Engineering Applications of Artificial Intelligence 64 (2017) 1–12

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Efficient local monitoring approach for the task of background subtraction

Brahim Farou [a,*], Med Nadjib Kouahla [a], Hamid Seridi [a], Herman Akdag [b]

[a] LabSTIC, Guelma University, POB 401, 24000 Guelma, Algeria
[b] LIASD, Paris 8 University, 93526 Saint-Denis, France

ARTICLE INFO

Keywords:
GMM
Background subtraction
Motion detection
Machine vision
Video surveillance
Pearson's coefficient

ABSTRACT

We present in this paper a novel and efficient method that will significantly reduce GMM drawbacks in the presence of complex and dynamic scene. The main idea is to combine global and local features to remove local variations and the instant variations in the brightness that, in most cases, decrease the performance of background subtraction models. The first step is to divide the extracted frames into several equal size blocks. Then, we apply an adaptive local monitoring algorithm for each block to control local variation using Pearson similarity measurement. When a significant environment changes are detected in one or more blocks, the parameters of GMM assigned to these blocks are updated and the parameters of the rest remain the same. We also proposed merging adjacent and invariant blocks to reduce processing time and splitting the blocks that have an intense movement to improve accuracy. Experimental results on several datasets demonstrate that the proposed approach is effective and efficient under the common problems found in background modeling, outperforming the most referred state-of-the-art background subtraction methods.

1. Introduction

Computer vision Systems dedicated to video processing require some paramount procedures such as detection and tracking of moving objects. When the objects of interest have a well-defined shape, the advanced classifiers can be used to segment objects directly from the image. These techniques work well for objects with well-defined contours, but are difficult to carry out for objects with flexible contour (Farou et al., 2015). The challenge in such systems is to achieve high sensitivity in the detection of moving objects while maintaining a good discrimination rates and low processing time (Farou et al., 2013). The intrinsic nature of environment with illumination changes, shadows, waving flags, dust, bootstrapping and ghosts make tasks even more difficult. Recently, an important efforts in this field have been focused on developing theories, methods and systems to deal with these problems and the most widely adopted techniques for handling these issues are optical flow, frame differencing and background subtraction (Farou et al., 2013). Background subtraction process usually assumes that the images extracted from video are static and can be described by a statistical model. In this case, the appearance of a new object in background will make this part inappropriate with the building model. The main idea in such approach is to model each pixel separately by a probability density function. Gaussian mixture models (GMM) are among the most commonly used approaches for detecting moving objects in a video sequence and provides a good compromise between quality and execution time compared to other methods (Yu et al., 2010). The GMM for modeling the background was proposed for the first time by Friedman and Russell (1997), and then an efficient update equations of the standard GMM was developed by Stauffer and Grimson (1999). Since that, many extensions are given to improve the model adaptation speed (Power and Schoonees, 2002; Hayman and olof Eklundh, 2003; Kaewtrakulpong and Bowden, 2001). Other works as MOG with PSO (White and Shah, 2007), MOG-IHLS (Setiawan et al., 2006), Improved MOG-FD (Wang and Suter, 2005), MOG with MRF (Schindler and Wang, 2006), S-TAPPMOG (Cristani and Murino, 2007) and ASTNA (Cristani and Murino, 2008) were also proposed to remove GMM drawback by both intrinsic and extrinsic improvements (Bouwmans, 2011). Unfortunately, local variations and instant changes in brightness remains the major problem of GMM (Hedayati et al., 2010; Zhang and Liang, 2010). In the last decade, several studies have attempted to improve the performance of GMM in environments with multiple dimming and high condensation background. Initial ideas focused on substitution of using color characteristics (Caseiro et al., 2010; Setiawan et al., 2006) or infrared camera (Seki et al., 2006). Hybrid models such as GMM and K-means (Charoenpong et al., 2010), GMM and fuzzy logic (Baf et al., 2009; Bouwmans, 2012), Markov Random Fields (Schindler and Wang,

* Corresponding author.
E-mail addresses: farou@ymail.com (B. Farou), kouahla.nadjib@yahoo.fr (M.N. Kouahla), seridihamid@yahoo.fr (H. Seridi), Herman.akdag@ai.univ-paris8.fr (H. Akdag).

http://dx.doi.org/10.1016/j.engappai.2017.05.013
Received 16 October 2016; Received in revised form 25 April 2017; Accepted 18 May 2017
0952-1976/

2006), GMM and adaptive background (Doulamis et al., 2010; Sheng and Cui, 2008) have been proposed to overcome GMM drawbacks. Other works have focused on improving the learning speed through an adaptive learning rates (Kan et al., 2010; Suo and Wang, 2008; Wang and Suter, 2005) or using a better settings (Zang and Klette, 2003). The use of real parallel operations on multiprocessor machines is another way to increase the learning speed (Li and Jing, 2011). Other systems use two backgrounds (Cheng et al., 2011) or a Multi-level approaches (Cristani et al., 2002; Cristani and Murino, 2007, 2008) to solve the problem of brightness between day and night.

Another way to improve the efficiency and robustness of the original GMM consist in using graph cuts (Sun et al., 2006; Mahamud, 2006; Tang and Miao, 2008; Chang and Hsu, 2009; Li et al., 2009, 2010; Chang et al., 2013; Miron and Badii, 2015) due to their ability to minimize the energy functions and consequently to ensure a convergence towards a global minimum with modest computational cost and without any a priori knowledge on the global shape model. Despite the fact that the graph cut has boosted the performance of GMMs (Bouwmans, 2012), they are unable to circumvent GMM problems related to local variations and instant variations in brightness.

More recently, many background subtraction techniques have been proposed to improve accuracy in a complex environment. Most of the works developed in this context try to enhance GMM either by using a spatio-temporal distribution obtained by neighborhood random sampling (Xia et al., 2016) or by a non-parametric background modeling technique that utilizes a single spatio-temporal Gaussian extracted from the previous frames for image wrapping (Viswanath et al., 2015). In the same way, a new approach based on both temporal and spatial information is proposed to reflect temporal variations of the background and to preserve multiple modes (Sun et al., 2015). For reducing computational cost, a multi-channel background model is proposed by using Gaussian filters with different variance, and updating rules for channel selection (Hu et al., 2014). In other works, memory access patterns are optimized through a low-cost adaptive background-modeling algorithm based on fine grain data parallelism (Azmat et al., 2016). Another method, combining GMM and the expectation maximization algorithm, also achieved a fast and high accuracy results with respect to standard variant (Kanagamalliga et al., 2016).

In complex scenes, some works used a local context descriptor to represent each pixel, and the choice of the pixel having the closest representation of the background is made according to a calculated distance between the pixel and each model of the system (Yang et al., 2014). Detection of the salient foreground region, which is another problem, can be identified by graph Laplacian defined over the color similarity of the image and treated by Fiedler Vectors (Perazzi et al., 2015).

Methods based on Codebook have also taken their part in the proposed improvements. The first one is done through an adaptive dictionary learning strategy based on sparse representation (Luo and Zhang, 2015), and the second one uses a linear transformation from the RGB to the YCbCr color space to reduce the effect of illumination changing and to improve the convergence (Zhou et al., 2016). Despite many algorithms have been proposed in the literature, the detection of moving objects in complex and dynamic environments is still far from being completely solved.

In this research, we focus on the detection of moving objects in video surveillance through a fixed camera. To overcome the problems mentioned above, we propose a new and efficient background subtraction method based on GMM and local background monitoring.

The remainder of this paper is organized as follows: Section 2 shows the creation of regions. Section 3 explains the contour detection and color-space transformation. Section 4 describes the block matching. Sections 5–7 give a detailed description of the proposed background subtraction method. Section 8 presents the split and merge algorithm. Section 9 presents results and discussion. Finally, Section 10 concludes the paper.

2. Establishing monitoring regions

This operation is only used in the initialization step. We divide the first frame into several equal size blocks to minimize local variations and to simplify the monitoring task. We noticed that the number of areas greatly influences on system quality. A large number of areas lead us to the starting point (pixel-based approach). In case where the number of areas is small (the size of the area is large), local variations accumulated in the same area force the system to consider the latter as an intense variation (Farou et al., 2016). In this way, all pixels belonging to the area will be updated. However, the number of blocks may change in processing time to improve system performance (see Fig. 1).

3. Contour detection and color-space transformation

In this phase, we transform the captured video into a set of images expressed in RGB color space; however, this kind of representation is not adequate because light affects the colors of objects (Suo and Wang, 2008). For this reason, we convert the RGB images to HSL model recognized to be one of the closest model of human perception and it provides a direct control of chromaticity.

The contour corresponds to local variability of the intensity values of pixels in an image. Contour detection is applied to preserve local features despite the influence of brightness. There are many contour detection techniques, but to obtain a quality results, it is preferable to use a contour detection algorithm with inherent smoothing properties that can be adapted to different conditions of noise and artifacts. Unfortunately, the context of real time processing does not allow such an operation because it requires a fast contour detection method. For this reason, Sobel filter is employed because it returns the best results in real time.

4. Block matching

The similarity between two sequences of measurement is a measure that quantifies the dependency between them. The use of similarity measure requires solving three major problems. The first one is to find the saved image that best matches the observed image; the second is to locate an object of interest in an observed image, and the third is to detect the rotational and scaling differences between the stored and observed image.

In our case, the two first problems are similar and resolved by using contour detection algorithm. Indeed, the original image is divided into a set of blocks and the similarity is applied, not to detect any type of object, but to measure the blocks dependence at the same position between the reference image and the observed image. The use of a binary image containing only contours, reduces the brightness change effect since the contours are invariant to the latter. The third problem is not probable since the camera is static and it has no zoom effect.

There are a large number of similarity measures proposed in the literature. However, each measure has its own strengths and weaknesses and a measure that performs well on one type of images may fail on other types. In this paper, we use Pearson's correlation coefficient, which is reported in the literature as the best similarity measure and works well on a wide range of images.

5. Pearson's correlation coefficient

Lets I_A the image being treatment and I_B the reference image. Pearson's correlation coefficient between I_A and I_B is defined as (Pearson, 1894):

$$r = \frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}} \tag{1}$$

where

n: is the number of pixel in the Image