Optik 125 (2014) 1809–1815

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

# Spatio-temporal segmentation of moving objects using edge features in infrared videos

Chaobo Min, Junju Zhang*, Benkang Chang, Bin Sun, Yingjie Li

*School of Electronic Engineering and Optic-Electronic Technology, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, PR China*

ARTICLE INFO

*Article history:*
Received 17 April 2013
Accepted 3 September 2013

*Keywords:*
Spatio-spatial segmentation
Markov Random Field
Performance evaluation
Infrared video

ABSTRACT

In this paper, a novel method is proposed for spatio-temporal segmentation of moving objects using edge features in infrared videos. We define motion saliency of edge (MSoE) to generate the MSoE-map. The seeds of moving objects are extracted from the MSoE-map by using Otsu's method and subsequently compensated by historical data. An improved layer-based region growing method is applied to the seeds to achieve spatial segmentation of moving objects. The region growing method has an adjustable growing threshold. So, one of the focuses of our work is how to determine the best growing threshold. A Markov Random Field (MRF) based criterion with maximum a posterior (MAP) estimation principle is proposed for performance evaluation of moving object segmentation without ground truth (GT) in infrared videos. This criterion can be considered as an object function of threshold determination during global searching. The global optimum is accomplished by using simulated annealing (SA) algorithm to obtain the best growing threshold. The final segmentation mask of moving objects is grown from the seeds with the best growing threshold. Experimental results are provided to illustrate that the proposed method has better performance for moving object segmentation with fewer effects of object-background misclassification in infrared videos.



## 1. Introduction

Infrared cameras can be used both day and night to produce infrared spectrum images [1]. Infrared images provide useful information for many applications, such as military navigation, automatic target recognition and tracking. Applications of infrared moving object detection are numerous and they span a wide range of applications including surveillance system, pedestrian tracking and aerospace application, to name a few. There are many difficulties with moving object segmentation in infrared videos, such as noise of the target, change in scene illumination, camera occlusion and a short standstill of moving objects, to mention a few.

The past decades have witnessed tremendous growth of the algorithms of infrared target segmentation. Many spatial segmentation methods focusing on extracting the thermal target regions from background have been explored. Some spatial segmentation methods are based on thresholding, for instance, one-dimensional Otsu [2], two-dimensional Otsu [3], two-dimensional entropy [4], etc. The methods based on thresholding are effective when the image gray histogram presents "double humps and single valley". However, this gray histogram is impossible in the infrared image [5]. Region growing method is a simple, steady and effective method of image segmentation. But it is very sensitive to noise and the gray level of targets. Its performance is affected by the seed and the growing threshold.

The changed information of moving objects can be detected by using a temporal segmentation scheme – frame subtraction scheme [6]. However, the results of frame subtraction maybe contain a lot of noise pixels and incomplete moving objects. It is difficult to extract out the complete pixels of moving objects with frame subtraction scheme. Background subtraction [7] scheme is a temporal method commonly used to segment interested objects in image sequences. It can extract the interested regions by comparing new frames to a background model. Mixture of Gaussian (MOG) background modeling method is developed into the current form by Stauffer et al. [8]. However, the background subtraction scheme cannot achieve satisfactory results for the infrared image with complex background and low signal-to-noise ratio (SNR). The motion-based segmentation method proposed in [9] can estimate optical flow vector and divide video frames into fixed-size blocks for the analysis of likelihoods of observations by using hidden Markov models (HMMs). But this method may filter out valuable information about abnormal events [10]. In addition, optical flow needs to assume that the luminance of any point of any object is fixed [11].

* Corresponding author at: Institute of Electronic Engineering and Optic-Electronic Technology, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, PR China. Tel.: +86 025 84315437.
*E-mail address:* mcbnjust@gmail.com (J. Zhang).

0030-4026/$ – see front matter. Crown Copyright © 2013 Published by Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.09.080

Due to high uncertainty and high ambiguity of the gray level of pixels, it is still challenging to segment moving objects with more accuracy by non-statistical segmentation method. Some kind of stochastic method is applied to model the important attributes of an image in order to get a better segmentation performance. In this regard, Markov Random Field (MRF) model is proved to be a better framework [12]. MRF theory provides a convenient and consistent way for modeling context dependent entities such as image pixels and correlated features. In [13], a method of object detection based on MRF is proposed, where a smooth transition of segmentation results is obtained by temporal constraints and temporal local intensity adaptations. A robust method for spatio-temporal segmentation of moving objects is proposed in [14]. Spatial segmentation is achieved by attribute modeling with MRF–MAP estimation based on distributed genetic algorithm (DGA). Temporal segmentation is achieved by direct combination of video object plane (VOP) of the previous frame with the change detection mask (CDM) of the current frame. A spatio-temporal segmentation of moving object detection based on compound MRF is presented in [15], where a compound MRF model is used as the prior image attribute model which takes care of the spatial distribution of color, temporal color coherence and edge map in the temporal frames to achieve a spatio-temporal segmentation.

In this work, we propose a novel spatio-temporal algorithm of moving object segmentation using edge features in infrared videos. We have defined motion saliency of edge (MSoE) to form a MSoE-map which can be considered to be a special combination of spatial and temporal edge features in an infrared image. The higher the MSoE-value of an edge point is, the more likely that point is to be the edge point of moving objects. So based on the MSoE-map, the edge points belonging to moving objects are extracted by Otsu's thresholding method, and subsequently compensated by historical data. Those edge points represent most of the edge points of moving objects. We consider them as the seeds of moving objects. Then, an improved layer-based region growing method is applied to the seeds to achieve spatial segmentation of moving objects. This region growing method has an adjustable growing threshold. So, one of the focuses of our work is how to determine the best growing threshold. A MRF based criterion is proposed for performance evaluation of moving object segmentation without GT in infrared videos. The proposed criterion is utilized to determine the best growing threshold, where the performance of moving object segmentation is measured by that of the corresponding segmentation mask's edge. We consider a MRF modeling for each edge point of a segmentation mask in spatial and temporal directions. This problem is formulated using MAP estimation principle. The MRF–MAP based criterion is used as an object function to determine the best growing threshold during global searching. At last, the global optimum is achieved by using SA algorithm to obtain the best growing threshold. The final segmentation mask of moving objects is grown from the seeds with the best growing threshold.

The results obtained by the proposed spatio-temporal segmentation method are compared with those of frame subtraction using Otsu's thresholding, MOG method [8] and a compound MRF based spatio-temporal segmentation method [15]. It is found that the proposed scheme produces better results toward moving object segmentation with less effect of object-background misclassification and noise.

The organization of this paper is as follows. In Section 2, a MRF–MAP based criterion is proposed for performance evaluation of moving object segmentation without GT in infrared video. In Section 3, we introduce a moving object segmentation scheme using performance evaluation, where seed determination based on MSoE is narrated in Section 3.1, an improved layer-based region growing method is proposed in Section 3.2, and the procedure of global optimum based on SA algorithm is shown in Section 3.3. Section 4 provides simulation results and analysis. Conclusion is presented in Section 5.

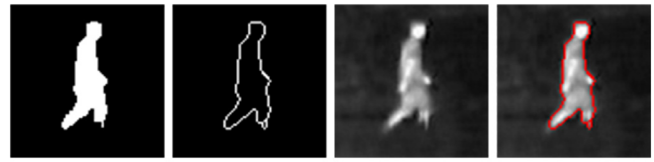


**Fig. 1.** The illustration of a realization: (a) the segmentation mask $M_t$; (b) the edge map of the segmentation mask $M_t^e$; (c) the original image $I_t$; (d) the realization $x_t$ (the edge points are labeled in red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

## 2. A new criterion for evaluation of moving object segmentation

It is a difficult task for performance evaluation of moving object segmentation without GT in infrared videos. Many attributes of a segmentation mask of moving objects have been applied to performance evaluation, such as gray, texture or optical flow, etc. The edge of a segmentation mask represents the estimated contour of moving objects. Thus, the performance of a segmentation mask's edge can reflect the similarity between the contours of a segmentation mask and actual moving objects. In this section, an edge of a segmentation mask is employed to evaluate the performance of moving object segmentation without GT.

### *2.1. Spatio-temporal MRF modeling*

Let $I_t$ represent a frame image of an infrared video at time $t$. The size of $I_t$ is $m \times n$. The spatial site of each pixel is denoted by $s$ and $I_t(s)$ indicates the value of each pixel in $I_t$, where $s=(i,j)$, $1 \le i \le m$, $1 \le j \le n$. A segmentation mask of $I_t$ is denoted by $M_t$. Edge maps of $M_t$ and $I_t$ are obtained by considering a $3 \times 3$ Laplacian window and they are defined as $M_t^e$ and $I_t^e$, respectively. Let $L_t = \{z_k : z_k \in R^2\}_{k=1}^N$ be a set of edge points in $M_t^e$, where $z_k$ is the spatial site of an edge point and $N$ is the number of edge points. An edge map of a segmentation mask can be considered to be a labeling problem of regular sites with discrete labels. In order to preserve the edge features, we consider a circular neighborhood for each edge point both in spatial and temporal MRF models. The circular neighborhood centers on $z_k$ with $r_N$ in radius and is defined as $S_{kt}$. In the neighborhood $S_{kt}$, the region belonging to foreground of a segmentation mask is denoted by $S_{kt}^{\text{in}}$, and the region belonging to background of a segmentation mask is denoted by $S_{kt}^{\text{out}}$. $S_{kt}^{\text{in}}$ is called as inner neighborhood and $S_{kt}^{\text{out}}$ is called as outer neighborhood. Let $X_t$ represent the MRF with realization $x_t$ consisting of $M_t^e$ and $I_t$, such as Fig. 1. Similarly, pixels in the temporal direction are also modeled as MRFs. Since we have considered MRF modeling for edge, $X_t$ satisfies the Markovianity property in spatial direction, that is

$$\begin{aligned} P[X_t(z_k) &= M_t^e(z_k)|X_t(s) = I_t(s), \forall s \in V, s \neq z_k] = P[X_t(z_k) \\ &= M_t^e(z_k)|X_t(s) = I_t(s), \quad s \in S_{kt}], \qquad z_k \in L_t, \end{aligned} \tag{1}$$

where $V$ denotes a rectangular lattice of $X_t$ and it can be denoted by $V=\{s=(i,j)|1 \le i \le m, 1 \le j \le n\}$. For temporal MRF, the following Markovianity property is also satisfied:

$$\begin{aligned} P[X_t(z_k) &= M_t^e(z_k)|X_h(s) = I_h(s), \forall s \in V, \forall h \in \{1, \ldots, \quad t-1\}, \quad s \neq z_k] \\ &= P[X_t(z_k) = M_t^e(z_k)|X_h(s) \\ &= I_h(s), \quad s \in S_{kt}, h \in \{t-1\}], \quad z_k \in L_t. \end{aligned} \tag{2}$$