Multi-modal Background Model Initialization

Domenico D. Bloisi[1(✉)], Alfonso Grillo[2], Andrea Pennisi[1], Luca Iocchi[1], and Claudio Passaretti[2]

[1] Sapienza University of Rome, via Ariosto, 25, 00185 Rome, Italy
{bloisi,pennisi,iocchi}@dis.uniroma1.it
[2] WT Italia, Rome, Italy
{alfonsogrillo,claudiopassaretti}@wtitalia.com

Abstract. Background subtraction is a widely used technique for detecting moving objects in image sequences. Very often background subtraction approaches assume the availability of one or more clear frames (i.e., without foreground objects) at the beginning of the image sequence in input. This strong assumption is not always correct, especially when dealing with dynamic background. In this paper, we present the results of an on-line and real-time background initialization method, called IMBS, which generates a reliable initial background model even if no clear frames are available. The accuracy of the proposed approach is calculated on a set of seven publicly available benchmark sequences. Experimental results demonstrate that IMBS generates accurate background models with respect to eight different quality metrics.

1 Introduction

Background subtraction (BS) is a popular and widely used technique that represents a fundamental building block for multiple Computer Vision applications, ranging from automatic monitoring of public spaces to augmented reality.

This work is motivated by the development of the Audio-Video Analytics Software (AVAS), joint work between WT Italia company and Sapienza University of Rome, which is designed to be a highly modular and flexible software framework for audio-video analytics. AVAS also aims at including state-of-the-art cutting-edge software components, which are properly integrated within the framework and to provide clear performance metrics and evaluation in challenging scenarios, as the one provided by scientific challenges and competitions.

A notable amount of work in BS has been done and many techniques have been developed for tackling the different aspects of the problem (see, for example, the surveys in [5,6]). In addition to the large literature, some open-source software libraries have been released, so that also non-experts can exploit BS techniques for developing Computer Vision systems. However, a number of open issues in BS still need to be addressed, in particular how to deal with sudden and gradual illumination changes (e.g., due to clouds), shadows, camera jitter (e.g., due to wind), background movement (e.g., waves on the water surface, swaying trees), and permanent and temporary changes in the background geometry (e.g., moving furniture in a room, parked cars).

V. Murino et al. (Eds.): ICIAP 2015 Workshops, LNCS 9281, pp. 485–492, 2015.
DOI: 10.1007/978-3-319-23222-5_59

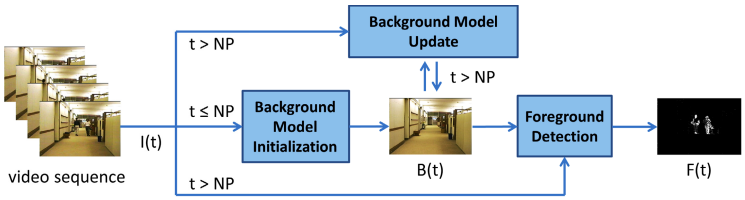

Fig. 1. Background subtraction process. *I(t)* is the current frame at time *t*, *B(t)* the background model, and *F(t)* the foreground mask. *P* is the sampling period and *N* is the total number of frames.

The BS process is carried out by comparing the current input frame with the model of the scene background and considering as foreground points the pixels that differ from the model. Thus, the problem is to generate a background model that is as reliable as possible. More formally, the BS process can be divided into three phases [4]: *background initialization*, *foreground detection*, and *model update*. Phase (1) is carried out only once, exploiting N frames at the beginning of the video sequence in input. Phases (2) and (3) are executed repeatedly as time progresses (see Fig. 1).

In contrast to the widely studied background model representation and model maintenance routines, limited attention has been given to the problem of initializing the background model [4]. In particular, often BS methods assume the availability of one or more *clean* frames at the beginning of the sequence, i.e., frames without foreground objects [11]. This is a strong assumption that is not always correct, because of continuous clutter presence. Generally, the model is initialized using the first frame or a background model over a set of training frames, which contains or do not contain foreground objects.

In this paper, we focus on the background initialization phase and describe the results of an on-line and real-time method, called Independent Multimodal Background Subtraction (IMBS) [1], when dealing with sequences where no clean frames are available. The software module evaluated in this paper is an extended version of IMBS with multi-thread optimization. For the evaluation, the test sequences provided by the Scene Background Modeling and Initialization (SBMI)[1] data set are used. Quantitative experimental results demonstrate that IMBS generates accurate background models for all the seven image sequences in SBMI with respect to eight different quality metrics, as well as very fast computation performance (over real-time).

The remainder of the paper is organized as follows. Related work, with particular emphasis on clustering-based BS methods and on existing software libraries, is discussed in the next Section 2 and the proposed method is summarized in Section 3. The results of the quantitative evaluation of IMBS and two other methods on the SBMI data set are reported in Section 4. Summary and conclusions are given in Section 5.

[1] http://sbmi2015.na.icar.cnr.it

2 Related Work

Background subtraction (BS) has been extensively studied and a rich literature with different approaches for generating accurate foreground masks exists. Some recent surveys have been realized by Hassanpour *et al.* [8], Cristani *et al.* [6], and Bouwmans [3]. From the large literature on BS algorithms, we have decided to discuss some methods adopting clustering-based solutions, since our algorithm is based on the same idea. This section describes also a set of implemented BS approaches for which open-source code and/or development libraries are available. We believe that the possibility of having the code for the algorithms, together with challenging benchmarks, is a fundamental requirement for achieving more and more reliable BS modules.

BS Clustering Approaches. Fan *et al.* in [7] perform a k-means clustering and single Gaussian model to reconstruct the background through a sequence of scene images with foreground objects. Then, based on the statistical characteristics of the background pixel regions, the algorithm detects the moving objects. In addition, an adaptive algorithm for foreground detection is used in combination with morphological operators and a region-labeling mechanism. Li *et al.* in [10] propose a method for background modeling and moving objects detection based on clustering theory. An histogram containing the pixel value over time is used to extract the moving objects, with each peak in the histogram considered as a cluster. Kumar and Sureshkumar in [9] propose a modification of the k-means algorithm for computing background subtraction in real-time processing. In their experimental results, the algorithm shows that selecting centroids can lead to a better background subtraction and it results efficient and robust for dynamic environment with new objects in it.

Differently to the above-cited clustering methods, *time* is a key factor in IMBS. Indeed, the background model is built by considering N frame samples that are collected on the basis of a time period P. The details about IMBS are given in Section 3.

Open Source BS algorithms. The possibility of having the source code of the BS methods described in the literature represents a key point towards the goals of generating more and more accurate foreground masks and widely applying this technology. OpenCV[2] library version 3 provides the source code for two BS methods:

1. MOG2: An improved adaptive Gaussian mixture model [13];
2. KNN: K-nearest neighbors background subtraction described in [14].

Bgslibrary[3] is an OpenCV based C++ BS library containing the source code for both native methods from OpenCV and several approaches published in literature. The author also provides a JAVA graphical user interface (GUI) that can be used for comparing different methods.

[2] http://opencv.org
[3] https://github.com/andrewssobral/bgslibrary

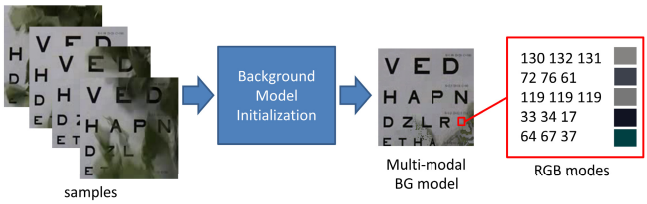

Fig. 2. IMBS stores multiple background values for each pixel.

3 IMBS Background Model Initialization

In this section, we briefly summarize the IMBS (Independent Multimodal Background Subtraction) background subtraction method experimented in this paper. Additional details can be found in [1], while the source code of IMBS is publicly available[4]. Although the method has been specifically realized for the maritime domain [2], which is characterized by non-regular and high frequency noise, IMBS can be successfully applied to many benchmark sequences, as demonstrated in Section 4.

The main idea behind IMBS is the discretization of the color distribution for each pixel, by using an on-line clustering algorithm. More specifically, for each pixel $p(i, j)$ the analysis of a set of N sample image frames is used to determine the background mode $\mathfrak{B}(i, j)$ for that pixel. $\mathfrak{B}(i, j)$ is a set of pairs $\langle c, f(c)\rangle$, where c is a value in the chosen color space (e.g., a triple in RGB or HSV space) and $f(c)$ is the number of occurrences of the value c in the sample set (see Fig. 2). After processing all the samples, only those color values that have enough occurrences are maintained in the background model. In this way, the background model contains, for each pixel, a discrete and compact multi-modal representation of its color probability distribution over time.

IMBS does not require fitting the data in some predefined distributions (e.g., Gaussian). This is the main difference with respect to a Mixture of Gaussians approach [12,13], where fitting Gaussian distributions is required and typically the number of Gaussians is limited and determined *a priori*.

Once the background model $\mathfrak{B}$ is computed, the foreground mask is determined by using a quick thresholding method: A pixel $p(i, j)$ is considered as a foreground point if the current color value is not within the distribution represented in the model, i.e., its distance from all the color values in $\mathfrak{B}(i, j)$ is above a given threshold A. IMBS requires a time $R = NP$ for creating the first background model. Then a new model, independent from the previous one, is built continuously, according to the same refresh time R.

For coping with the model update problem, IMBS adopts a conditional update policy: Given a scene sample S_k and the current foreground binary mask F, if $F(i, j) = 1$ and $S_k(i, j)$ is associated to a background mode in the background model under development, then it is labeled as a "foreground mode".

[4] http://www.dis.uniroma1.it/~bloisi/software/imbs.html

When computing the foreground, if $p(i, j)$ is associated with a foreground mode, then p is classified as a potential foreground point. Such a solution allows for identifying regions of the scene representing not moving foreground objects (i.e., temporarily static foreground objects).

4 Experimental Results

In order to experimentally evaluate the performance of our method, seven different image sequences, provided in the SBMI data set, have been used. The SBMI sequences have been extracted from multiple publicly available sequences that are frequently used in the literature to evaluate background initialization algorithms. In addition to our method, we used for comparison the results generated by two other BS methods, i.e., KNN and MOG2, whose implementation is available in OpenCV 3. For computing the results, we maintained the default parameters for KNN and MOG2 and we used the following parameters for IMBS: $P = 500ms$, $N = 20$, $D = 2$, and $A = 5$.

Accuracy Evaluation. The proposed method has been evaluated both qualitatively and quantitatively, by using the SBMI scripts for computing the results. Qualitative evaluation is illustrated in Fig. 3, where the first column contains a sample frame for each sequence; the second column contains ground truth images, included in the SBMI data set, that have been manually obtained by either choosing one of the sequence frames free of foreground objects (not included into the subsets of used frames) or by stitching together empty background regions from different sequence frames. The background model images computed with the KNN, MOG2, and IMBS methods are shown in the last three columns, respectively. For KNN and MOG2, we created the model images with the *getBackgroundImage* function, while for IMBS we created the model image by selecting, for each pixel p, the mode with the minimum distance d from the ground truth:

$$d = \arg\min_{k} |r_k - r_{GT}| + |g_k - g_{GT}| + |b_k - b_{GT}|$$

where (r_k, g_k, b_k) is one of the modes in $\mathfrak{B}$ for the pixel p and (r_{GT}, g_{GT}, b_{GT}) is the corresponding ground truth value.

Table 1 shows quantitative results obtained on the seven sequences with respect to the quality metrics suggested in SBMI (bold font is used to denote best performance in each metric). The quantitative results demonstrate that, when the nature of the scene is static (as in the first four sequences), then the three methods obtain comparable results. However, when the nature of the scene becomes highly dynamic (as in the last three sequences), then IMBS outperforms the other two methods. In particular, for the last two sequences it can be noted that IMBS obtains better results on all the considered metrics. This is due to the specific capacity of IMBS to model scenes with dynamic background, since IMBS does not consider a predefined distribution of the pixel values in the background. The last row in Table 1 shows that, when computing the average results on the

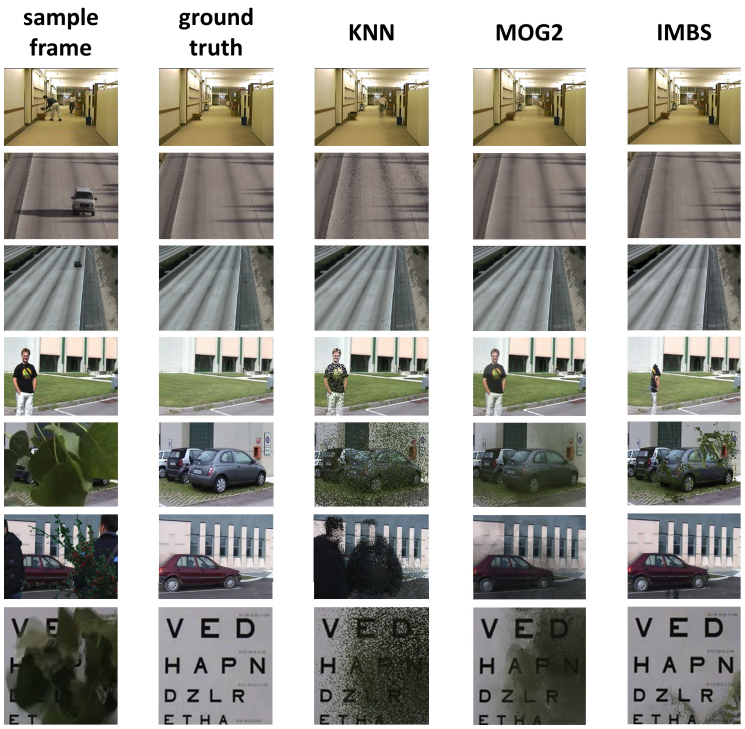

Fig. 3. Qualitative results on the Scene Background Initialization (SBI) data set.

complete SBMI data set, IMBS performs better than KNN and MOG2 with respect to all the eight considered metrics.

Computational Performance. The complete pipeline shown in Fig. 1 has been implemented to take advantage of parallel execution by using the OpenMP[5] libraries. Indeed, IMBS can process in parallel the operations for creating the background model, since an independent color distribution is generated for each pixel. In the same way, it is also possible to obtain a fast computation of the foreground mask by exploiting the parallel execution.

In order to ensure real-time performance, we measured the computational speed of IMBS by using an Intel(R) Core(TM) i7-3610QM CPU @ 2.30GHz, 8 GB RAM on 9 high-resolution video sequences of an urban scenario from the AVAS system. The results for three different high-resolution computer display standards are reported in Table 2. A computational speed of more than 30 frame per seconds can be achieved with Full High-Definition (Full HD) images. In addition, we measured also the performance on 352×240 images, obtaining a very high processing speed, i.e., more than 450 frames per second.

[5] http://openmp.org

Table 1. Results on the Scene Background Modeling and Initialization (SBMI) data set.

Sequence	Method	AGE	EPs	pEPs	CEPs	pCEPS	MS-SSIM	PSNR	CQM
Hall	KNN	3.9413	1019	0.0121	174	0.0021	0.9519	28.2208	37.4907
&	MOG2	2.4506	917	0.0109	378	0.0045	0.9833	34.3943	45.9714
Monitor	IMBS	**1.5762**	**80**	**0.0009**	**0**	**0.0000**	**0.9953**	**38.8691**	**48.3614**
	KNN	6.1277	4728	0.0616	24	0.0003	0.8506	25.1521	34.8174
HighwayI	MOG2	2.6031	174	0.0023	15	0.0002	0.9753	35.8635	**58.2889**
	IMBS	**1.9224**	**49**	**0.0006**	**7**	**0.0001**	**0.9889**	**39.5607**	54.8991
	KNN	3.2112	649	0.0085	4	0.0001	0.9851	32.0981	39.6454
HighwayII	MOG2	**2.0893**	305	0.0040	**0**	**0.0000**	**0.9946**	**36.1190**	**45.2643**
	IMBS	3.2424	**52**	**0.0007**	**0**	**0.0000**	0.9894	35.1272	38.3185
	KNN	15.9267	2212	0.0813	**345**	**0.0127**	0.8241	18.2332	30.9930
CaVignal	MOG2	16.9327	3031	0.1114	2277	0.0837	0.8136	18.5891	34.5104
	IMBS	**3.7573**	**839**	**0.0308**	532	0.0196	**0.9039**	**24.2900**	**38.9337**
	KNN	34.5615	11410	0.3962	1109	0.0385	0.6281	14.1761	25.6845
Foliage	MOG2	32.3624	19252	0.6685	15914	0.5526	**0.8038**	16.5991	31.5282
	IMBS	**19.4927**	**4305**	**0.1495**	**1093**	**0.0380**	0.8035	**18.1790**	**31.9055**
People	KNN	48.4920	36231	0.4718	22782	0.2966	0.4238	10.9196	19.8121
&	MOG2	33.8442	54590	0.7108	47110	0.6134	0.8584	16.2252	27.4728
Foliage	IMBS	**13.6299**	**8074**	**0.1051**	**1211**	**0.0158**	**0.9748**	**23.9064**	**31.8531**
	KNN	61.9389	14166	0.6832	8975	0.4328	0.4493	10.6164	22.5804
Snellen	MOG2	58.8159	15790	0.7615	14182	0.6839	0.5336	11.4143	27.0312
	IMBS	**17.5073**	**3785**	**0.1825**	**2828**	**0.1364**	**0.9521**	**21.0135**	**41.6055**
	KNN	24.8856	10059	0.2449	4773	0.1146	0.7304	19.9309	30.1462
Average	MOG2	21.2997	13437	0.3242	11411	0.2769	0.8518	24,2254	38.5810
	IMBS	**8.7326**	**2455**	**0.0671**	**810**	**0.0299**	**0.9439**	**28.7065**	**40.8395**

Table 2. IMBS computational load for different computer display standards. A comparison between mono and multi-thread solutions has been reported.

Video Standard	Frame size	FPS mono	FPS multi
Video CD	352×240	30.75	455.23
HD	1360×768	10.44	125.07
HD+	1600×900	7.62	65.46
Full HD	1920×1080	5.73	30.22

It is worth nothing that, the possibility of working with Full HD data allows for using high-level image processing routines after the foreground extraction, such as face recognition and plate identification. Moreover, with the computational speed achieved by IMBS it is possible to process simultaneously up to four HD video streams in real-time on a single PC.

5 Summary and Conclusions

In this paper, we presented the results of a fast clustering-based background subtraction method, called IMBS [1], when dealing with the problem of background initialization. The key aspect of IMBS is the capacity of generating an accurate background model even if no clear frames (i.e., without foreground objects)

are present in the image sequence in input. Experimental results on the challenging sequences of the SBMI data set demonstrate that IMBS can generate highly accurate initial background models. The results obtained by IMBS have been compared with two state-of-the-art BS methods implemented in OpenCV 3, i.e., KNN and MOG2, obtaining better results in average with respect to eight different quality metrics.

References

1. Bloisi, D., Iocchi, L.: Independent multimodal background subtraction. In: Proc. of the Third Int. Conf. on Computational Modeling of Objects Presented in Images: Fundamentals, Methods and Applications, pp. 39–44 (2012)
2. Bloisi, D.D., Iocchi, L.: ARGOS - a video surveillance system for boat trafic monitoring in venice. International Journal of Pattern Recognition and Artificial Intelligence **23**(7), 1477–1502 (2009)
3. Bouwmans, T.: Recent advanced statistical background modeling for foreground detection: A systematic survey. Recent Patents on Computer Science **4**(3), 147–176 (2011)
4. Bouwmans, T.: Traditional and recent approaches in background modeling for foreground detection: An overview. Computer Science Review **1112**, 31–66 (2014)
5. Bouwmans, T., El Baf, F., Vachon, B.: Statistical background modeling for foreground detection: A survey. In: Handbook of Pattern Recognition and Computer Vision, pp. 181–199. World scientific Publishing (2010)
6. Cristani, M., Farenzena, M., Bloisi, D., Murino, V.: Background subtraction for automated multisensor surveillance: A comprehensive review. EURASIP J. Adv. Sig. Proc. **2010**, 1–24 (2010)
7. Fan, T., Li, L., Tian, Q.: A novel adaptive motion detection based on k-means clustering. In: IEEE Int. Conf. on Computer Science and Information Technology (ICCSIT), vol. 3, pp. 136–140 (2010)
8. Hassanpour, H., Sedighi, M., Manashty, A.: Video frameś background modeling: Reviewing the techniques. Journal of Signal and Information Processing **2**(2), 72–78 (2011)
9. Kumar, A., Sureshkumar, C.: Background subtraction based on threshold detection using modified k-means algorithm. In: Int. Conf. on Pattern Recognition, Informatics and Medical Engineering (PRIME), pp. 378–382 (2013)
10. Li, Q., He, D., Wang, B.: Effective moving objects detection basedon clustering background model for video surveillance. In: Congresson Image and Signal Processing (CISP), vol. 3, pp. 656–660 (2008)
11. Maddalena, L., Petrosino, A.: Background model initialization for static cameras. In: Handbook on Background Modeling and Foreground Detection for Video Surveillance, pp. 3-1-3-16. Chapman and Hall/CRC (2014)
12. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. Int. Conf. on Computer Vision **2**, 246–252 (1999)
13. GZivkovic, Z.: Improved adaptive gaussian mixture model forbackground subtraction. In: Int. Conf. on Pattern Recognition,vol. 2, pp. 28–31 (2004)
14. Zivkovic, Z., van der Heijden, F.: Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recognition Letters **27**(7), 773–780 (2006)