J. Vis. Commun. Image R. 25 (2014) 184–194

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

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

Region based foreground segmentation combining color and depth sensors via logarithmic opinion pool decision

Jaime Gallego *, Montse Pardàs

Universitat Politècnica de Catalunya (UPC), Department of Signal Theory and Communications, Barcelona, Spain

A R T I C L E I N F O

Article history:
Available online 1 April 2013

Keywords:
Foreground segmentation
Space-color models
Space-depth models
Kinect camera
GMM
Color and depth combination
Logarithmic Opinion Pool
Hellinger distance

A B S T R A C T

In this paper we present a novel foreground segmentation system that combines color and depth sensors information to perform a more complete Bayesian segmentation between foreground and background classes. The system shows a combination of spatial-color and spatial-depth region-based models for the foreground as well as color and depth pixel-wise models for the background in a Logarithmic Opinion Pool decision framework used to correctly combine the likelihoods of each model. A posterior enhancement step based on a trimap analysis is also proposed in order to correct the precision errors that the depth sensor introduces. The results presented in this paper show that our system is robust in front of color and depth camouflage problems between the foreground object and the background, and also improves the segmentation in the area of the objects' contours by reducing the false positive detections that appear due to the lack of precision of the depth sensors.

1. Introduction

Foreground segmentation is an important image processing area that is needed in video applications where the output depends completely or partially on the visualization of the segmentation. In some applications, the objects segmentation is the objective itself (i.e. realistic video-conferencing), in others, it serves as a first processing stage for vision systems which monitor real-world activity (i.e. human behavior understanding, video surveillance applications).

New devices suitable for capturing the depth of the scene, which have been developed in the recent years, are creating a new trend on the foreground segmentation area towards the new available depth information of the scenes. ToF and structured light depth cameras are an example of these kind of devices that have offered an alternative to the stereo systems and their complex problems in the disparity estimation.

For several years, many authors have been working in foreground segmentation using static color camera devices. Real-time foreground segmentation techniques based on pixel-wise background modeling are proposed in [1–3]. Recently, other authors have proposed some techniques based on foreground modeling: Yu et al. [4] propose to model the foreground and the background classes by means of spatial-color GMM models, while in [5] we proposed a combination between pixel-wise background model and region based color-spatial foreground model. Despite foreground modeling methods improve the performance of the color foreground segmentation, all these methods present problems when foreground objects have similar color to the background, the camouflage problem, or when lighting or shadow affect the foreground and background.

Depth data allows a more robust segmentation of the object of interest towards the color camouflage problem than the systems based on color segmentation: Kolb et al. [6] shows an exhaustive revision of the ToF technology, applications and limitations. Guomundsson et al. [7], Stone and Skubic [8] use a pixel-wise exception to background segmentation using MoG background model, while Xia et al. [9] defines a depth region growing in a depth thresholding condition.

Nevertheless, all these systems present other problems concerning the segmentation with depth sensors: lack of precision in the segmentation due to the noisy and low resolution depth maps obtained by the sensors, and presence of camouflage errors when foreground and background present similar depth. How to solve these problems combining both color and depth sensors is becoming an important topic in order to achieve a precise and robust objects segmentation that uses the best characteristics of each sensor, avoiding, as far as possible, color and depth camouflage problems. Some authors have proposed some solutions in this research line combining depth thresholding segmentation and detection refinement:

In [10] the authors propose a simple depth thresholding segmentation followed by a color-depth trimap analysis to improve the precision of the segmentation in the borders of the object,

* Corresponding author.
E-mail addresses: jgallego@upc.edu (J. Gallego), montse.pardas@upc.edu (M. Pardàs).

1047-3203/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jvcir.2013.03.019

while in [11] the depth thresholding segmentation allows to automatically obtain a pentamap that is used to make a more efficient color graph cut regularization. Frick et al. [12] uses a thresholding technique to separate foreground from the background in multiple planes, and a posterior trimap based refinement based on hierarchical grab-cut segmentation to reduce the artifacts produced by the depth noise.

These kind of methods allow to obtain correct results under limited constraints on the scenario set and the depth thresholding, but present some segmentation errors in presence of difficult situations provided by depth camouflage problem.

Other authors have addressed the problem trying to combine the color and depth sensors in a more robust framework:

In [13] the authors propose a color-depth Mean Shift segmentation system, of the overall image, based on the depth noise analysis in order to weight the depth reliability, while Schiller and Koch [14] uses a pixel-wise probabilistic background model in color-depth domain, to perform a more complete exception to background segmentation. Although these kind of methods present a more robust and general framework in front of camouflage situations than the thresholding approaches, they still present some problems for correctly combining the color and depth sensors information when camouflage situations appear because of the lack of foreground objects information to detect it and thus, to improve the final segmentation results. In this way, Wang et al. [15] proposes a probabilistic fusion framework between foreground and background classes for color and depth cues, which achieve correct results in close-up sequences. The probabilistic models of each one of the classes are combined according to the foreground–background histogram similarity.

Combining different sensors for foreground segregation is an important topic in order to achieve a precise and robust objects segmentation. In this area, Pinheiro and Lima [16] propose a sensor fusion combination based on Multi Bayesian utility functions, while several authors in statistics have well addressed the fusion of information provided by several sensors in a Bayesian framework [17,18].

In this paper, we present a novel foreground segmentation system that belongs to the last group of proposals. We propose a system that combines in a probabilistic framework both, color and depth sensors information to perform a more complete Bayesian segmentation between foreground and background classes. The system, suitable for static color-depth camera sequences in a close-up and long-shot views, achieves a correct segmentation results taking into account the spatial context of the models showing a combination of color-spatial and depth-spatial region-based models for the foreground and a color and depth pixel-wise models for the background in a Bayesian Logarithmic Opinion Pool decision model. In order to improve the foreground segmentation precision, we add a final segmentation refinement based on a trimap analysis.

1.1. Problem statement

The choice of the correct sensor to achieve a correct foreground segmentation will depend on the characteristics of the scene that we want to segment. In this paper, we address both color and depth camera sensors:

– Color cameras are based on sensors like CCD or CMOS among others, which allow us a more reliable representation of the scene with high resolution. The segmentation using these kind of sensors results in more precise separation between foreground and background if there are no color camouflage problems between both classes.
– Depth cameras are based on IR transmitter/receiver sensors. Despite new precise sensors based on laser technologies are appearing nowadays, so far, the resultant depth maps obtained using ToF and kinect devices are images with lack of precision on the definition of the objects ([6] gives an in-depth technical analysis of ToF limitations). The segmentation using these devices is a more robust segmentation against color problems, though errors with depth camouflage will be present.

Both, color and depth sensors work with different technologies that can be used together at the same time without suffering any interference between them, thus presenting non-correlated errors each other. Therefore, a correct combination of these sensors will allow us to improve the overall performance of the system. For that, we calibrate and register the depth and color cameras projecting the depth map onto the color image, allowing a color-depth pixel correspondence. It should be noted that some problems of miss association can appear due to:

- Camera centers are different and some blind regions appear for each one of the sensors because of the projection process and the parallax computation between cameras.
- The low resolution of the depth measurements produces that several color pixels are associated to only one depth map value.
- The lack of precision of the depth sensor is more pronounced in the borders of the objects, and produces many depth-color association errors in these regions.

In order to show the limitations of each sensor, Fig. 2 shows an example of segmentation with a simple exception to background analysis. We can observe how color segmentation (Fig. 2(c)) gives us a reduced false positive detections in the segmentation although some false negative errors appear due to the foreground–background color similarity. When using depth segmentation (Fig. 2(d)), robustness against color similarity is present, but some false positive detections appear in the borders of the object due to the lack of precision of the depth sensor.

All the sequences used in this paper has been recorded by means of the kinect sensor, developed by Microsoft, combined with the OpenNI SDK configured with the factory calibration presets, in order to obtain the synchronized and registered color and depth maps of the sequences under analysis.

1.2. Proposed system

We propose a segmentation system that exploits the advantages of each sensor type. With this aim we propose an algorithm where parametric foreground models in color-space and depth-space domains are evaluated against pixel-wise color and depth background models. The improvements of the proposed method in the segmentation process are twofold: first, we combine probabilistic models of color, space and depth, in order to obtain a correct pixel classification according to the color and depth sensors and in a posterior step, we correct the errors of precision that the depth sensor introduces to the overall process and are converted into some false positive detections in the borders of the foreground object. Once the approximate position of the borders in the current image is known, it is better to disregard the depth information at those positions where the color sensor provides enough discrimination. The reliability of the sensors, based on the Hellinger distance [19] between foreground and background models is used in both stages.