Pattern Recognition Letters 33 (2012) 1601–1613

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Background subtraction based on phase feature and distance transform

Gengjian Xue, Jun Sun, Li Song *

Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
Shanghai Key Laboratory of Digital Media Processing and Transmissions, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 5 September 2010
Available online 22 May 2012
Communicated by R. Davies

Keywords:
Background subtraction
Phase feature
Phase based background model
Distance transform

ABSTRACT

A novel background subtraction method that can work under complex environments is presented in this paper. The proposed method consists of two stages: coarse foreground detection through the phase based background model we present, and foreground refinement using the distance transform. We first propose a phase feature which is suitable for background modeling. The background model is then built where each pixel is modeled as a group of adaptive phase features. Although the foreground detection result produced by the background model only contains some sparse pixels, the basic structure of the foreground has been captured as a whole. In the next stage, we adopt the distance transform to aggregate the pixels surrounding the foreground so that the final result is more clear and integrated. Our method can handle many complex situations including dynamic background and illumination variations, especially for sudden illumination change. Besides, it has no bootstrapping limitations, which means our method is without background initialization constraints. Experiments on real data sets and comparison with the existing techniques show that the proposed method is effective and robust.

1. Introduction

Background subtraction is often the first task in vision-based applications, such as security and surveillance. The output of background subtraction is usually an input to higher level processes, making it a critical part of the system. Background subtraction consists of two phases: building a statistical representation of the background scene, and detecting the foreground by "subtracting" the background from the scene. The performance of background subtraction depends mainly on the background modeling technique it uses. Natural environments make background subtraction a challenging task since they usually contain complex scenes including rippling water, waving trees, illumination variations, etc.

In the last decade, many kinds of approaches have been proposed for moving object detection. These techniques have used pixel intensity, texture or other effective information for background modeling. However, background modeling techniques utilizing phase information are rarely seen. Perhaps the phase wrapping property and the narrow value range restrict its applications.

In this paper, we propose an efficient background subtraction technique based on a phase feature and the distance transform. Our method consists of two stages: modeling the phase based background for coarse foreground detection, and foreground refinement by using the distance transform. We choose the phase for background modeling because it has the property of being insensitive to illumination variations. In order to overcome the inherent limitations of phase wrapping and its narrow value range, a new phase feature which is suitable for background modeling is proposed. After the image patch is convolved with local Gabor filters, the phase feature is constructed by adding up the Gabor phases corresponding to the first largest amplitudes. Assuming the feature value of a particular pixel over time as a pixel process, we model its current value using a mixture of Gaussian distributions. In addition, the adaptive updating scheme ensures that the model has no bootstrapping limitations. The proposed model can detect the basic structure of the foreground but with a sparse representation, the distance transform is then applied to aggregate the pixels surrounding the foreground in order to get more integrated result. We will justify our method by experiments.

The rest of this paper is organized as follows: Section 2 provides a brief review of existing works. A new phase feature for background modeling is proposed in Section 3. In Section 4, our phase based background model is described in detail. The distance transform for foreground refinement is given in Section 5. Experimental results and evaluations are given in Section 6. Conclusions are finally drawn in Section 7.

2. Related work

One of the most common methods of background description is based on a Gaussian distribution. Wren et al. (1997) represented

* Corresponding author at: Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. Tel.: +86 21 34205492; fax: +86 21 34204155.
E-mail addresses: xgjsword@sjtu.edu.cn (G. Xue), sunjun@cdtv.org.cn (J. Sun), song_li@sjtu.edu.cn (L. Song).

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.05.009

the intensity distribution of each background pixel with one Gaussian distribution. In order to describe more complicated scenes, a Gaussian mixture model (GMM) was proposed (Stauffer and Grimson, 1999). The model for each pixel intensity consisted of a few Gaussians, and an online *K*-means approximation technique instead of the exact EM algorithm was adopted for updating. The GMM technique was then modified by several researchers. For example, Zivkovic and van der Heijden (2006) extended the model by constantly selecting the appropriate number of Gaussian components for each pixel while updating the model parameters. Lee (2005) presented an adaptive learning rate calculated for each Gaussian at every frame to improve the model convergence speed.

Another popular technique is the nonparametric statistical approach. Elgammal et al. (2002) utilized a kernel density estimation (KDE) technique for background modeling, where the probability density function (PDF) of the pixel intensity was estimated directly from the data without any distribution assumptions. In (Mittal and Paragios, 2004), an estimation method with an adaptive kernel size for each data point was used. Based on the assumption that ergodicity in time often holds spatially, Jodoin et al. (2007) performed pixel kernel density estimation with only one background frame. This method had a lower memory requirement.

Some authors have proposed region-based techniques for background modeling. Heikkilä and Pietikäinen (2006) modeled the background using local binary pattern (LBP) histograms calculated over a circular region around the pixel. In (Zhang et al., 2008), the LBP feature was computed considering both spatial and temporal information. Mason and Duric (2001) adopted edge and color histograms calculated over the block area as the features to describe the block. Since a region can capture more global information than a single pixel, region-based approaches are more robust under dynamic background scenes.

Several other effective models and methods have also been used for background subtraction. Zhong and Sclaroff (2003) cast the dynamic background region in time as an autoregressive moving process, and they used a robust Kalman filter to estimate the region intrinsic appearance. In (Stenger et al., 2001), a Hidden Markov Model (HMM) with an online parameter estimation scheme was proposed to model the background. Patwardhan et al. (2008) proposed to detect the foreground using pixel layers. Inspired by the biological mechanisms of motion-based perceptual grouping, Mahadevan and Vasconcelos (2010) treated background subtraction as a saliency detection problem, and they proposed a spatiotemporal saliency algorithm that worked well with dynamic background scenes. Assuming background and foreground were two mutual independent signals, Tsai and Lai (2009) adopted the independent component analysis (ICA) technique to extract the foreground. Maddalena and Petrosino (2008) proposed a neural network architecture to model the background, but this technique needed more memory space. Casting background subtraction as a sparse error recovery problem, Dikmen and Huang (2008) presented a sparse representation framework for foreground detection, then they further discussed the different base selection methods (Dikmen et al., 2009).

3. New phase feature for background modeling

Phase contains a wealth of information, and its great importance has been introduced in detail by Oppenheim and Lim (1981). In recent years, phase as a feature has been successfully applied to several fields, such as palmprint identification (Zhang et al., 2003), face recognition (Zhang et al., 2007), etc. However, little work has utilized phase information for background modeling.

In this section, we propose a new phase feature for background modeling. The input image is first convolved with local Gabor filters so that each pixel has a group of features containing multiple amplitudes and corresponding phase values. For each pixel, we select the effective phase information according to the criteria that higher amplitude value in the feature group means more accurate local structure information has been captured, and its corresponding phase information is more representative. The new phase feature is then defined as the sum of the selected phase values.

Due to the properties of spatial localization, orientation selectivity, and spatial-frequency selectivity, Gabor filters have been widely used to extract pixel amplitude and phase information. A two-dimensional Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. The Gabor wavelets can be defined as follows (Zhang et al., 2007):

$$\psi_{\varphi,v}(z) = \frac{\|k_{\varphi,v}\|^2}{\sigma^2} e^{(-\|k_{\varphi,v}\|^2\|z\|^2/2\sigma^2)}[e^{ik_{\varphi,v}z} - e^{-\sigma^2/2}] \tag{1}$$

where $\overrightarrow{k_{\varphi,v}} = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v cos\phi_\varphi \\ k_v sin\phi_\varphi \end{pmatrix}$, $k_v = f_{max}/2^{v/2}$, $\phi_\varphi = \varphi(\pi/\varphi_{max})$, $v = 0,\ldots,v_{max}-1$, $\varphi = 0,\ldots,\varphi_{max}-1$, v is the frequency and φ is the orientation. v_{max} and φ_{max} represent the number of frequencies and orientations, respectively. The first term in the square brackets in (1) determines the oscillatory part of the kernel, and the second term compensates for the DC value. σ determines the ratio of the Gaussian window width to wavelength. The Gabor transformation of a given image is defined as its convolution with the Gabor functions:

$$G_{\varphi,v}(z) = I(z) * \Psi_{\varphi,v}(z) \tag{2}$$

where the symbol "$*$" represents the convolution operator, $z = (x,y)$ denotes the image position, and $G_{\varphi,v}(z)$ is the convolution result corresponding to the Gabor kernel at frequency v and orientation φ. $G_{\varphi,v}(z)$ is a complex value which is composed of one amplitude item $A_{\varphi,v}(z)$ and one phase item $\theta_{\varphi,v}(z) \in [0,2\pi)$. It can be written as:

$$G_{\varphi,v}(z) = A_{\varphi,v}(z) \cdot exp(i\theta_{\varphi,v}(z)) \tag{3}$$

Since the phase value varies quickly with image locations, applying the Gabor filter to the entire image would produce results too coarse to effectively represent the phase feature. In order to more precisely extract the phase information for each pixel, we divide the image into non-overlapping partitions and execute patch based convolutions. Based on our experience, the Gabor filters are designed with four frequencies and six orientations, which means $v_{max} = 4$, $\varphi_{max} = 6$. Fig. 1 shows the amplitude responses of Gabor function with different frequencies. The frequency of Gabor wavelet is computed according to the formula: $k_v = f_{max}/2^{v/2}$, $v = 0, \ldots, v_{max}-1$.

It can be seen that the frequency value is higher and the Gabor wavelength is shorter when v has a lower value. While the value v is increasing, the frequency value is decreasing and the Gabor wavelength is becoming longer. We select four different frequencies in our experiments because the wavelength of the central envelope is about eight pixels when v equals to 3, which is capable of capturing image local structures; and more frequencies would contribute very little to our problem. The number of orientations is set to 6 to maintain the discriminability of the Gabor filters.

The patch size is determined by considering the Gabor wavelength. Larger patch size may result in a coarse information representation, while smaller patch size may not capture image local structures. Varying the patch size from 3 to 6, we observe the results so as to choose the appropriate value. Taking a public sequence which features a sudden illumination change (Toyama et al., 1999) as an example, Fig. 2 shows the convolution results at a randomly selected point from the 1865th frame using different patch sizes.