Signal Processing 89 (2009) 685–691

Contents lists available at ScienceDirect

Signal Processing

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

Fast communication

The analysis of the color similarity problem in moving object detection

Xiang Zhang *, Jie Yang

University of Shanghai Jiaotong, Dochuan Road 800, Shanghai, China

ARTICLE INFO

Article history:
Received 3 March 2008
Received in revised form
27 September 2008
Accepted 27 October 2008
Available online 5 November 2008

Keywords:
Moving object detection
Foreground segmentation
Color similarity
Confusion point
Model accuracy

ABSTRACT

The color similarity between the background and foreground causes serious misdetections in moving object detection from video sequences. In this paper, we point out that the existence of a confusion point and the model inaccuracy are the reasons for the misdetections due to the color similarity. Accordingly, the solutions of the color similarity are to shift the confusion point and to improve the model accuracy. Based on this conclusion, a simple algorithm by combining a weighting technique and a new foreground model is presented, and improved results are generated. More accurate weighting techniques and foreground models are expected to be developed in the future based on the solutions.

1. Introduction

Serious misdetections are caused in moving object detection when the foreground and background have similar color distributions. An example is shown in Fig. 1a, where the skirt and hair of the pedestrian exhibit similar color to the background. Many foreground pixels are misclassified in Fig. 1b by Sheikh's algorithm [1], which is one of the most excellent detection algorithms nowadays. The work of this paper aims to explore the reasons for the misdetections caused by the color similarity, and to present solutions of the color similarity problem.

The moving object detection plays an important role in a wide range of computer vision applications. For example, some works [2,3] regarding sport video analysis and some other works [4,5] regarding traffic monitoring are reported in recent years. The environment is well-constrained in sports video analysis. Algorithms used in intelligent transportation are often designed to detect specific objects, such as the pedestrian [4] and shadow [5]. More algorithms are designed to deal with general scenes, such as the dynamic background and illumination changes.

The background modeling is the early criterion exploited for moving object detection. A survey of the background modeling can be found in [6]. Many background models have been developed in recent years, such as the Gaussian mixture model [7], the nonparametric statistical model [8,9], the predictive models [10,11], and et al. The foreground modeling is exploited for more accurate detection in recent years. The foreground model can be constructed in a consistent fashion with the background model [12]. The nonparametric statistical model is the most widely used model now, for it is capable of modeling arbitrary probability distributions. In order to accelerate the computation of the kernel density estimator (KDE) used in nonparametric modeling, fast KDE [13,14] are developed. The thresholding is formerly used to classify new observations into foreground and background, whereas the energy minimization tools [15,16] have become the standard classification tools now.

In this paper, further analysis of the example in Fig. 1 reveals that the confusion point and the model inaccuracy

* Corresponding author. Tel./fax: +86 021 34204033.
E-mail addresses: hover_chang@sjtu.edu.cn (X. Zhang), jieyang@sjtu.edu.cn (J. Yang).

0165-1684/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2008.10.027

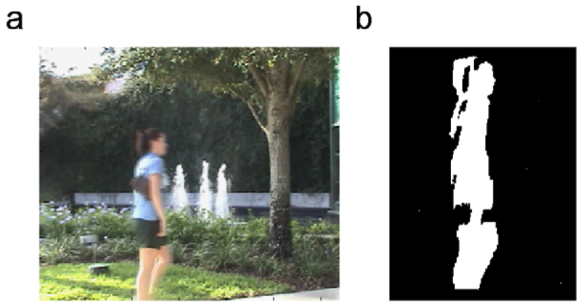

Fig. 1. (a) Is a frame from a video sequence with serious color similarity problem and (b) is segmentation by Sheikh's algorithm.

are the reasons for the misdetections due to the color similarity problem. Accordingly, the color similarity can be solved by shifting the confusion point and improving the model accuracy. Based on this conclusion, a simple but effective algorithm is presented. More elaborate algorithms are expected to be developed in the future. This paper is organized as follows. The confusion point and the model accuracy are discussed in Sections 2 and 3, respectively. Experimental results are given in Section 4, followed by the conclusion in Section 5.

2. Confusion point

As stated in introduction, the nonparametric statistical model is the most popular model nowadays. In order to improve the detection, multiple features have been used for statistical analysis in many algorithms. A combination of color and optical flow is used [17], a combination of color and spatial-temporal derivatives is used [18], and another combination of spectral, spatial and temporal features is used in [19]. The use of multiple features leads to distinct performance improvement, but the color similarity is still a difficult problem. For example, both the color and spatial features are used in Sheikh's algorithm, but many foreground pixels are still misclassified in Fig. 1b. No matter how many features are used, finally a probability model is computed. Since we are interested in how to use the probability model for classification but not how to construct the probability model, only a simple model with the color feature is used in this paper.

Let I_n be a pixel of an image in the RGB color space, where n is the index of the image lattice. Moving object detection aims to assign each pixel I_n a label from the set (background, foreground). Considering pixel I_n at time instant t, before which all pixels labeled background in a K^*K neighborhood of position n form the background model $\varphi_{bn} = (y_1,\ldots,y_m,\ldots,y_M)$, where M is the total number of pixels in φ_{bn}. The background probability, the probability of pixel I_n belonging to the background, can be computed with the KDE as

$$\hat{p}(I_n|\varphi_{bn}) = K^{-2}\sum_M f_H(I_n - y_m) \tag{1}$$

where f is a kernel function with a variance matrix H. In this paper, we assume that all color components are independent of each other and have the same variance. Considering the same pixel I_n at time instant t, before which all pixels labeled foreground in a K^*K neighborhood of position n form the foreground model $\varphi_{fn} = (y_1,\ldots,y_m,\ldots,y_M)$. The foreground probability can be computed as the background probability. In fact the foreground probability is smaller than the background probability most of the time, even if I_n is a foreground pixel. In order to enhance the foreground probability, the foreground probability is modeled as a mixture of the KDE and a uniform likelihood in [1] as

$$\hat{p}(I_n|\varphi_{fn}) = w\tau + (1-w)K^{-2}\sum_M f_H(I_n - y_m) \tag{2}$$

where w is a weight coefficient, τ is a random variable with uniform probability. In order to adapt to the background and foreground changes, sliding windows of length ρ_b frames and ρ_f frames are maintained for the background and foreground models, respectively. More discussions about nonparametric modeling can be found in [1] and [8].

Given the membership probabilities computed with Eqs. (1) and (2), traditionally a likelihood ratio classifier is used to classify new observations into background and foreground. The likelihood ratio η of I_n is defined as

$$\eta(I_n) = \ln(\hat{p}(I_n|\varphi_{fn}))/\ln(\hat{p}(I_n|\varphi_{bn})) \tag{3}$$

The likelihood ratio classifier χ is defined as

$$\chi(I_n) = \begin{cases} \text{background} & \text{if } \eta(I_n) > k \\ \text{foreground} & \text{otherwise} \end{cases} \tag{4}$$

Because segmentations obtained by the likelihood ratio classifier are often noisy, energy minimization tools are used to classify observations instead of the likelihood ratio classifier by introducing prior knowledge into the decision framework. Although much better results are generated by energy minimization from visual observation, in fact the decision for most pixels by energy minimization is the same as that by the likelihood classifier, except a few pixels that are classified as noise points by the likelihood classifier. In this paper, a popular energy minimization tool, the graph cut, is used to classify observations as in [1].

Theoretically the sum of $\hat{p}(I_n|\varphi_{fn})$ and $\hat{p}(I_n|\varphi_{bn})$ is 1, but it is not equal to 1 most of the time. Keeping the ratio between $\hat{p}(I_n|\varphi_{fn})$ and $\hat{p}(I_n|\varphi_{bn})$ unchanged and normalizing the sum to 1, the normalized probabilities are denoted as $p(I_n|\varphi_{fn})$ and $p(I_n|\varphi_{bn})$, respectively. According to Eqs. (3) and (4), the normalization will not change the final classification of pixels.

The negative log-probability $-\ln(p(I_n|\varphi_{fn}))$ can be denoted as

$$-\ln(p(I_n|\varphi_{fn})) = -\ln(1 - p(I_n|\varphi_{bn})) \tag{5}$$

The definition area of $p(I_n|\varphi_{bn})$ is (0,1). Let $p(I_n|\varphi_{bn})$ be the x axis, the two functions, $-\ln(p(I_n|\varphi_{fn}))$ and $-\ln(p(I_n|\varphi_{bn}))$, are shown in Fig. 2a. The confusion point δ, as shown in Fig. 2a, is the cross-point of the two functions. When a pixel is at the confusion point, the probability of the pixel belonging to the background is equal to the probability of the pixel belonging to the