J. Vis. Commun. Image R. 21 (2010) 453–461

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Multi-view video based multiple objects segmentation using graph cut and spatiotemporal projections

Qian Zhang *, King Ngi Ngan

Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

ARTICLE INFO

Article history:
Received 1 July 2009
Accepted 14 September 2009
Available online 22 September 2009

Keywords:
Multi-view pre-processing
Depth reconstruction
Saliency model
Object extraction
Graph cut
Spatiotemporal projections
Multi-view segmentation
Multi-view video

ABSTRACT

In this paper, we present an automatic algorithm to segment multiple objects from multi-view video. The *Initial Interested Objects* (IIOs) are automatically extracted in the *key view* of the *initial frame* based on the saliency model. Multiple objects segmentation is decomposed into several sub-segmentation problems, and solved by minimizing the energy function using binary label graph cut. In the proposed novel energy function, the color and depth cues are integrated with the data term, which is then modified with *background penalty with occlusion reasoning*. In the smoothness term, *foreground contrast enhancement* is developed to strengthen the moving objects boundary, and at the same time attenuates the background contrast. To segment the multi-view video, the coarse predictions of the other views and the successive frame are projected by pixel-based disparity and motion compensation, respectively, which exploits the inherent spatiotemporal consistency. Uncertain band along the object boundary is shaped based on *activity* measure and refined with graph cut, resulting in a more accurate *Interested Objects* (IOs) layer across all views of the frames. The experiments are implemented on a couple of multi-view videos with real and complex scenes. Excellent subjective results have shown the robustness and efficiency of the proposed algorithm.

1. Introduction

In the recent decades, image/video segmentation has become an active research topic in image processing, computer vision and computer graphics, leading to significant breakthroughs on the development of its theories and technologies. Robust and accurate separation of foreground object from background has turned out to be a crucial prerequisite for many applications such as face segmentation in videotelephony [1], video object cut for pasting [2], and 3D modeling and reconstruction by joint segmentation [3]. Current segmentation methods can be categorized into two groups, region-based segmentation and boundary-based segmentation. Region-based segmentation methods aim to directly construct the region itself, while boundary-based segmentation methods tend to represent each region by its boundary. Some of the classical region-based segmentation methods are mean-shift [4], region growing [5], and graph partition (graph cut [6], grab cut [7]), as well as some popular image cutout tools such as Magic Wand in Photoshop. Active contour (snake) [8], level set [9] and GVF [10] are the representative approaches for boundary-based segmentation. Lazy snapping [11] designs a novel user interface for image cutout by inheriting the advantages of region-based and boundary-based methods.

Most of the interest has been focused on the research of single view segmentation, thus many advanced algorithms have emerged [12–15]. On the contrary, multiple view segmentation has not attracted much attention due to the limitation of image capturing technology and the difficulty to segment all the images simultaneously in real-time. However, multi-view images capturing the real-world environment from arbitrary viewpoints are capable of describing dynamic scene from different angles and can provide the observer more vivid and extensive viewing experience than the single-view image, resulting in more realistic and exciting visual effect. Additionally, depth information in the 3D scene can be reconstructed from multi-view images and assists in characterizing the visual objects more efficiently than the conventional 2D representation. Furthermore, efficient segmentation of IOs has played an important role in many multi-view applications, such as image-based rendering and 3D object model reconstruction. In image-based rendering, multi-view images are available for good visual rendering quality. The end-users may desire to render only the IOs instead of the whole scene, which makes the accurate segmentation of the objects desirable. For 3D object model reconstruction, integrating the 2D images captured from different views to reconstruct the 3D object model is a challenging problem. The first task is the efficient removal of background from these objects.

* Corresponding author.
E-mail addresses: qzhang@ee.cuhk.edu.hk (Q. Zhang), knngan@ee.cuhk.edu.hk (K.N. Ngan).

1047-3203/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2009.09.005

With the recent growing capability of the capturing devices, multi-view capturing system with dense or sparse camera array [16,17] can be built with ease, which motivates the development of multi-view techniques and its related applications. A multi-view image segmentation algorithm proposed in [18] aims to segment foreground object from a collection of 2D images taken from different viewpoints for 3D object reconstruction. It incorporates some useful and well-known algorithms including graph cut image segmentation, volumetric graph cut and learning shape priors. Quan et al. [19] investigated the issue of image-based plant modeling. They propose a plant modeling system for generating 3D models of natural-looking plant from a number of images captured by a hand-held camera with different views. Segmenting the leaves of a plant is a tough problem because of the occlusion and similarity of color between different overlapping leaves. In their approach, leaf segmentation problem is formulated as graph-based optimization aided by 3D and 2D information. To reconstruct the 3D geometry of static scene, an algorithm in [20] simultaneously deals with the depth map estimation and background separation in multi-view setting with several calibrated cameras. By exploiting the strong interdependency of two problems and minimizing a discrete energy functional using graph cut, this combined approach yields more correct depth estimate and better background separation on both real-world and synthetic scenes. The state-of-the-art work for bi-layer segmentation of the stereo video sequence is presented in [21]. By probabilistic fusion of stereo, color and contrast cues, it efficiently separates the foreground from background layer in real-time, and successfully applies to background substitution.

2. Overview of the proposed framework

In this paper, we propose an automatic and efficient algorithm to segment multiple objects from multi-view video. Fig. 1 shows the algorithm framework composed of three components: data pre-processing, offline-operations and online segmentation. We built a five-view camera system to capture the multi-view video data. Given the multi-view image sets I_t^v captured at time instances t from five different views $v \in \{0,1,2,3,4\}$, the objective is to obtain the labeling field f_t^v. After data acquisition, the raw sources undergo two pre-processing stages: color equalization and geometric calibration. Color equalization uniformizes the color responses across all views. Geometric calibration calculates the multiple camera parameters by the nonlinear algorithm in [22], used for correction of geometric distortion and for disparity estimation based on epipolar constraint.

In off-line operations, auxiliary information is calculated beforehand to support the online segmentation. Images with far views will lead to large search range of the disparity value, which makes the stereo matching error-prone and disparity estimation time-consuming. In order to reduce the projection error and avoid extensive computational load, we select view 2 as the *key view* to start the segmentation process. Motion field $M_{t,t-1}^v$ between successive frames, disparity field $D_t^{v_i,v_j}$ (target view v_i with respect to reference view v_j) and occlusion map $O_t^{v_i,v_j}$ between two neighboring views are estimated offline. Based on the camera geometry and perspective projection model, depth can be reconstructed using the multiple disparity maps and the calibrated camera parameters. Depth maps DE_t^v are reconstructed using two disparity maps between a particular view and its two neighboring views. The occluded pixels in either of the occlusion maps between v and its two neighboring views are defined as occluded in the combined occlusion map CO_t^v.

The remainder of this paper focuses on the online segmentation. In Section 3, we introduce the multiple objects segmentation in the *key view* of the multi-view images. Section 4 is devoted to the multi-view video segmentation. Experimental results shown in the Section 5 validate the efficiency and robustness of the proposed algorithm. Finally, conclusions are drawn in Section 6.

3. Multiple objects segmentation for key view

In computer vision, image segmentation generally can be formulated as an energy minimization problem. Graph cut as a powerful energy minimization tool, has been widely used for solving many related vision and graphic problems with great success, such as stereo matching [23], multi-view reconstruction [24] and texture synthesis [25]. With its efficiency in segmentation as demonstrated by Boykov and Jolly [6], graph cut has generated extensive interest for image segmentation and spawned many related works [26–28].

3.1. Automatic IOs extraction based on saliency model

Most of the classical and start-of-the-art graph cut based segmentation algorithms require user's interventions to specify the initial foreground and background regions as hard constraints. Even though user's assistance is helpful to achieve good segmentation results, a major drawback is the dependence on such guidance. Initialization itself may be annoying to the user especially large quantities are needed. Furthermore, graph cut based segmentation

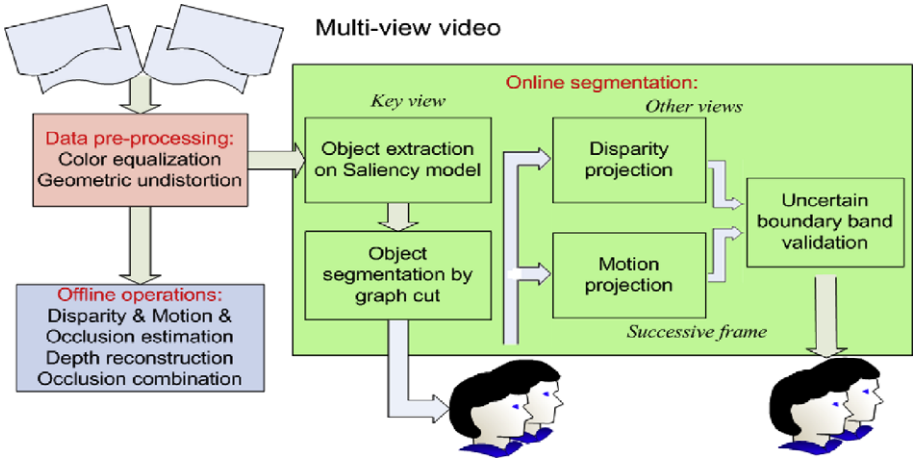

Fig. 1. The framework of the proposed algorithm.