Robust Carving for Non-Lambertian Objects

Gang ZENG[1] Sylvain PARIS[2] Long QUAN[1]
[1] Dep. of Computer Science, HKUST, Clear Water Bay, Kowloon, Hong Kong
{zenggang, quan}@cs.ust.hk
[2] ARTIS/ GRAVIR-IMAG, INRIA Rhône-Alpes, 38334 Saint Ismier, France
sylvain.paris@imag.fr

Abstract

This paper presents a new surface reconstruction method that extends previous carving methods for non-Lambertian objects by integrating the smoothness and image information of different aspects. We introduce a robust multi-view photo-consistency function and a single-view visual-consistency function. The former considers a general specularity that may introduce both reflectance models and robust statistics, while the latter is based on pixel homogeneity and continuity in individual images. The new consistent shape, with both multi-view and single-view, extends previous photo hull and visual hull, and is proven to be included in both photo hull and visual hull. This approximate shape is then refined by a graph-cut optimization method. Sample reconstructions from real sequences for both global and local optimizations are demonstrated.

1. Introduction

Three-dimensional surface reconstruction from images is a fundamental but challenging problem and has been intensively studied. Most existing techniques are based on the Lambertian assumption: the intensity values of reflected radiance of a surface point are equal in all directions following a cosine rule, producing view-independent effects. However, this rule is clearly insufficient for most common materials that have more complex behaviors like highlight, reflection, transparency or subsurface scattering. Handling all these effects in a unified way still appears unachievable, as some of them duplicate the characteristics of other objects. This paper focusses on opaque and non-reflective objects and proposes a general approach without any specific constraint. The range of materials we consider is therefore broad: plastic, unpolished metal, general cloth, skin, etc. We first describe some existing work without being exhaustive.

Methods based on non-Lambertian assumptions A practical way of handling non-Lambertian objects is to remove highlights to transform them into Lambertian ones. Lin et al. [13] remove highlights from images with a color histogram difference. Magda et al. [15] propose two original methods that exploit specularity instead of simply removing its. However, for their method, a specific light and camera setup is needed. Carceroni and Kutulakos [6] propose a method of recovering a 3D shape using dynamic surfel. Georghiades [9] proposes a surface reconstruction method based on the Torrance-Sparrow reflectance Model. Bhat and Nayar [2] have studied the relationship among camera configurations, surface properties, and specular intensity variations. Their work is based on a specific model, and gives major qualitative ideas: the matching tolerance has to be higher with a wider baseline and/or with shinier surfaces.

Methods based on different image information Silhouette is employed to construct an approximate shape called visual hull [12,16]. More precise approach [22] is proposed to evaluate local curvatures along occluding contours. Photo-consistency is widely used for volumetric reconstruction methods [11, 18, 19] by comparing colors under the Lambertian assumption. However, photo-consistency is not robust, it fails as soon as strong non-Lambertian objects are present. Recently, several methods [1,3,5,7] have been proposed to improve the above carving methods based on a probabilistic framework. Texture correlation considers texture similarities and can be used in level set methods [8] or graph-cut methods [4, 10, 17, 23]. Cross-correlation, particularly its zero-mean normalized version, ZNCC, is more robust as it is invariant to local linear transformation of lighting. The computation of ZNCC is better if the surface orientation is taken into account since it is operating on a larger window.

Overview of our approach We propose a novel approach to integrate various techniques into two steps to combine their advantages while avoiding their drawbacks. We first use multi-view and single-view consistency criteria to reconstruct an overall shape of the object without any previous information but lacks precision. We then use ZNCC to refine this shape into a finely detailed surface. This second step relies on the orientation information of the shape in the first step. We also introduce two novel graph-cut techniques that make the refinement more precise and robust.

2. Robust multi-view photo-consistency

The first step toward reconstructing non-Lambertian objects is to define a robust carving criterion from multiple views using reflectance models. We choose the Phong reflectance model for our first attempt. Although it is not a physical model, it is simple and works correctly under the specificity of our setup. In the case of surface reconstruction from multiple images, we combine the Lambertian model and the Phong model to describe both view-independent and view-dependent effects:

$$I(\mathbf{v}) = \alpha + \sum \beta_i (\mathbf{r}_i \cdot \mathbf{v})^\sigma$$

where $\mathbf{v}$ is the unit vector in the viewing direction and $\mathbf{r}_i$ is the unit vector in the mirror reflection (maximum reflection) direction of each light source, with α and β_i controlling the quantity of the Lambertian and non-Lambertian effects. σ indicates the shininess of the material.

Like traditional photo-consistency, the criterion of multi-view consistency is defined as the mean deviation between the colors predicted by the reflectance model and the colors actually seen from the images:

$$MVC = \min_{\alpha,\beta_i,\sigma} \left\{ \sqrt{\frac{1}{k} \sum (I_j - I(\mathbf{v}_j))^2} \right\}$$

where k is the number of visible cameras, with I_j being the projection on the jth visible image and $\mathbf{v}_j$ being the corresponding viewing direction.

In our algorithm, σ is selected from a set of user-defined parameters. For example, we need two parameters for head reconstructions to indicate the shininess of hair and skin separately (e.g. [21] provides some typical values for skin). We also notice that in most cases, one light source is enough to model the reflection because there is only one significant highlight. Thus, we approximate $\mathbf{r}$ by the viewing direction of the camera in which the point appears the brightest.

Finally, we simply ignore the worst matches to allow a significant but not major deviation from the Phong model. We handle image noise: we estimate the criterion through a process that starts with the above classical form and refines it by iteratively singling out samples whose distance to the current model is higher than the others. We will consider more accurate models in our future work.

$$MVC_r = \min_{\alpha,\beta,\sigma} \left\{ \sqrt{\frac{1}{k-r} \sum w_j (I_j - I(\mathbf{v}_j))^2} \right\}$$

$$with \;\; w_j = 0, 1 \;\; and \;\; \sum w_j = k - r$$

3. Single-view visual-consistency

Even if multi-view photo-consistency is made robust as above, one of the major problems is that the continuity of the object has never been taken into consideration. One of the possibilities for enforcing the continuity is to go back to individual images, as these sophisticated non-Lambertian view-dependent effects are stable in individual images. Our goal is then to enforce that, in every image, the silhouette of the object corresponds to some image discontinuities. This is what we call *single-view consistency* [24].

We combine the multi-view and single-view consistency as follows. When a voxel is found to be inconsistent and carved by multi-view, we check immediately whether the current shape is single-view consistent or not. Since we have an over-estimated 3D shape, all projected 2D silhouettes are also over-estimations of the object boundaries and thus they have to move inward to ensure that crossed color-homogeneous regions become purely background regions. Therefore, the parts outside the "visual hull" of the shrunken silhouettes are carved, and the new 3D estimation is enforced to inside this hull. The algorithm terminates when all surface voxels are multi-view consistent, and we then have a consistent shape, called *photo-visual hull*.

4. Surface refinement by texture correlation

The photo-visual hull gives a tight estimation of the object, which also provides visibility and geometry information for the surface points. To obtain a more detailed surface, we let the current surface evolve in the surrounding narrow band and aim at finding the surface for which the texture correlation gives the best score. This score has been adapted to account for the non-Lambertian specificity of our technique. We choose the Zero-mean Normalized Cross-Correlation (ZNCC) as it is invariant to a local linear transformation of lighting. Thus, it gives an accurate measurement of the texture correlation even if view-dependent effects are present. Based on the visibility and geometry information of the photo-visual hull, we select the two most front-facing visible cameras to compute this correlation score in order to limit the perspective distortion.

However, we cannot straightforwardly select the best matched points because they may introduce irregularity. This caveat is solved by the introduction of a smoothing constraint through an optimization process which removes the irregularity according to the neighborhood.

4.1. Global optimization

We choose the general graph-cut optimization introduced in [20]. In the case of a single object with surrounding viewpoints, we naturally employ a cylindrical parametrization (h, θ, r) of the object and aim at finding an optimal surface $r(h, \theta)$ minimizing the following functional:

$$\iint_{\mathcal{N}} \left(c(h, \theta, r(h, \theta)) + a \left| r_h(h, \theta) \right| + \frac{b}{r} \left| r_\theta(h, \theta) \right| \right) r \, d\theta \, dh \tag{1}$$

where $c(\cdot)$ is the consistency term of the surface point, with a and b controlling the smoothing constraint.

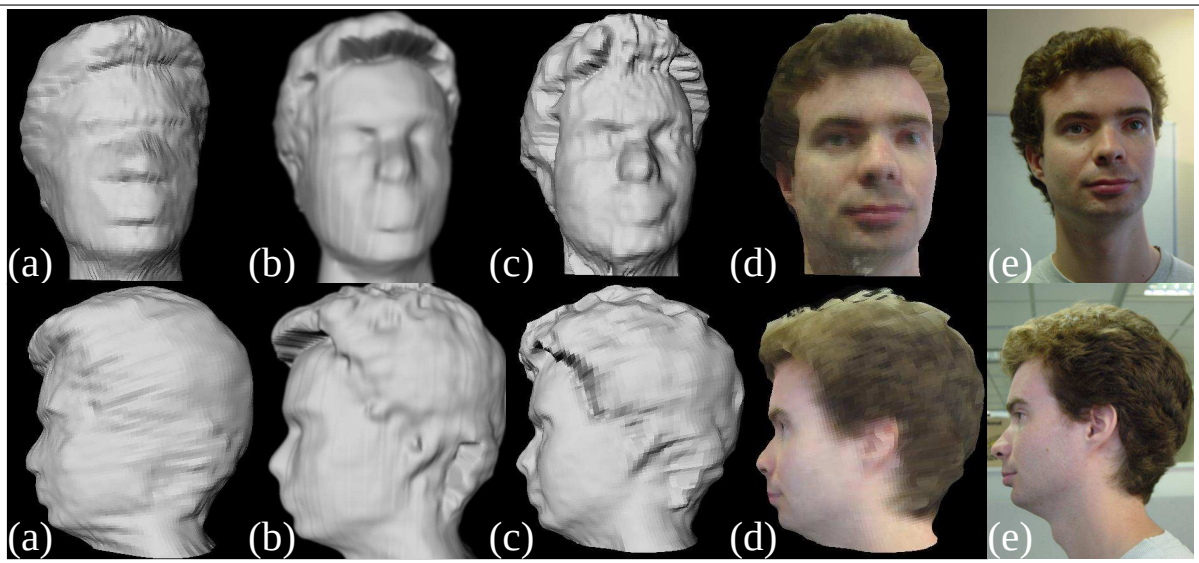

Figure 1. (a) Photo-visual hull. (b) Global optimization. (c-d) Local optimization. (e) Comparison with inputs.

This technique introduces a 3D embedded graph whose main property is the equivalence between its *minimal cut* and the optimal surface of the functional equation (1). Its purely geometric formulation makes its adaptation to our setup straightforward.

4.2. Local optimization

Another possibility is to enforce the optimization to run in a local context since sub-parts of the object may be independent and therefore can be reconstructed separately. For instance, in the case of head reconstructions, the hair can be reconstructed without considering the eyes. Thus, we run local optimizations several times in different positions instead of a global optimization. Although these local optimizations are run independently, adjacent optimizations are linked since their neighborhoods largely overlap.

Based on the tight estimation, $r_e(h, \theta)$, we perform the following local optimization for every surface point $(h_i, \theta_i, r_e(h_i, \theta_i))$. We first define its local context $\mathcal{N}$ to be of size $\triangle h \times \triangle\theta \times \triangle r$ and centered at this point. Then, the graph-cut optimization [20] is performed within this local context by minimizing the functional (1). Finally, the value $r(h_i, \theta_i)$ is assigned according to the center of this optimal surface patch.

5. Experimental results and discussions

We usually acquire about 35 images roughly in all directions with a hand-held camera by circumnavigating the object. The lighting and the background are arbitrarily unknown. The geometry of the sequences is automatically computed and self-calibrated from a standard uncalibrated approach [14]. It is also important to note that these kinds of sequences are typically difficult for traditional carving methods: the image appearance changes quite a lot from one view to another; skin and hair are widely known to be highly non-Lambertian due to numerous complex effects. Our implementation runs successfully on many examples. Here we show two typical head modeling examples, one with a more textured face and the other with less. The reconstruction results are shown in Figures 1 and 2 including the tight estimation and two optimized shapes. The role of each step of the algorithm is clearly put into evidence. The tight estimation (fig. 1,2-a) gives a very robust but yet not sufficiently accurate location of the object. The single-view consistency has segmented out the background while the multi-view consistency recovers some concavities. This has been achieved fully automatically without any prior information of the background. Then, detailed surface geometry (fig. 1,2-b,c,d) is carved out in the second step using graph-cut optimizations. We see particularly the very accurate recovery of concavities around the hair and eye areas.

Global optimization vs local optimization Comparing the global optimization (fig. 1,2-b) with the local optimization (fig. 1,2-c), we see the smoothness of the former and the sharpness of the latter. The global one yields more reasonable results on smooth regions, including the forehead and the cheeks, but fails to get details on sharp regions, including the nose, the eyes, the mouth and the boundary of the hair. The local one produces detailed geometry with local continuity, resulting in satisfactory surfaces.

6. Conclusions

We have proposed a novel approach for reconstructing non-Lambertian objects from an arbitrary set of calibrated images. To achieve our goal, we first generalize the photo-consistency criterion into a robust multi-view consistency, which combines reflectance models and robust statistics.

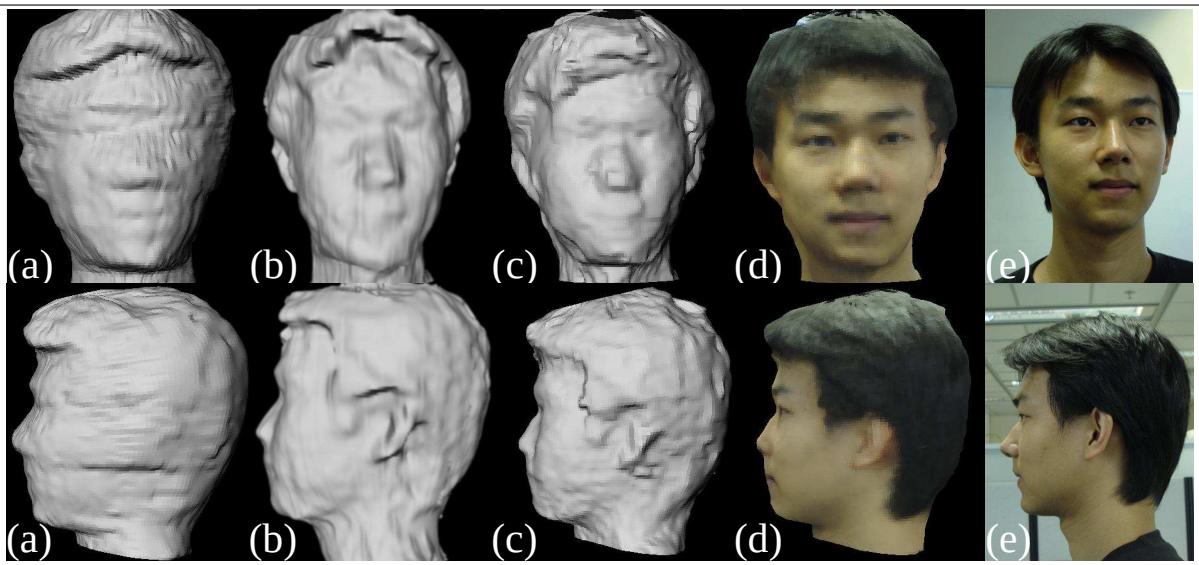

Figure 2. (a) Photo-visual hull. (b) Global optimization. (c-d) Local optimization. (e) Comparison with inputs.

Then, we introduce a new single-view aspect, which intrinsically overcomes view-dependent caveats. This is mainly related to robust computation of silhouettes and appears as a complementary tool to multi-view consistency. The above two steps result in a tighter shape estimation which includes both photometric and contour information. This shape is finally refined thanks to the usage of graph-cut optimizations driven by texture correlation. This approach has been validated on various reconstructions, and is shown to be especially efficient on human heads, which are widely known to be highly non-Lambertian.

Acknowledgements

This project is supported by the Hong Kong RGC grant HKUST 6188/02E.

References

[1] M. Agrawal and L. Davis. A probabilistic framework for surface reconstruction from multiple images. *CVPR 01*.

[2] D. Bhat and S. Nayar. Stereo in the presence of specular reflection. *ICCV 95*.

[3] R. Bhoitka, D. Fleet and K. Kutulakos. A probabilistic theory of occupancy and emptiness. *ECCV 02*.

[4] Y. Boykov and V. Kolmogorov. Computing Geodesics and Minimal Surfaces via Graph Cuts. *ICCV 03*.

[5] A. Broadhurst, T. Drummond and R. Cipolla. A Probabilistic Framework for Space Carving. *ICCV 01*.

[6] R. Carceroni and K. Kutulakos. Multi-view scene capture by surfel sampling. *ICCV 01*.

[7] J. DeBonnet and P. Viola. Poxels: Probabilistic voxelized volume reconstruction. *ICCV 99*.

[8] O. Faugeras and R. Keriven. Complete dense stereovision using level set methods. *ECCV 98*.

[9] A. Georghiades. Incorporating the Torrance and Sparrow Model of Reflectance in Uncalibrated Photometric Stereo. *ICCV 03*.

[10] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. *ECCV 02*.

[11] K. Kutulakos and S. Seitz. A theory of shape by space carving. *ICCV 99*.

[12] A. Laurentini. The visual hull concept for silhouette-based image understanding. *TPAMI 94*.

[13] S. Lin, Y. Li, S. Kang, X. Tong, and H. Shum. Diffuse-specular separation and depth recovery from image sequences. *ECCV 02*.

[14] M. Lhuillier and L. Quan. Quasi-dense reconstruction from image sequence. *ECCV 02*.

[15] S. Magda, T. Zickler, D. Kriegman, and P. Belhumeur. Beyond lambert: Reconstructing surfaces with arbitrary BRDFs. *ICCV 01*.

[16] W. Matusik, C. Buehler, R. Raskar, S. Gortler and L. McMillan. Image-based visual hulls. *SIGGRAPH 00*.

[17] S. Roy. Stereo without epipolar lines : A maximum-flow formulation. *IJCV 99*.

[18] S. Seitz and C. Dyer. Photorealistic scene reconstruction by voxel coloring. *IJCV 99*.

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

[20] S. Paris, F. Sillion and L. Quan. A Surface Reconstruction Method Using Global Graph Cut Optimization. *ACCV 04, extended version submitted to IJCV*.

[21] S. Paris, F. Sillion and L. Quan. Lightweight Face Relighting. *Pacific Graphics*, 2003.

[22] R. Vaillant and O. Faugeras. Using extremal boundaries for 3-D object modeling. *TPAMI 92*.

[23] G. Zeng, S. Paris, L. Quan and M. Lhuillier. Surface Reconstruction by Propagating 3D Stereo Data in Multiple 2D Images. *ECCV 04*.

[24] G. Zeng and L. Quan. Silhouette extraction from multiple images of an unknown background. *ACCV 04*.