

Unsupervised visual hull extraction in space, time and light domains [☆]



Maxim Mikhnevich ^{a,*}, Patrick Hébert ^b, Denis Laurendeau ^a

^a Computer Vision and Systems Laboratory, Pavillon Adrien-Pouliot, Université Laval, 1065 avenue de la Médecine, Québec, Québec G1V 0A6, Canada

^b Creaform, 5825 rue Saint-Georges, Lévis, Québec G6V 4L2, Canada

ARTICLE INFO

Article history:

Received 2 March 2012

Accepted 8 March 2014

Available online 18 March 2014

Keywords:

Shape from silhouette

Multi-view image segmentation

Multi lighting

Visual hull

Graph cut

ABSTRACT

This paper presents an unsupervised image segmentation approach for obtaining a set of silhouettes along with the visual hull (VH) of an object observed from multiple viewpoints. The proposed approach can deal with mostly any type of appearance characteristics such as texture, similar background color, shininess, transparency besides other phenomena such as shadows and color bleeding. Compared to more classical methods for silhouette extraction from multiple views, for which certain assumptions are made on the object or scene, neither the background nor the object appearance properties are modeled. The only assumption is the constancy of the unknown background for a given camera viewpoint while the object is under motion. The principal idea of the method is the estimation of the temporal evolution of each pixel over time which provides a stability measurement and leads to its associated background likelihood. In order to cope with shadows and self-shadows, an object is captured under different lighting conditions. Furthermore, the information from the space, time and lighting domains is exploited and merged based on a MRF framework and the constructed energy function is minimized via graph cut. Experiments are performed on a light stage where the object is set on a turntable and is observed from calibrated viewpoints on a hemisphere around the object. Real data experiments show that the proposed approach allows for robust and efficient VH reconstruction of a variety of challenging objects.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The reconstruction of an object's geometry from multiple images is still a challenging problem in computer vision. This problem, also referred to as multi-view geometry, consists of processing a set of 2D images of an object, taken from several viewpoints, in order to produce an interpretation of its geometrical structure. In the real world, object interaction with light involves many effects such as: shadows, self-shadows, color bleeding, light inter-reflection, transparency, subsurface scattering and others. All these phenomena influence the object's appearance in the captured image, and as a result, its analysis is not trivial in the general case.

At some point, the reconstruction process involves the segmentation of the object and the background in all images, resulting in a set of silhouettes. The aim of our work is to obtain these silhouettes automatically in the more general case, that is: without any assumption about object appearance or structure. For this objective a roboticized system was designed. It allows the rotation of an object on a turntable, the control of the lighting conditions

around an object and the selection of the camera viewpoint on a hemisphere as illustrated in Fig. 1. Since images can be captured from any viewpoint on the hemisphere above the object, it is practically not possible to model the background at the pixel level before positioning the camera even if the viewpoints are calibrated. It is also not possible to exploit an active background for all possible viewpoints, like in [1]. Instead we address the problem with a more fundamental approach where the object moves while the camera remains static at each viewpoint. Moreover, we exploit the possibility to change the – unknown – lighting conditions in the scene. Thus, the only assumption for any type of object and background is that background pixels should, more or less, remain constant while the object is moving and is being observed from a given viewpoint. In contrast to classical approaches where the background model is known *a priori* and a camera is fixed [2,1] or object photometric consistency between views is requested [3,4], the proposed approach makes it possible to address very difficult situations such as transparent objects.

Controlling the lighting conditions also helps in interpreting the silhouettes since low intensity areas will change their location in the image and it will be possible to correctly analyze these areas when they are well illuminated. Nevertheless, phenomena such as color bleeding from the object to background could lead to

[☆] This paper has been recommended for acceptance by Edwin R. Hancock.

* Corresponding author.

E-mail addresses: maxim.mikhnevich.1@ulaval.ca (M. Mikhnevich), patrick.hebert@creaform3d.com (P. Hébert), denis.laurendeau@gel.ulaval.ca (D. Laurendeau).

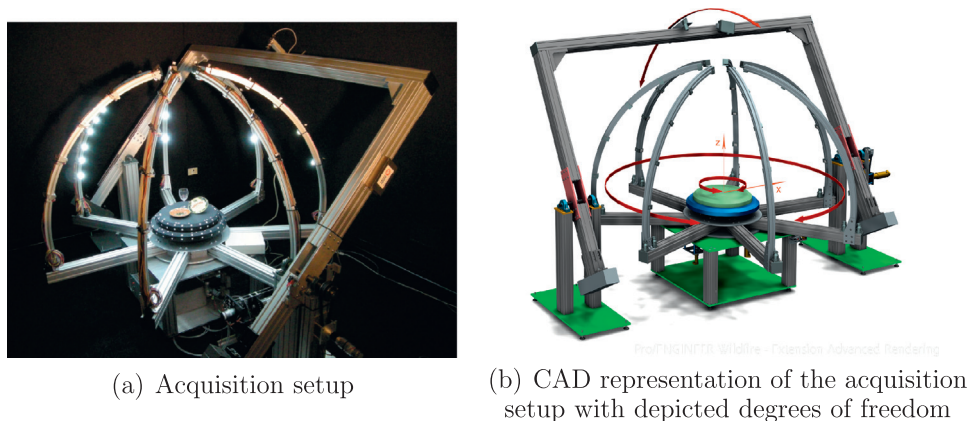


Fig. 1. Acquisition setup. (a) Image of the real acquisition setup. (b) CAD model of the setup, red arrows show the direction of possible rotation of each element of the setup. The turntable can be rotated around the z axis by 360° . The lighting system can be rotated independently of the turntable around the z axis by 360° . The camera system can be rotated by 180° around the x axis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

misinterpretation of background pixels. In this case, the combination of the silhouettes captured from several viewpoints into a visual hull (VH) will be a minimal 3D interpretation to eliminate such view-dependent effects.

The basic experiment takes place under the following conditions: the object is put on a turntable and the camera is stable. The turntable is rotated and, each time, an image of an object is captured under several lighting conditions. Then all the captured images are grouped by light sources and each group is processed as a time sequence. The difference between images in one group is characterized by the motion of an object and its interaction with the light. Based on these observations and the constant background assumption, the background likelihood at each pixel is estimated by our method. Finally, the resulting background likelihood is used to compute the object's silhouette and the process is repeated for another viewpoint.

The key idea of the method is to examine the evolution of each pixel over time. The pixels for which the intensity is stable with respect to their temporal neighbors are more likely to belong to the background, while pixels whose intensities vary are more likely to belong to the foreground due to object motion. It can be observed in Fig. 2(a) that the intensity of each pixel (blue curve) in the intervals from 0 to 140 and from 240 to 359 is similar to that of its temporal neighbors. These constant intervals correspond to our expectation of background behavior. On the other hand, the intensity signal in the interval from 140 to 240 is not stable, the value of each point is different from that of its temporal neighbors and such a signal is assumed to correspond to an object. These observations are reflected in the estimated stability signal (red curve). When the intensity signal is stable and uniform, the value of the estimated background likelihood is high, while in the other case it is low.

One of the key features of the experimental setup that has to be emphasized is the possibility to switch on and off light sources. This procedure helps the method to cope with shadows present in the scene and, for that reason, the background likelihood is estimated for different lighting conditions (see Fig. 2(a) and (d)). In addition we show the benefits of switching on light sources sequentially instead of turning all of the light sources on simultaneously. More generally, the acquisition process switches between steps in time and light changes. In the latter case, no spatial neighborhood is defined but we still consider the processing over a volume of images, parameterized in time and light source index.

The ideas stated above are integrated into a Markov Random Field (MRF) optimization framework, and the optimization is performed through graph cut. This optimization framework makes it

possible to further consider spatial neighborhood in images. The proposed approach will thus produce a set of silhouettes along with a visual hull from the simplest to the most difficult situations.

The contributions of our work are:

- The temporal analysis of a pixel intensity signal which leads to automatic estimation of background likelihood. The processing of this signal also integrates a new type of filtering that preserves edges in a signal.
- The fusion of information from different domains such as multiple space, time, light conditions and its integration into an MRF framework.
- The practical implementation of a general framework capable of producing visual hulls as a first step in multiview geometry for various and highly challenging situations.

The paper is organized as follows: in Section 2 an overview of the related work is given. Section 3 introduces the hypothesis and notation used in the paper. In Sections 4 and 5 the details of the estimation of background and object likelihoods as well as the segmentation framework are presented. In Section 6 some experiments are described and results are demonstrated. In the last Section we conclude and propose a direction for future work.

2. Related work

2.1. Overview

The concept of VH was first introduced by Baumgart [5]. This concept is strongly linked to image segmentation into background and foreground, since the object's silhouettes are required for VH construction. One of the classical and simplest techniques for silhouette extraction is chroma keying. The basic idea is to take a picture of an object against a uniform or known background, then by using color thresholding or background subtraction, a silhouette of an object is obtained. This approach was formally described in [6]. This idea is often used in 3D reconstruction systems for silhouette extraction [1,7]. Even though this method is easy to implement and provides fairly good results, there are some drawbacks. First the environment must be modified in order to set up the background. This limits possible camera placement on a hemisphere since the background has to be visible from all viewpoints. Second, an object part with the same color as the background may create holes in the object's silhouette.

Download English Version:

<https://daneshyari.com/en/article/527582>

Download Persian Version:

<https://daneshyari.com/article/527582>

[Daneshyari.com](https://daneshyari.com)