DIPARTIMENTO
DI SCIENZE APPLICATE

Università
degli
Studi
di Napoli
Parthenope

Centro Direzionale di Napoli
Isola C4
80143 Napoli

dsa@uniparthenope.it

P. IVA 01877320638

Further Experimental Results with 3D_SOBS Algorithm for Moving Object Detection

Lucia Maddalena,
Alfredo Petrosino

Technical Report No.:
RT-DSA-UNIPARTHENOPE-12-01

Date:
January 2012

Dipartimento di Scienze Applicate, Università degli Studi di Napoli "Parthenope",
Centro Direzionale Isola C4, 80143, Napoli.
ICAR, National Research Council, Via P. Castellino 111, 80131, Napoli.

Further Experimental Results with 3D_SOBS Algorithm for Moving Object Detection

Lucia Maddalena and Alfredo Petrosino

I. Introduction

In [1] we propose a system that is able to distinguish moving and stopped objects in digital image sequences taken from stationary cameras. Our approach is based on self organization through a 3D artificial neural network to construct a model of the scene background and a model of the scene foreground that can handle scenes containing moving backgrounds or gradual illumination variations, helping in distinguishing between moving and stopped foreground regions.

In [2] we extend our previous research, including the complete description of a model-based framework to segment stopped foreground objects against moving foreground objects, that is independent from the model chosen for the scene background and foreground. Moreover, a formal and detailed description is provided for the 3D neural image sequence model, named 3D_SOBS (3D Self-Organizing Background Subtraction), that is targeted for modeling the background and the foreground, finalized at the detection of stopped objects.

However, the main objective of both the above described publications is the segmentation of moving and stationary objects, rather than basic moving object detection. Therefore, in this technical report our aim is to provide a thorough analysis of moving object detection accuracy achieved by the 3D_SOBS neural image sequence model for background subtraction, also through extensive experimental results on real image sequences. Specifically, qualitative and quantitative results will be described in the following and compared with those obtained by other existing approaches.

II. Qualitative Evaluation

Experimental results for moving object detection achieved by the *3D_SOBS* algorithm have been produced for several image sequences. Here we report results obtained on sequences belonging to the PETS2001 dataset, publicly available at ftp://ftp.pets.rdg.ac.uk/ pub/ PETS2001, that represent typical situations critical for video surveillance systems and have been adopted in recent literature (e.g. [3], [4]).

The testing sequence *Dataset3* consists of 5336 frames of size 768×576, taken by two different cameras, and is challenging in terms of multiple targets and significant lighting variations. For such sequence a ground truth is publicly available at http://limu.ait.kyushu-u.ac.jp/en/dataset/, consisting of binary detection masks for several sequence frames (one every fifteen frames), sub-sampled at a resolution of 320×240.

In Fig. 1 we report selected sequence frames taken from Camera1 and Camera2 (columns (a) and (c), respectively) and the corresponding moving object detection results computed by the 3D_SOBS algorithm (columns (b) and (d), respectively), where green pixels superimposed on the original image indicate detected moving pixels.

Generally, the achieved detection is quite accurate, and moving background (such as the waving shadow of the tree on the left in all the frames, as well the waving tree in the bottom center in column (c)) is successfully modeled by the 3D neural background model. Moreover, although non-uniform strong illumination changes affect the entire image sequence, the 3D neural model well adapts to such changes, including their contribution to the background model through updated weight vectors.

Few false positives can still be observed. Specifically, part of the shadows cast on the ground by moving people (e.g., in frames 1446 and 2691 - first and third row of column (b)) is recognized as foreground: this is due to the fact that in such frames illumination is quite strong and the projected shadows are too dark as compared to the ground. Lighter shadows (such as the one of the moving man in frame 1881 - second row) have been successfully recognized. Moreover, in frame 3096 taken from Camera2 (fifth row, column (d)) illumination changes are only slowly included into the background model. This is due to the fact that in previous frames (see frame 3066 - fourth row, column (d)) the area covered by the shadow was correctly detected as foreground due to a moving man; successively, illumination strongly decreased due to clouds, and selective update of 3D_SOBS algorithm has prevented the rapid inclusion of such changes into the background model.

L. Maddalena is with the National Research Council of Italy, Institute for High-Performance Computing and Networking, Naples, 80131 ITALY e-mail: lucia.maddalena@cnr.it.

A. Petrosino is with the University of Naples Parthenope, Department of Applied Science, Naples, 80143 ITALY e-mail: alfredo.petrosino@uniparthenope.it.

(a) (b) (c) (d)

Fig. 1. Selected frames from sequence *Dataset3*: (a) Original frame from Camera1 and (b) corresponding moving object detection result achieved by the 3D_SOBS algorithm (green pixels); (c) Original frame from Camera2 and (d) corresponding moving object detection result achieved by the 3D_SOBS algorithm.

TABLE I
COMPARISON OF ACCURACY VALUES FOR SEQUENCE *Dataset3-Camera1*.

	Recall	Precision	F_1
MoG [5]	0.750	0.456	0.567
Adaptive MoG [6]	0.711	0.569	0.632
Parzen [7]	0.563	0.516	0.538
RRF [9]	0.375	0.224	0.280
LBP [10]	0.343	0.565	0.427
MoG+LBP [3]	0.711	0.846	0.773
Integrated [4]	0.779	0.691	0.732
SOBS [12]	0.738	0.682	0.709
3D_SOBS [1], [2]	0.643	0.798	0.712

III. Quantitative Evaluation

Results obtained by the proposed 3D_SOBS algorithm on PETS2001 sequences have been compared with those obtained by other existing algorithms, that will be referred to as *MoG, Adaptive MoG, Parzen, RRF, LBP, MoG+LBP*, *Integrated*, and *SOBS*.

The *Mixture of Gaussian* (MoG) method [5] uses a fixed number of Gaussian distributions as a model for the values of the background pixels and an on-line approximation to update the model. The Gaussian distributions are then evaluated to determine which are the most likely to result from a background process. It can handle gradual (but not abrupt) illumination changes.

The *Adaptive MoG* method [6] is a computationally effective variant of MoG that automatically changes the number of Gaussian distributions for each pixel depending on the sequence background changes. It can handle gradual (but not abrupt) illumination changes.

The *Parzen* method [7] allows to estimate the probability density function (pdf) for probabilistic modeling of pixel values faster than the classical Kernel Density Estimation [8]. Here the adopted kernel function is a rectangular function and the pdf is estimated by incremental updating using its value in the previous frame. It can handle gradual (but not abrupt) illumination changes.

With the *Radial Reach Filter* (RRF) method [9] each pixel is classified as foreground or background based on the Radial Reach Correlation defined to evaluate local texture similarity without suffering from illumination changes.

The *Local Binary Pattern-based* (LBP) method [10] uses LBP histograms for modeling image sequence blocks. A procedure similar to the one adopted for MoG [5] is used for choosing the most probable background histograms and for updating them.

In the *Hybrid MoG and LBP-based* (MoG+LBP) method [3] the authors construct two different background models that are robust to long-term and short-term illumination changes, respectively. The first background model is the pixel-based Adaptive MoG presented in [6], while the second is a temporal variation of the spatial region-based model based on the LBP [10] obtained including predictive values into the LBP. Suitable blending rules between the two models allow to classify pixels as either foreground or background.

In the *Integrated* method [4] background modeling, based on spatio-temporal features, is obtained by a suitable combination of three complementary approaches: pixel-level, region-level, and frame-level background modeling. Pixel-level modeling is achieved by the *Parzen* method [7], region-level modeling is based on an improved RRF method where the background model is updated according to background changes, while frame-level modeling is based on a brightness normalized background model that is robust to non-uniform illumination changes.

The *SOBS* algorithm is the original 2D version of the 3D_SOBS algorithm. It has been shown [12] that SOBS adaptive model can handle scenes containing moving backgrounds, gradual illumination variations and camouflage, can include into the background model shadows cast by moving objects, and achieves robust detection for different types of videos captured with stationary cameras.

Parameter values for the 3D_SOBS algorithm have been experimentally chosen as follows: the number n of model layers has been fixed to 5; the number K of training frames has been fixed to 30; the neighborhoods size for the model updates has been fixed to 3; the variance of the Gaussian low-pass filters specifying the weights for model updates has been fixed to 0.75; the segmentation thresholds have been fixed as 0.1 for training and 0.0025 for testing; finally, the learning rates have been fixed as 1 for training and 0.05 for testing.

In Tables I and II we report accuracy results on the two views of sequence Dataset3 for all considered methods; results for the 3D_SOBS and SOBS algorithms have been obtained as average measures over the whole image sequences, while other results are those published in [3] and in [4]. The adopted metrics are defined as

$$Recall = \frac{TP}{TP+FN}, \quad Precision = \frac{TP}{TP+FP}, \quad F_1 = \frac{2 * Recall * Precision}{Recall + Precision},$$

where *TP, FN, FP* are the total number of true positive, false negative, and false positive pixels, respectively. *Recall* and *Precision* give the percentage of detected true positive pixels as compared to the total number of true positive pixels in the

TABLE II
COMPARISON OF ACCURACY VALUES FOR SEQUENCE *Dataset3-Camera2*.

	Recall	Precision	F_1
Parzen [7]	0.728	0.468	0.570
RRF [9]	0.207	0.248	0.226
MoG+LBP [3]	0.429	0.765	0.550
Integrated [4]	0.655	0.697	0.675
SOBS [12]	0.715	0.610	0.658
3D_SOBS [1], [2]	0.613	0.747	0.673

ground truth and of pixels detected by the method, respectively. Using such metrics, generally a method is considered *good* if it reaches high *Recall* values, without sacrificing *Precision*. The F_1 metric is the weighted harmonic mean of *Precision* and *Recall*, and allows to obtain a single measure that can be used to "rank" different methods. All the considered measures attain values in [0, 1], and the higher is the value, the better is the accuracy.

Results reported in Tables I and II confirm our previously reported qualitative analysis, highlighting that the 3D_SOBS algorithm performs quite well on the *Dataset3* sequence taken from both cameras. Only the methods presented in [3] and in [4], which have been specifically designed for non-uniform and abrupt illumination changes, attain in some cases higher accuracy values. The Precision-Recall curve reported in Fig. 2 confirms our analysis of the the 3D_SOBS algorithm accuracy.

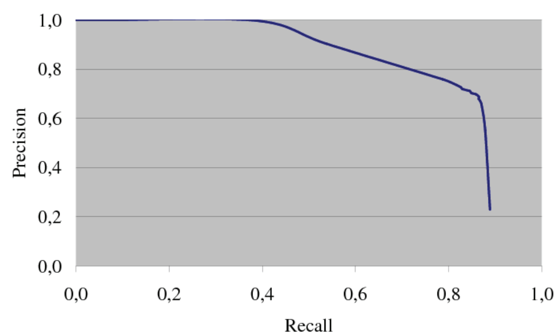

Fig. 2. Precision-Recall curve for the 3D_SOBS moving object detection algorithm on *Dataset3*.

A. Further Comparisons

Further comparisons with other existing algorithms have been performed on Wallflower sequences, publicly available at http://research.microsoft.com/ en-us/um/people/jckrumm/WallFlower/TestImages.htm. Such sequences represent some of the *canonical problems* for background subtraction highlighted in the paper of Toyama et al. [13], such as *light changes*, *moving background*, *cast shadows*, *bootstrapping*, and *camouflage*. Hand-segmented background is given for one test frame of each sequence, allowing to compare the obtained results on a pixel-by-pixel basis.
Accuracy results are reported in Table III, where results other than ours are those published in [13]. From such results we can conclude that the adopted 3D neural model is almost perfectly suited for the representation of moving background (sequence Waving Trees) and for the continuous background adaptation to gradual illumination changes (sequence Time of Day). Moreover, the color space and the metric adopted for background subtraction, chosen as in [12], are quite adequate for the discrimination of moving objects (sequence Camouflage). Although the background model is generally quite accurate, its selective update prevents the approach to quickly adapt to rapid illumination changes (sequence Light Switch), and leads to slow adaptation of the background model to the empty scene in the absence of initial empty frames (sequence Bootstrap). However, the update of weight vectors of pixels that are close to the ones erroneously detected as foreground allows to mitigate such effects in the long run.

Finally, it should be emphasized that the proposed low-level background subtraction approach does not include any higher level computer vision module (such as tracking and object recognition); this means that higher level tasks, such as the recognition of stationary moved objects (sequence Moved Object) or the recognition of new moving objects previously detected as belonging to the background (sequence Foreground Aperture), obviously cannot be quickly achieved.

Fig. 3. Selected frames from sequence *Dataset3*: (a) Original frame from Camera1 and (b) corresponding moving object detection result achieved by the 3D_SOBS algorithm (green pixels); (c) Original frame from Camera2 and (d) corresponding moving object detection result achieved by the 3D_SOBS algorithm.

TABLE III
PERFORMANCE EVALUATION IN TERMS OF FALSE NEGATIVE PIXELS (FN) AND FALSE POSITIVE PIXELS (FP) ON THE WALLFLOWER DATASET [13].

		Moved Object	Time of Day	Light Switch	Waving Trees	Camou-flage	Boot-strap	Foreground Aperture
Mean +	FN	0	949	1857	3110	4101	2215	3464
covariance	FP	0	535	15123	357	2040	92	1290
Mixture of	FN	0	1008	1633	1323	398	1874	2442
Gaussians	FP	0	20	14169	341	3098	217	530
Block	FN	0	1030	883	3323	6103	2638	1172
correlation	FP	1200	135	2919	448	567	35	1230
Temporal	FN	0	1151	752	2483	1965	2428	2049
derivative	FP	1563	11842	15331	259	3266	217	2861
Bayesian	FN	0	1018	2380	629	1538	2143	2511
decision	FP	0	562	13439	334	2130	2764	1974
Eigen-	FN	0	879	962	1027	350	304	2441
background	FP	1065	16	362	2057	1548	6129	537
Linear	FN	0	961	1585	931	1119	2025	2419
prediction	FP	0	25	13576	933	2439	365	649
Wallflower	FN	0	961	947	877	229	2025	320
	FP	0	25	375	1999	2706	365	649
SOBS	FN	0	478	1201	182	529	461	1817
	FP	1070	40	11169	120	318	286	488
3D_SOBS	FN	0	483	1036	70	408	1352	1419
	FP	1063	58	11285	195	258	270	428

REFERENCES

[1] L. Maddalena and A. Petrosino, "3d neural model-based stopped object detection," in *Image Analysis and Processing ICIAP 2009*, ser. Lecture Notes in Computer Science, P. Foggia, C. Sansone, and M. Vento, Eds. Springer Berlin / Heidelberg, 2009, vol. 5716, pp. 585–593.

[2] ——, "Stopped object detection by learning foreground model in videos," *submitted*, 2012.

[3] A. Shimada and R. Taniguchi, "Hybrid background model using spatial-temporal lbp," in *Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance AVSS 2009*, sep. 2009, pp. 19 –24.

[4] T. Tanaka, A. Shimada, R. Taniguchi, T. Yamashita, and D. Arita, "Towards robust object detection: Integrated background modeling based on spatio-temporal features," in *Asian Conference on Computer Vision*, 2009, pp. 201–212.

[5] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 1999, p. 252.

[6] A. Shimada, D. Arita, and R. Taniguchi, "Dynamic control of adaptive mixture-of-gaussians background model," in *IEEE International Conference on Video and Signal Based Surveillance AVSS 2006.*, nov. 2006, p. 5.

[7] T. Tanaka, A. Shimada, D. Arita, and R. Taniguchi, "A fast algorithm for adaptive background model construction using parzen density estimation," in *IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS 2007*, sep. 2007, pp. 528 –533.

[8] A. M. Elgammal, D. Harwood, and L. S. Davis, "Non-parametric model for background subtraction," in *ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part II*. London, UK: Springer-Verlag, 2000, pp. 751–767.

[9] Y. Satoh, S. Kaneko, Y. Niwa, and K. Yamamoto, "Robust object detection using a radial reach filter (rrf)," *Syst. Comput. Japan*, vol. 35, no. 10, pp. 63–73, 2004.

[10] M. Heikkila, M. Pietikainen, and J. Heikkila, "A texture-based method for detecting moving objects," in *British Machine Vision Conference*, 2004, pp. 187–196.

[11] L. Maddalena and A. Petrosino, "Object motion detection and tracking by an artificial intelligence approach," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 22, no. 5, pp. 915–928, January 2008.

[12] ——, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1168 –1177, jul. 2008.

[13] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," in *Seventh IEEE International Conference on Computer Vision*, vol. 1, 1999, pp. 255 –261.