Combining Structure from Motion and Photometric Stereo: A Piecewise Formulation

Reza Sabzevari, Alessio Del Bue, and Vittorio Murino

PAVIS Department, Istituto Italiano di Tecnologia, Genova, Italy

Abstract

This paper introduces a novel piecewise formulation for Photometric Stereo on static scenes exploiting a set of multiple view constraints on the surface. Such constraints will help to recover geometrical properties of the object and to eliminate the global bas-relief ambiguity. On the other hand, the proposed piecewise formulation will help to model more complex photometric properties of the surfaces using local lighting models. The experimental results show performance of the proposed method against 3D ground truth.

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Photometry; I.4.5 [Image Processing and Computer Vision]: Reconstruction—Transform methods;

1. Introduction

One of the most challenging scenarios in 3D reconstruction is recovering shape using images taken from different views with different cameras under unknown lighting conditions. Many approaches have been proposed to solve this problem exploiting different cues and considering various constraints to simplify such general problem [HVC08]. Photometric Stereo (PS) computes dense surfaces by analyzing the variations of pixel values with respect to different light sources up to an unknown bas-relief ambiguity. In contrast, Structure from Motion (SfM) [Sze11] can recover very accurate 3D models from a sparse set of image correspondences but it can not recover dense surfaces. Some works have been done on dense 3D reconstruction using both SfM and PS [JK07], but the proposed methods have introduced a set of strong constraints and/or have problem with complex object surfaces. The proposed pipeline combines SfM with a novel piecewise formulation of PS to recover a dense surface preserving the metric properties of the shape from uncalibrated camera under uncalibrated light sources. Comparing to other methods used PS to recover dense surfaces, our proposed Piecewise Photometric Stereo (PPS) can model more complex photometric properties of the surface thanks to individual photometric models assigned to each surface patch.

2. Pipeline Architecture

The proposed scheme uses two different sequences of images and it is summarized in Fig. 1. A multi-view sequence, imaging a static object from different views, is used to track a set of 2D features from the object's surface among different frames. The resulted set of 2D correspondences is used to recover a sparse 3D point cloud using SfM. Then the 3D point cloud is used to create a 3D mesh employing Delaunay triangulation. The second sequence has images from fixed viewpoint with varying lighting condition and it is used to recover the dense surface of the shape. The 3D mesh is projected on the multi-illumination sequence and the triangular image patches are decomposed from that sequence. Afterward, the proposed PPS is applied to images from multi-illumination sequence and an individual photometric model is assigned to each patch to recover its dense surface. At the final stage, the surface constraints defined by the 3D mesh are used to assemble the surface patches taking the advantage of the proposed photo-geometric alignment. Our piecewise formulation of PS is based on the first order spherical harmonic representation introduced in [BJK07]. A set of binary masks are defined to decompose the images in m different patches and to formulate m PS problems in a way that $S = H_1 \cdot S_1 + \cdots + H_m \cdot S_m$, where S is the complete surface, S_i $(i = 1 \cdots m)$ is surface of patch i and H_i is the photo-geometrical transformation required to align S_i on the 3D mesh. The surface patches are recovered in parallel, which makes PPS faster than classical PS, while the photo-geometric transformations are computed via a linear system of equations representing the relations between vertices of triangular surface patches and corresponding vertices on the 3D mesh.

DOI: 10.2312/conf/EG2012/posters/033-034

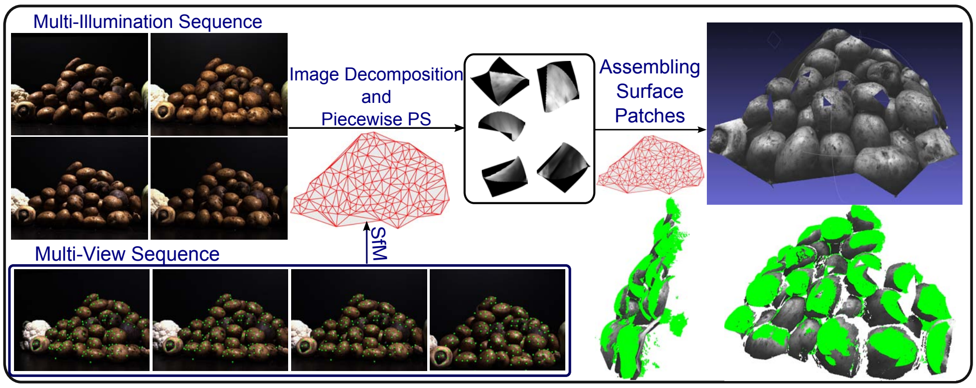

Figure 1: *The PPS pipeline: 2D features are tracked among the multi-view sequence to recover the shape geometry and PPS is applied to multi-illumination sequence to recover the surface details. Results are compared with ground truth (in green).*

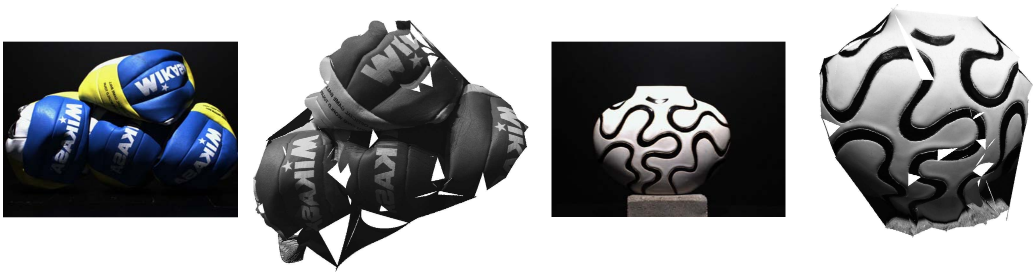

Figure 2: *Results on Jar and Ball sequences: the reconstruction is presented together with one image from the sequence.*

3. Results

The proposed 3D reconstruction pipeline is applied on sequences of the Robot data set [ADP11]. The data set is completely uncalibrated and includes images with 30 different views and 19 different lighting conditions for each view. Each single image has a frame size of 1600×1200. All the available views are used to recover the 3D mesh. The frontal views with 10 different lighting conditions are used as the multi-illumination sequence. The data set provides 3D ground truth given by a structured light acquisition system. Fig. 1 shows results of the proposed pipeline on one of the most challenging samples of this data set. The scene has a very sophisticated surface with complex surface bendings and strong shadow effects. Fig. 2 shows more results on two other sequences of this dataset. In this pipeline, selection of features and mesh generation is completely automatic which may cause improper selection of feature points. This effect will make the system unable to model photometric properties of a few patches, causing some spikes on the surface that can be easily detected by checking the surface relief. Since these patches are relatively small considering the scale of the images, they can be replaced with flat triangles without a major affect on the global surface. In Fig. 1 and 2, these patches are shown as some holes on the reconstructed surface for more visibility. Evaluating the results with ground truth, we obtained an average 3D error of 11.22% for the Potatoes sequence using the PPS algorithm, whereas the classic PS approach (applied to the whole images and without decomposition) results to an error of 17.34%. For the Jar and Ball sequence the proposed pipeline reduced the reconstruction error to 8.69% and 10.35% respectively. Such results prove that the depth of obtained surface is improved with respect to the metric properties of the scene.

Further improvement of this method could be considered as a better image decomposition, taking into account the photometric information of each patch.

References

[ADP11] Aanæs H., Dahl A., Pedersen K. S.: Interesting Interest Points. *International Journal of Computer Vision* (2011), 1–18. 2

[BJK07] Basri R., Jacobs D., Kemelmacher I.: Photometric Stereo with General, Unknown Lighting. *International Journal of Computer Vision 72*, 3 (2007), 239–257. 1

[HVC08] Hernández C., Vogiatzis G., Cipolla R.: Multiview photometric stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 30*, 3 (2008), 548–554. 1

[JK07] Joshi N., Kriegman D.: Shape from varying illumination and viewpoint. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on* (2007), IEEE, pp. 1–7. 1

[Sze11] Szeliski R.: *Computer vision: Algorithms and applications*. Springer-Verlag New York Inc, 2011. 1