lambertian one. Li et al. [12] detect highlights as failures of a Lambertian technique. But the robustness of the detection is not clearly demonstrated. Lin et al. [13] use the same idea by first removing highlights from images with a color histogram difference. However, restrictive assumptions have to be made on the scene colors. As the highlights are handled in a binary mode, these methods implicitly assume that the surfaces have a strong shiny lobe and follow the Lambertian rule out of the highlight. These methods would most probably fail on materials like skin or mat plastic that do not match those constraints. Lin and Shum [14] refined the analysis of the color spectrum to single out the specular component from the images. The method can therefore be applied on a wide variety of materials but it still requires that the highlights do not spatially overlap, which is difficult to satisfy in short baseline stereo configurations.

Magda et al. [15] propose two original methods that exploit specularities instead of simply removing them. A specific light and camera setup is however needed. The highlights then become a information source rather than an obstacle. The requirement for the specific setup is too restrictive to be achieved in our context.

Bhat and Nayar [16] have studied the relationship between camera configuration, surface properties, and specular intensity variations. Their work is based on the specific model of Torrance-Sparrow. Although this model is quite large in its applications for image rendering, it restricts the practical use of analysis since it is quite hard to determine if a real surface is well approximated by this model. However it gives major qualitative ideas: the matching tolerance has to be higher with a wider baseline and/or with shinier surfaces.

C. Methods based on other image information

Besides photo-consistency, silhouette and cross-correlation are also widely used for surface reconstruction.

Silhouettes have been widely used to construct an approximate shape called visual hull [17], [18], [19]. More precise approaches [20], [21], [22] have been proposed to evaluated local curvature along occluding contours but they still seem numerically unstable. This shape-from-contour or silhouette approach is independent of any lighting condition since they do not consider surface appearance, they rely only on edges detected from images. More often, silhouettes are simply extracted from a known background using chroma-key technique.

Cross-correlation is usually computed over a larger neighborhood. It takes therefore local variations into account, it is very sensitive to local texture variation. Cross-correlation, particularly its zero-mean normalized version ZNCC, is more robust than photo-consistency as it is invariant to a local linear transformation of lighting. The computation of ZNCC is better if the surface orientation can be taken into account as it is operating on a larger window. The ZNCC between the locations at point $X_1 = (x_1, y_1)^T$ in the first image and at point $X_2 = (x_2, y_2)^T$ in the second image is defined to be

$$\frac{1}{n\sigma_1\sigma_2} \sum_{p \in \mathcal{N}(X_1)} (I_1(p) - \overline{I}_1)(I_2(\pi(p)) - \overline{I}_2)$$

where p is a point in the neighborhood $\mathcal{N}$ of X_1, n the number of points in $\mathcal{N}$, $\pi(p)$ is the corresponding point of p in the neighborhood of X_2 potentially accounting for perspective distortion, I_i is the intensity in image $i \in \{1; 2\}$, $\overline{I}_i$ and σ_i are the mean and standard deviation of the intensity in $\mathcal{N}(X_i)$.

Ishikawa [23] proposes an approach based on entropy of the criterion responses to make use of these different criteria. But it results in different criteria for different surface points and thus it makes almost impossible for any further optimization as the values at different points can no longer be compared.

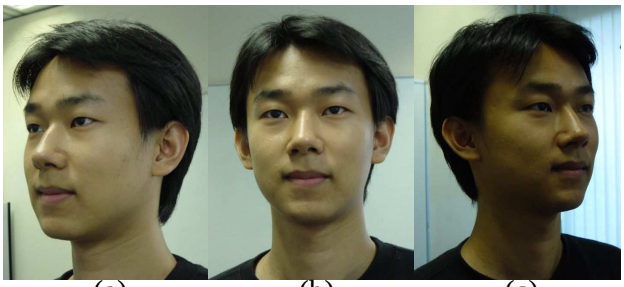
III. Discussion and overview

Considering all these various approaches, we can formulate some remarks and then try to sketch a first algorithm to face reconstruction.

First of all, most of the general methods exploiting the specificities of the appearance of a non-Lambertian scene in a sequence seem to require a too specific setup [15] and/or too restrictive hypotheses [16], [14], [12]. Those issues do not fit with our goal, we target a simple framework: we want to handle the scene "as is" without moving a light neither changing the background color. Nonetheless, some interesting points are worth considering: a non-Lambertian scene is almost a Lambertian scene but with some additional features like highlights, translucency, etc. These features can be either removed [14], [12] or accounted for [16]. We here follow this approach: non-Lambertian cases will be considered as an extension of the Lambertian ones. This implies that we are more likely to adapt and extend existing tools than to develop new ones.

Then among the existing tools, silhouettes and texture matching are more robust to non-Lambertian effects. However, curvature estimation using silhouettes [20], [21] is not stable enough to be used with real images, so we restrict to visual hull estimation [17], [18], [19]. Using this criterion alone would obviously bring nothing more than the existing methods which are proved to result in robust but poor approximation of the real objects. This introduces a new idea: silhouettes may be fruitfully combined with other criteria like photo-consistency or cross-correlation. On the one hand, photo-consistency is more sensitive to non-Lambertian artifacts and on the other hand, cross-correlation needs a surface orientation estimate to be computed. Our current view is to first use a mix between silhouettes and photo-consistency to build a shape estimation which is refined in a second pass using cross-correlation.

The use of photo-consistency is not trivial because this criterion directly stems from a Lambertian hypothesis: the color of a Lambertian surface does not depend on the viewpoint. To overcome that difficulty and extend this definition, we are inspired by the work of Bhat and Nayar [16] that shows to tweak thresholds in a Lambertian approach to take into account highlights and by the approach of Li et al. [12] that can be seen as an outlier classification which may lead to robust statistics [24]. Another approach to adapt photo-consistency is the work of Bhat and Nayar [16]: the use of a specific reflectance model to match the input data. Such a model may be helpful to extend photo-consistency: the color would be no longer expected to similar throughout the sequence but expected to match a given model. Nonetheless, this raises the new difficulty: To determine how the real material correspond to such a model. The question deserves an clear answer to be sure that the match error between the image sequence and the model is not due to the difference between the model and the real surface properties. For instance, Marschner et al. [25] have shown that skin cannot be approximated by classical models.



(a) (b) (c)

Fig. 1. Three of the input images. The image appearance changes quite a lot from one view to another; skin and hair are highly non-Lambertian due to numerous complicating effects.

The last issue to observe is the use of ZNCC. By estimating the surface orientation from the current shape, we could use two most front-facing visible cameras to compute the correlation value. This makes a more accurate texture matching. It is important to know that it is not perfect due to the irregularities and textureless regions. Therefore, we plan to use a regularizing approach like level sets [26] or graph cuts [27] which introduce some smoothing criteria that should compensate for the lack of texture. A complete review of these techniques is out of the scope of this paper.

Finally our approach can be summarized as follow:

- Silhouettes to build the visual hull.
- Photo-consistency to improve the visual hull. It may be adapted with robust statistics or with a specific reflectance model.
- ZNCC to refine the previous shape estimation. It may used with level sets or graph cuts to handle the irregularities and textureless regions.

IV. Algorithm and preliminary implementation

Today we have implemented a first demonstration algorithm which illustrates some of the main ideas developed in the previous section. We here describe roughly each step of the process.

A. Input

We usually require 25 images from a hand-held camera by completely turning around the object. The lighting and the background of the object are arbitrarily unknown. The geometry of the sequence is automatically computed and self-calibrated from a standard uncalibrated approach [28]. It is also important to notice that this kind of sequences is typically difficult for traditional space carving methods: the image appearance changes quite a lot from one view to another; skin and hair are well-known to be highly non-Lambertian due to numerous complicating effects (see Figure 1).

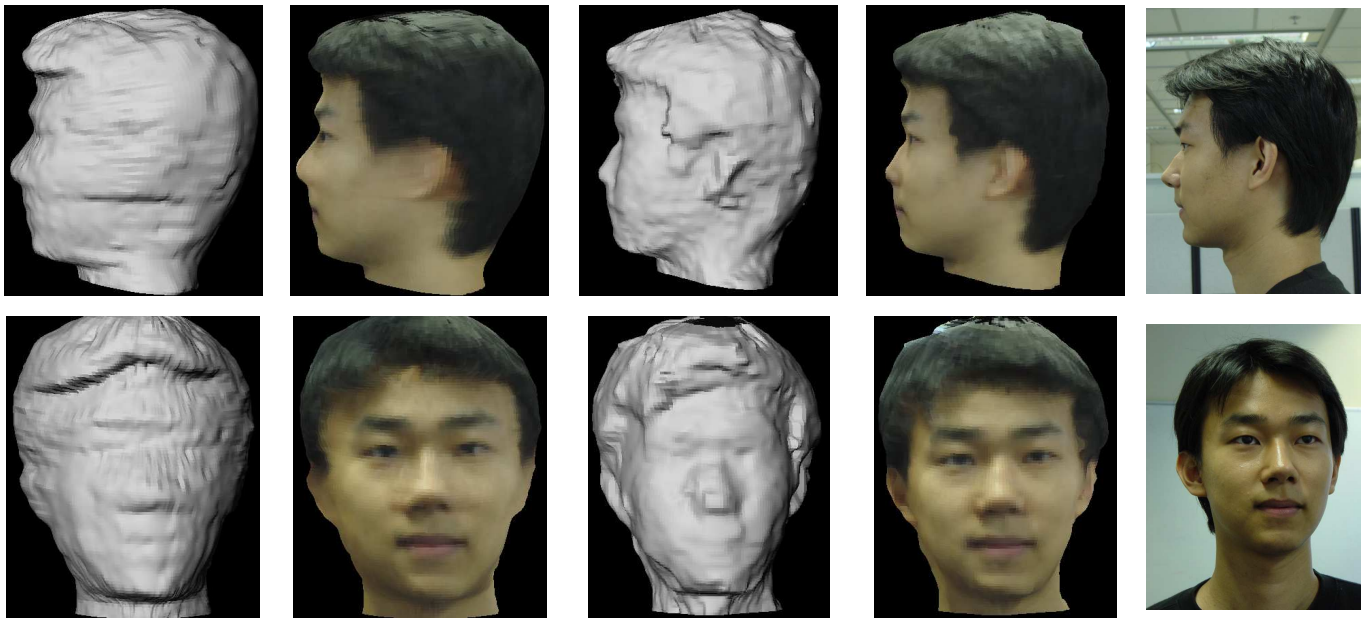


Fig. 2. Head reconstruction. Notice the various highlights. (a-b) The estimation. (c-d) Final result. (e) Comparison with original images.

B. Initialization

All this algorithm is based on a discrete voxel space. This discretization takes place in the classical cylindrical coordinate system (r, θ, h) which is suitable to parameterize a head as long as the axis is vertical and goes through the center of the head.

C. Visual hull

The visual hull is computed by intersecting the visual cones resulting from the objects silhouettes in the input images. These silhouettes are extracted by hand. We have also tested a new technique which automatically detects the silhouettes in images. This technique still needs to be formally and rigorously stated before being exposed in details. The underlying idea is to impose that all the silhouettes correspond to the same 3D object while being consistent with the input images.

D. Space carving

The previous visual hull is then carved using the photo-consistency criterion. In our current implementation, we use a rather naive approach to robust statistics: the outliers are determined relatively to the standard deviation of the color point set.

This leads to satisfying results but there is obviously room for improvements *e.g.* outliers of blue surface can be red which does seem reasonable under common circumstances.

We have also tried to match a Phong reflectance model. This also results in satisfying shapes but the overall consistency of the model is still not ensured: parameters can change from point to point which is not an acceptable behavior. We have therefore kept the robust statistics approach until we have a better technique.

E. Refinement with ZNCC

Finally the above estimation is refined with a graph-cut technique [29] driven by ZNCC. This optimization technique solves exactly the problem by providing the global minimum cost function.

The advantage of this technique is that it adds a regularization term through the derivatives that smooth the surface in textured regions and ensure continuity in textureless areas.

ZNCC is computed using the two most front-parallel visible cameras with an 11×11 window leading to acceptable results.

Nonetheless, this technique still deserves in-depth studies to handle events like occlusions or topological changes.

V. Results

Figure 2 shows the reconstruction of the head of the first author, including the estimation and the final optimized shape. The role of each step of the algorithm is clearly put into evidence. The estimation (Figure 2-a,b) is built by silhouettes and photo-consistency, which gives very robust but yet not sufficiently accurate localization of the object. Then detailed surface geometry (Figure 2-c,d) is carved out by the second

step of graph-based optimization driven by cross-correlation. We see particularly the very accurate recovery of concavities around the hairs and eyes areas.

One important feature of our technique is that it reconstructs a full head rather than a face. This implies handling hair which is highly non-Lambertian. Moreover, notice that our images are fairly taken under unknown background and complex lighting environment. These first results are promising and encourage us to explore further this approach.

VI. Conclusions

We have exposed the first steps of a novel approach to reconstructing faces and even entire heads from an arbitrary set of calibrated cameras.

An interesting point is that we make very few assumptions about the surface properties so we can expect to handle very general materials in the near future. Our approach mixes various traditional techniques while improving each of them to ensure a compatibility with the non-Lambertian surfaces. In the case of face reconstruction, this results in a satisfying head estimation which includes both photometric and contour information. This shape is finally refined thanks to the use of an optimization process driven by the local texture of the objects. All these points form a consistent set of tools which is very promising for the near future. Moreover, we have raised numerous questions that initiate a fruitful reflection which will hopefully result in powerful techniques.

The method has been shown to be especially efficient on human heads which are well-known to be highly non-Lambertian in many aspects (mainly skin and hair). Our approach is very general in its formulation: we aim at widening its applicability to propose a general method to reconstruct non-Lambertian objects.

References

[1] Y. Shan, Z. Liu, and Z. Zhang, "Model-Based bundle adjustment with application to face modeling," in *International Conference On Computer Vision (ICCV 01)*, 2001, pp. 644–651.

[2] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *SIGGRAPH*, 1999.

[3] A. Georghiades, P. Belhumeur, and D. Kriegman, "Illumination-based image synthesis: Creating novel images of human faces under differing pose and lighting," in *IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes*, 1999.

[4] B. Guenter, C. Grimm, D. Wood, H. Malvar, and F. Pighin, "Making faces," in *SIGGRAPH*, 1998.

[5] R. Szeliski and P. Golland, "Stereo matching with transparency and matting," in *International Conference on Computer Vision (ICCV 98)*, 1998.

[6] M. Oren and S. K. Nayar, "A theory of specular surface geometry," in *International Conference on Computer Vision (ICCV 95)*, 1995.

[7] S. M. Seitz and C. R. Dyer, "Photorealistic scene reconstruction by voxel coloring," *International Journal of Computer Vision (IJCV 99)*, 1999.

[8] K. N. Kutulakos, "Approximate n-view stereo," in *European Conference on Computer Vision (ECCV 00)*, 2000.

[9] G. Slabaugh, B. Culbertson, T. Malzbender, and R. Schafer, "A survey of methods for volumetric scene reconstruction from photographs," in *VolumeGraphics 01*, 2001.

[10] G. Slabaugh, B. Culbertson, T. Malzbender, and R. Schafer, "Improved voxel coloring via volumetric optimization," Center for Signal and Image Processing, Georgia Institute of Technology, Tech. Rep., 2000.

[11] W. Culbertson, T. Malzbender, and G. Slabaugh, "Generalized voxel coloring," in *Vision Algorithms: Theory and Practice*, ser. Lecture Notes in Computer Science, B. Triggs, A. Zisserman, and R. Szeliski, Eds., vol. 1883. Springer-Verlag, 2000, pp. 100–114.

[12] Y. Li, S. Lin, H. Lu, S. Kang, and H. Shum, "Multi-baseline in presence of specular reflections," in *International Conference on Pattern Recognition (ICPR'02)*, 2002.

[13] S. Lin, Y. Li, S. B. Kang, X. Tong, and H. Shum, "Diffuse-specular separation and depth recovery from image sequences," in *European Conference on Computer Vision (ECCV 02)*, 2002.

[14] S. Lin and H. Shum, "Separation of diffuse and specular reflection in color images," in *Computer Vision and Pattern Recognition (CVPR'01)*, 2001.

[15] S. Magda, T. Zickler, D. Kriegman, and P. Belhumeur, "Beyond lambert: Reconstructing surfaces with arbitrary brdfs," in *International Conference on Computer Vision (ICCV 01)*, 2001.

[16] D. N. Bhat and S. K. Nayar, "Stereo in the presence of specular reflection," in *International Conference on Computer Vision (ICCV 95)*, 1995.

[17] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *Transactions on Pattern Analysis and MAchine Intelligence (TPAMI 94)*, 1994.

[18] W. Matusik, C. Buehler, R. Raskar, L. McMillan, , and S. Gortler, "Image-based visual hulls," in *SIGGRAPH*, 2000.

[19] W. Matusik, C. Buehler, R. Raskar, and L. McMillan, "Polyhedral visual hulls for real-time rendering," in *SIGGRAPH*, 2001.

[20] R. Vaillant and O. D. Faugeras, "Using extremal boundaries for 3-D object modeling," *Transactions on Pattern Analysis and machine Intelligence (TPAMI 92)*, 1992.

[21] R. Szeliski and R. Weiss, "Robust shape recovery from occluding contours using a linear smoother," Digital Equipment Corporation, Cambridge Research Lab, Tech. Rep. DEC-CRL-93-7, 1993.

[22] K.-Y. K. W. Paulo R. S. Mendona and R. Cipolla, "Camera pose estimation and reconstruction from imag profiles under circular motion," in *European Conference on Computer Vision (ECCV 00)* , vol. 2, 2000, pp. 864–878.

[23] H. Ishikawa, "Global optimization using embedded graphs," Ph.D. dissertation, New York University, 2000.

[24] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics*. Wiley, 1986.

[25] S. Marschner, E. Lafortune, S. Westin, K. Torrance, and D. Greenberg, "Image-based BRDF measurement," Program of Computer Graphics, Cornell University, Tech. Rep. PCG-99-1, January 1999.

[26] O. Faugeras and R. Keriven, "Variational principles, surface evolution, PDE's, level set methods and the stereo problem," *Transactions on Image Processing*, 1998.

[27] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *European Conference on Computer Vision (ECCV 02)*, 2002.

[28] M. Lhuillier and L. Quan, "Quasi-dense reconstruction from image sequence," in *European Conference on Computer Vision (ECCV 02)*, 2002. vol. 2, 2002, pp. 125–139.

[29] S. Paris, F. Sillion, and L. Quan, "A volumetric reconstruction method from multiple calibrated views using global graph cut optimization," INRIA, Tech. Rep. 2003.