

High-quality Region-based Foreground Segmentation Using a Spatial Grid of SVM Classifiers

Xiaohan Zhang , Carlos R. del Blanco , Carlos Cuevas , Fernando Jaureguizar , and Narciso García

Abstract—This paper presents a novel background modeling system that uses a spatial grid of Support Vector Machines classifiers for segmenting moving objects, which is a key step in many video-based consumer applications. The system is able to adapt to a large range of dynamic background situations since no parametric model or statistical distribution are assumed. This is achieved by using a different classifier per image region that learns the specific appearance of that scene region and its variations (illumination changes, dynamic backgrounds, etc.). The proposed system has been tested with a recent public database, outperforming other state-of-the-art algorithms.

I. INTRODUCTION

The number of consumer applications to analyze and extract valuable information from video sequences has rapidly increased along the recent years [1]. This fact is due to the explosion of electronic devices endowed with cameras. In many of these applications, foreground segmentation is a fundamental step that must be carried out to satisfactorily perform other high level tasks, such as object tracking and counting [2], classification, or augmented reality [3].

Several foreground segmentation approaches have been recently proposed [4]. On the one hand, some strategies focus on the usability and real time processing. However, they only obtain satisfactory results in low-complex sequences with no significant changes, failing in more challenging scenes characterized by dynamic backgrounds, illuminations changes, or shadows [5]. On the other hand, high-quality multimodal methods have been also proposed to improve the quality of the results in such complex scenarios [6]. These methods use kernel density estimation techniques to model the background distribution at pixelwise level. However, they have serious problems to scale up the background distribution using a region-based approach due to the famous “curse of dimensionality”. This fact avoids to use more complex distributions that could improve the final segmentation.

We propose a novel region-based spatial grid Support Vector Machines (SVM) classification system for high-quality foreground segmentation. Instead of modeling the background distribution at region level that has obvious problems due to the high dimensionality, a classification strategy is adopted, which is able to separate background and foreground classes in high dimensional spaces. An independent SVM classifier is

used per image region to predict the belonging to the background and foreground classes, using Histograms of Oriented Gradients (HOG) [7] and color histograms as inputs. Promising results have been obtained in a recent public dataset, outperforming other state-of-the-art strategies.

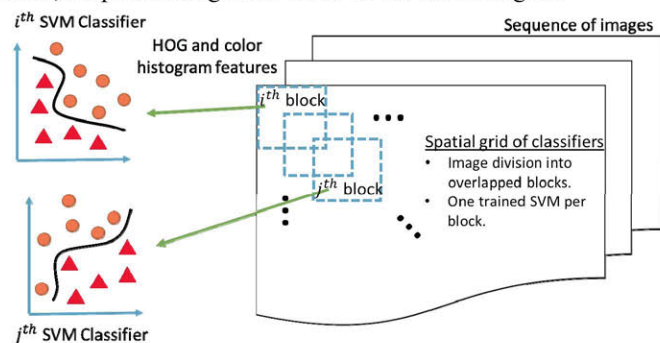


Fig. 1 Spatial grid of SVM classifiers.

II. SYSTEM DESCRIPTION

The background/foreground segmentation is accomplished via a novel classification-based framework. A spatial grid of SVM classifiers is used to segment every region of an image into background and foreground. Every classifier is applied over a specific image area with the purpose that it can learn the specific appearance of the underlying scene region and its variations. The main advantage is that no assumption is made about the expected distribution of the background/foreground, which allows to deal with multimodal backgrounds and strong changes in the scene appearance. Besides, unlike the typical approach that uses a unique classifier and a sliding window, the proposed grid of classifiers can learn the particular background appearance of every image region.

The spatial grid of SVM classifiers is structured as follows (see Fig. 1). Each acquired video frame is divided into blocks that are partially overlapped. For every block, there is a SVM classifier that learns and predicts the foreground/background classes of the underlying scene region. Since the straight output of the grid of classifiers is a downsampled version of the original image, a bilinear interpolation technique is used to obtain the final binary image, which represents the foreground/background segmentation. The used interpolation is the nearest neighbor.

The details about the used image features and the training procedure for the SVM classifiers are given as follows.

III. FEATURES AND TRAINING PROCEDURE

Color and gradient information is used to characterize

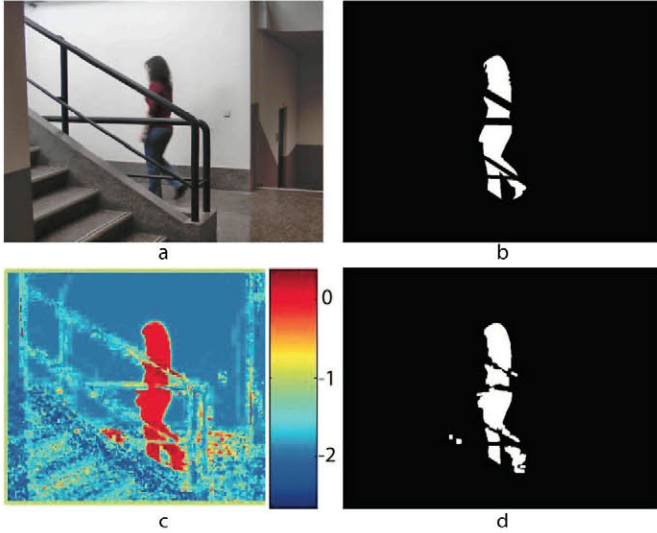


Fig. 2 (a) Original image, (b) Ground truth, (c) Confidence map, (d) Result.

the background/foreground. Color information is encoded by means of color histograms, and gradient information through HOG descriptors. Both region-based image descriptors are then combined and normalized to form the input feature vectors of the SVM classifiers.

Training samples for the background class are specific for every classifier, since they have to learn the appearance of different scene areas. This means that every classifier is trained with a different set of training features, which are extracted for corresponding image blocks. However, the same training samples are used for all the classifiers to train the foreground class. The reason is that the moving objects in the foreground could be located in any scene position. This training scheme produces much more foreground training samples than background ones. To balance both training datasets, the k-means clustering algorithm is used to reduce the number of foreground training examples.

IV. RESULTS

The proposed foreground segmentation algorithm has been tested with a public and recent dataset called Background Models Challenge. This dataset contains 29 sequences divided into 10 sequences for learning and 19 for testing. In addition, the following state-of-the-art background subtraction methods have been used for comparison purposes: Takavoli [8], Guyon [9], and Yoshinaga [10].

Figure 2 shows an example of the obtained results with the proposed algorithm. Fig. 2(a) shows the original image, Fig. 2(b) the ground truth, Fig. 2(c) the confidence map from the grid of classifiers (values greater than zero are foreground and otherwise background), and Fig. 2(d) the resulting foreground/background binary image segmentation. Note that the foreground is compact and accurate, in spite of the occluding railing and the illumination changes.

Table I shows the quantitative results for different sequences using the well-known F-measure. Observe that the proposed algorithm (called Grid-SVM) outperforms the other

state-of-the-art methods. Specifically, it achieves the best F-measure score in all the sequences with the exception of one case, where it has the second best score.

TABLE I
FOREGROUND SEGMENTATION RESULTS USING THE F-MEASURE

Sequence name	Grid-SVM	Takavoli	Guyon	Yoshinaga
112	0.887	0.713	0.867	0.870
122	0.878	0.689	0.853	0.875
212	0.870	0.699	0.867	0.851
312	0.859	0.692	0.858	0.732
322	0.874	0.695	0.861	0.904
512	0.834	0.679	0.711	0.822
522	0.899	0.695	0.812	0.896

V. CONCLUSION

A novel background subtraction system for video-based consumer applications is proposed. The main contribution is the use of a spatial grid of SVM classifiers that allows the system to adapt to a large range of dynamic background situations. No parametric model or statistical distribution are assumed, unlike other state-of-the-art approaches. Instead, a different classifier is used and adapted for every image region, which learns the appearance of a specific scene region and its variations. Promising results have been obtained with a public database, outperforming other state-of-the-art algorithms.

REFERENCES

- [1] C. Cuevas, D. Berjón, F. Morán, and N. García, "GPGPU Implementation of an Improved Nonparametric Background Modeling for Moving Object Detection Strategies", *IEEE Int. Conf. on Consumer Electronics*, pp. 23-24, 2013.
- [2] C.R. del Blanco, F. Jaureguizar, and N. García, "An efficient multiple object detection and tracking framework for automatic counting and video surveillance applications", *IEEE Trans. Consumer Electronics*, vol. 58, no. 3, pp. 857-862, 2012.
- [3] E. Komagal, A. Vinodhini, A. Srinivasan, and B. Ekava, "Real time background subtraction techniques for detection of moving objects in video surveillance system," *IEEE Int. Conf. Communication and Applications*, pp. 1-5, 2012.
- [4] S. Lee, J. Lee, M.H. Hayes, and J. Paik, "Adaptive background generation for automatic detection of initial object region in multiple color-filter aperture camera-based surveillance system," *IEEE Trans. Consumer Electronics*, vol.58, no.1, pp.104-110, 2012.
- [5] S.Y. Elhabian, K.M. El-Sayed, and S.H. Ahmed, "Moving object detection in spatial domain using background removal techniques - State-of-art," *Recent Patents on Computer Science*, vol. 1, no. 1, pp. 32-54, 2008.
- [6] E.G. Learned-Miller, M. Narayana, and A. Hanson, "Background modeling using adaptive pixelwise kernel variances in a hybrid feature space", *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2104-2111, 2012.
- [7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Conf. Computer Vision and Pattern Recognition*, 2005, vol.1, pp.886-893, 2005.
- [8] H.R. Takavoli, E.H. Rahtu, and J. Heikkilä, "Temporal saliency for fast background subtraction", *Asian Conf. Computer Vision, Workshops*, Part I, LNCS 7728, pp. 321-3362, 2013.
- [9] C. Guyon, T. Bouwmans, and E. Zahzah, "Foreground Detection via Robust Low Rank Matrix Decomposition Including Spatio-Temporal Constraint", *Asian Conf. Computer Vision, Workshops*, Part I, LNCS 7728, pp. 315-320, 2013.
- [10] S. Yoshinaga, A. Shimada, H. Nagahara, R. Taniguchi, "Background Model Based on Statistical Local Difference Pattern", *Asian Conf. Computer Vision, Workshops*, Part I, LNCS 7728, pp. 327-332, 2013.