Neurocomputing 103 (2013) 132–142

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Background subtraction driven seeds selection for moving objects segmentation and matting

Bineng Zhong [a,d,*], Yan Chen [a], Yewang Chen [a], Rongrong Ji [b], Ying Chen [c], Duansheng Chen [a], Hanzi Wang [d]

[a] Department of Computer Science and Engineering, Huaqiao University, Jimei Road, No.668, Xiamen, Fujian, China
[b] Department of Electronic Engineering, Columbia University, NY, USA
[c] Department of Basic Sciences, Beijing Electronic Science and Technology Institute, Beijing, China
[d] School of Information Science and Technology, Xiamen University, Xiamen, Fujian, China

ARTICLE INFO

Article history:
Received 20 April 2012
Received in revised form
4 August 2012
Accepted 6 October 2012
Communicated by: Ran He
Available online 26 October 2012

Keywords:
Segmentation
Matting
Background subtraction
Moving object detection
Dynamic scenes
Neighboring Image Patches Embedding

ABSTRACT

In this paper, we address the difficult task of moving objects segmentation and matting in dynamic scenes. Toward this end, we propose a new automatic way to integrate a background subtraction (BGS) and an alpha matting technique via a heuristic seeds selection scheme. Specifically, our method can be divided into three main steps. First, we use a novel BGS method as attention mechanisms, generating many possible foreground pixels by tuning it for low false-positives and false-negatives as much as possible. Second, a connected components algorithm is used to give the bounding boxes of the labeled foreground pixels. Finally, matting of the object associated to a given bounding box is performed using a heuristic seeds selection scheme. This matting task is guided by top-down knowledge. Experimental results demonstrate the efficiency and effectiveness of our method.

1. Introduction

Accurate moving objects detection and matting in video has long been a basic problem in computer vision that attracts much research interest due to its wide variety of applications, including intelligence video surveillance [36–38,42], image editing [39], human machine interfaces, special effects in motion pictures, and so on. Despite much progress has been made in the last two decades, the problem has remained challenging to this date. Challenges in the problems include dynamic background (e.g., water ripples, swaying trees and rain), illumination changes, tremendous amount of manual interaction to provide trimap seeds, etc.

To achieve robust moving objects detection even in dynamic scenes, a number of background subtraction (BGS) methods have been proposed. BGS method can be classified into three categories: (1) the pixel-based methods that use only the pixel color or intensity information to make the decision; (2) the patch-based methods that consider some statistics of neighborhoods; (3) the hybrid methods that combine the pixel-based and the patch-based BGS methods.

1.1. Pixel-based methods

One popular technique is to model each pixel's color in a video frame with a Gaussian distribution [1]. This model does not work well in the case of dynamic scenes. To deal with this problem, Gaussian Mixture Model (GMM) [2] is used to model each pixel. But it cannot adapt to the case where the background has quick variations [3]. Numerous improvements of the original method developed by Stauffer and Grimson [2] have been proposed over the recent years and a good survey of these improvements is presented in [4]. In the W4 system [5], the background scene is statically modeled by the minimum and maximum intensity values and maximal temporal derivative for each pixel recorded over some period. A non-statistical clustering technique to construct a background model is presented in [6]. The background is encoded on a pixel-by-pixel basis and samples at each pixel are clustered into a set of codewords.

1.2. Patch-based methods

Elgammal et al. [7] are among the first to utilize the kernel density estimation (KDE) technique to model the background's color distribution, which has been successfully applied in the BGS literature. Another significant contribution of this work is the

* Corresponding author at: Department of Computer Science and Engineering, Huaqiao University, Jimei Road, No.668, Xiamen, Fujian, China.
Tel.: +86 592 616 2556.
E-mail address: bnzhong@gmail.com (B. Zhong).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.10.001

incorporation of the spatial constraints into the formulation of foreground classification. In the second phase of their approach, the pixel values that could be explained away by distributions of neighboring pixels are reclassified as background, allowing for greater resilience against dynamic backgrounds. In [8], the background and foreground models are first constructed via a KDE technique separately, which are then used competitively in a MAP-MRF decision framework. Mittal and Paragios [9] propose the use of variable bandwidths for KDE to enable modeling of arbitrary shapes of the underlying density in a more natural way. Parag and Elgammal [10] use a boosting method (RealBoost) to choose the best feature to distinguish the foreground for each of the areas in the scene. However, one key problem with the KDE techniques is their high computational requirement due to the large number of samples needed to model the background. A Bayesian framework that incorporates spectral, spatial, and temporal features to characterize the background appearance is proposed in [11]. Under this framework, the background is represented by the most significant and frequent features, i.e., the principal features, at each pixel. Seki et al. [12] propose a BGS method based on the co-occurrence of image variations, which can be regarded as narrowing the background image variations by estimating the background image pattern in each image block from the neighboring image patterns in the input image. Some authors model the background using texture features. Heikkila and Pietikainen [13] propose an approach based on the discriminative LBP histogram. However, simple grayscale operations make LBP rather sensitive to noise and it is also not so efficient on uniform regions. Yao and Odobez [14] propose a multiple layer background model which makes use of the LBP texture feature and color feature. In [15], the background is firstly divided into three types of regions—flat, sketchable and textured region according to a primal sketch representation. Then, the three types of regions are modeled respectively by Mixture of Gaussians, image primitives and LBP histograms. Finally, the geometry information obtained from camera calibrations is used to further reduce the false alarms. Ko et al. [16] have developed a BGS scheme that analyzes the temporal variation of intensity or color distributions, instead of either looking at temporal variation of point statistics, or the spatial variation of region statistics in isolation. Dalley et al. [17] introduce a new image generation model that takes into account the spatial uncertainty of dynamic background textures. In their model, they allow the pixels to be generated from nearby Gaussians. Mahadevan and Vasconcelos [18] view BGS as a problem of saliency detection: the background points are those considered not salient by suitable comparison of object and background appearance and dynamics. Zhong et al. [19] propose a novel feature extraction method, Neighboring Image Patches Embedding (NIPE), for background modeling. The NIPE feature vector, whose components are similarities between current image patch and its neighbors, describes mainly the mutual relationship between neighboring patches.

1.3. Hybrid methods

In [3,20], a three-stage algorithm is separately presented, which operates respectively at pixel, region and frame level. Tian et al. [21] first use BGS to get a set of candidate foreground pixels, then use foreground analysis to remove the false alarm pixels of the detected foreground regions. In [22], a scene is coarsely represented as the union of pixel layers and the foreground objects are detected by propagating these layers using a maximum-likelihood assignment. However, the limitations of the method are high-computational complexity and the requirement of an extra offline training step. In [32], the background is modeled as a set of warping layers, where at any given time, different layers may be visible due to the motion of an occluding layer. Since low-rank subspaces have been a powerful tool in image processing and machine learning [40], Candes et al. [41] propose a robust PCA (principal component analysis) method for background subtraction. However, the SVD (singular value decomposition) used to perform the robust PCA is too slow for the real-time applications.

In a word, the pixel-based methods, though quite successful, can be hindered by their lack of explicit encoding of statistics of neighborhoods—one might, for example, generate many false-positives (i.e., false foreground pixels) in dynamic scenes. The patch-based methods, though insensitive to the noises and the small movement of the dynamic scene, may lead to distortion at the boundary of moving objects. Although the hybrid methods can further improve the detection results by exploiting the complementary strengths of pixel-based and patch-based BGS methods, accurately extracting moving objects in dynamic scenes remains a very challenging problem.

To accurately extracting moving objects in a video, a number of authors have viewed the problem as a video matting problem, in which a high-quality alpha matte and foreground are pulled from a video sequence. However, most of the techniques require a known background (e.g., a blue screen [23,33]), several key frames [24], or tremendous amount of manual interaction to provide trimap seeds [25,26,34] and scribbles [31,35]. Please refer to [27] for a more complete image and video matting survey.

The following question naturally arises: Can we unite the strengths of a BGS and an alpha matting technique to automate the process of accurate moving objects detection and matting in dynamic scenes? Our answer is yes. In this paper, we propose a new automatic matting way to extract a high-quality alpha matte and moving objects from the dynamic scenes driven by BGS-based seeds selection. First, we use a novel BGS method as attention mechanism, generating many possible foreground pixels by tuning it for low false-positives and false-negatives as much as possible. To achieve the two goals, we adopt the following strategy: (1) We design a novel feature extraction framework, Neighboring Image Patches Embedding (NIPE) for robust and efficient moving objects detection with low false-positives. (2) As the NIPE feature vector is a patch-based feature, the NIPE-based BGS method may lead to high false-negatives at the boundary of moving objects. We further exploit the complementary strengths of the pixel-based and the patch-based BGS methods via a foreground AND operator to get the low false-negatives. Second, a connected components algorithm is used to give the bounding boxes of the labeled foreground pixels. Finally, at each extracted bounding box, we compute local trimap matting via a heuristic seeds selection scheme, in which the labeled foreground pixels in the bounding box are used as the foreground seeds and the pixels in a window of slightly larger extent than the bounding box (that) are used as the background seeds. Experimental results show that this strategy leads to accurate moving objects detection and matting even in dynamic scenes.

The rest of the paper is organized as follows: Section 2 introduces the overview of our method. Then, we describe the framework of NIPE, verify its efficacy for background modeling, and discuss the construction and maintenance of the NIPE-based background modeling in Section 3. In Section 4, we briefly introduce the GMM-based background modeling method. The detailed algorithm of heuristic seeds selection for matting is described in Section 5. Experimental results are given in Section 6, and we conclude in Section 7.

2. Overview of our method

To accurately extract moving objects in dynamic scenes, we combine a background subtraction and an alpha matting