Multivalued Background/Foreground Separation for Moving Object Detection

Lucia Maddalena[1] and Alfredo Petrosino[2]

[1] ICAR - National Research Council,
Via P. Castellino, 111, 80131 Naples, Italy
[2] DSA - University of Naples Parthenope,
Centro Direzionale, Isola C/4, 80143 Naples, Italy
lucia.maddalena@na.icar.cnr.it, alfredo.petrosino@uniparthenope.it

Abstract. The detection of moving objects is usually approached by background subtraction, i.e. by constructing and maintaining an up-to-date model of the background and detecting moving objects as those that deviate from such a model. We adopt a previously proposed approach to background subtraction based on self organization through artificial neural networks, that has been shown to well cope with several of the well known issues for background maintenance, featuring high detection accuracy for different types of videos taken with stationary cameras. Here we formulate a fuzzy approach to the background model update procedure to deal with decision problems typically arising when crisp settings are involved. We show through experimental results that higher accuracy values can be reached for color video sequences that represent typical situations critical for moving object detection.

Keywords: moving object detection, background subtraction, multivalued background modeling, self organization, neural network.

1 Introduction

Many computer vision applications, such as video surveillance or video compression, rely on the task of detecting moving objects in video streams, that provides the segmentation of the scene into background and foreground components.

The usual approach to moving object detection is through background subtraction, that consists in maintaining an up-to-date model of the background and detecting moving objects as those that deviate from such a model. Compared to other approaches, such as optical flow (e.g. [3]), this approach is computationally affordable for real-time applications. The main problem is its sensitivity to dynamic scene changes, and the consequent need for the background model adaptation via background maintenance. Such problem is known to be significant and difficult [14], and it has to take into account several well known issues in background maintenance, such as *light changes*, *moving background*, *cast shadows*, *bootstrapping*, and *camouflage*. Due to its pervasiveness in various contexts, background subtraction has been afforded by several researchers, and plenty of literature has been published (see surveys in [4,10,11], and more recently in [6]).

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 263–270, 2009.

In [9] we proposed the Self-Organizing Background Subtraction (SOBS) algorithm, which implements an approach to moving object detection based on the background model automatically generated by a self-organizing method without prior knowledge about the involved patterns. Such adaptive model can handle scenes containing moving backgrounds, gradual illumination variations and camouflage, has no bootstrapping limitations, can include into the background model shadows cast by moving objects, and achieves robust detection for different types of videos taken with stationary cameras.

One of the main issues to be pursued in background subtraction is the uncertainty in the detection caused by the cited background maintenance issues. Usually, crisp settings are needed to define the method parameters, and this does not allow to properly deal with uncertainty in the background model. Recently several authors have explored the adoption of fuzzy approaches to tackle different aspects of detecting moving objects. In [16] an approach using fuzzy Sugeno integral is proposed to fuse texture and color features for background subtraction, while in [2] the authors adopt the Choquet integral to aggregate the same features. In [12] a fuzzy approach to selective running average background modeling is proposed, and in [1] the authors model the background by the Type-2 Fuzzy Mixture of Gaussian Model proposed in [15].

Here we propose a fuzzy approach to the background model update procedure of SOBS algorithm, where a fuzzy function is computed, pixel-by-pixel, on the basis of the background subtraction phase. The idea is to introduce into the update phase an automatic and data dependent mechanism for further reinforcing into the background model the contribution of pixels that belong to it. It will be shown that the fuzzy approach, implemented in what will be called MSOBS (Multivalued SOBS) algorithm, further improves the accuracy of the corresponding crisp moving object detection procedure.

The paper is organized as follows. In Section 2 we detail the proposed fuzzy approach to moving object detection, describing the basic model adopted from [9] and the proposed modifications. In Section 3 we give a qualitative and quantitative evaluation of the proposed approach accuracy, comparing obtained results with those obtained by the crisp analogous approach. Conclusions are drawn in Section 4.

2 MSOBS Algorithm

The background model constructed and maintained in SOBS algorithm [9], here adopted, is based on a self organizing neural network, inspired by Kohonen [7], organized as a 2-D flat grid of neurons. Each neuron computes a function of the weighted linear combination of incoming inputs, with weights resembling the neural network learning, and can be therefore represented by a weight vector obtained collecting the weights related to incoming links. An incoming pattern is mapped to the neuron whose set of weight vectors is most similar to the pattern, and weight vectors in a neighborhood of such node are updated.

For each pixel we build a neuronal map consisting of $n \times n$ weight vectors, all represented in the HSV color space, that allows to specify colours in a way that is close to human experience of colours. Each weight vector $c_i, i = 1, \ldots, n^2$, is therefore a 3D vector, initialized to the HSV components of the corresponding pixel of the first sequence frame I_0. The complete set of weight vectors for all pixels of an image I with N rows and M columns is represented as a neuronal map $\tilde{B}$ with $n \times N$ rows and $n \times M$ columns, where adjacent blocks of $n \times n$ weight vectors correspond to adjacent pixels in image I.

By subtracting the current image from the background model, each pixel p_t of the t-th sequence frame I_t is compared to the current pixel weight vectors to determine if there exists a weight vector that matches it. The best matching weight vector is used as the pixel's encoding approximation, and therefore p_t is detected as foreground if no acceptable matching weight vector exists; otherwise it is classified as background.

Matching for the incoming pixel p_t is performed by looking for a weight vector c_m in the set $C = (c_1, \ldots, c_{n^2})$ of the current pixel weight vectors satisfying:

$$d(c_m, p_t) = \min_{i=1,\ldots,n^2} d(c_i, p_t) \leq \epsilon \ , \tag{1}$$

where the metric $d(\cdot)$ and the threshold ϵ are suitably chosen as in [9].

The best matching weight vector $c_m = \tilde{B}_t(\overline{x}, \overline{y})$ and all other weight vectors in a $n \times n$ neighborhood of the background model $\tilde{B}$ are updated according to selective weighted running average:

$$\tilde{B}_t(i,j) = (1 - \alpha_{i,j}(t))\tilde{B}_{t-1}(i,j) + \alpha_{i,j}(t) p_t(x,y), \quad \begin{array}{l} i = \overline{x} - \lfloor \frac{n}{2} \rfloor, \ldots, \overline{x} + \lfloor \frac{n}{2} \rfloor \\ j = \overline{y} - \lfloor \frac{n}{2} \rfloor, \ldots, \overline{y} + \lfloor \frac{n}{2} \rfloor \end{array} \tag{2}$$

Values for $\alpha_{i,j}(t)$ chosen in [9] can be expressed as

$$\alpha_{i,j}(t) = M(p_t)\ \alpha(t)\ w_{i,j} \ , \tag{3}$$

where $w_{i,j}$ are Gaussian weights in the $n \times n$ neighborhood, $\alpha(t)$ represents the learning factor, that is the same for each pixel of the t-th sequence frame and depends on scene variability, and $M(p_t)$ is the crisp hard-limited function

$$M(p_t) = \begin{cases} 1 & \text{if} \quad d(c_m, p_t) \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

that gives the background/foreground segmentation for pixel p_t.

It should be observed that if the best match c_m is not found, the background model $\tilde{B}$ remains unchanged. Such selectivity allows to adapt the background model to scene modifications without introducing the contribution of pixels not belonging to the background scene.

In this paper we propose to substitute the crisp function $M(\cdot)$ in eq. (3) with a fuzzy function that allows to take into account uncertainty in the background model. Specifically, we modify eq. (3) as follows:

$$\alpha_{i,j}(t) = (1 - F(p_t))\ \alpha(t)\ w_{i,j} \ , \tag{5}$$

where $F(p_t)$ is a saturating linear function given by

$$F(p_t) = \begin{cases} \dfrac{d(c_m, p_t)}{\varepsilon} & \text{if} \quad d(c_m, p_t) \leq \varepsilon \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

Values of the function $F(\cdot)$ are normalized in $[0, 1]$; the closer is the incoming sample p_t to the background model $C = (c_1, c_2, \ldots, c_{n^2})$, the smaller is the corresponding value $F(p_t)$. Therefore, choosing $\alpha_{i,j}(t)$ as in eq. (5) ensures that the closer is the incoming sample p_t to the background model, the more it contributes to the background model update, thus further reinforcing the corresponding weight vectors.

Other choices for learning factors $\alpha_{i,j}(t)$ as a function of $F(p_t)$ could have been considered according to the above considerations; for example, in [12] the authors propose a fuzzy running average approach where learning factors are chosen as an exponential function that, adapted to our case and notation, is given by

$$\alpha_{i,j}(t) = \exp(-5 * F(p_t)) \; \alpha(t) \; w_{i,j} \; . \tag{7}$$

Summarizing, the background subtraction and update procedure considered in [9], as well as the modified versions proposed in the present paper, can all be stated in a similar way. Given an incoming pixel value p_t in sequence frame I_t, the estimated background model $\tilde{B}_t$ is obtained through the following algorithm:

Background subtraction and update algorithm

`Initialize weight vectors` C `for pixel` p_0 `and store it into` $\tilde{B}_0$
for `t=1, LastFrame`
  `Find best match` c_m `in` C `to current sample` p_t `as in eq. (1)`
  `Compute learning factors` $\alpha_{i,j}(t)$
  `Update` $\tilde{B}_{t-1}$ `in the neighborhood of` c_m `as in eq. (2)`
endfor

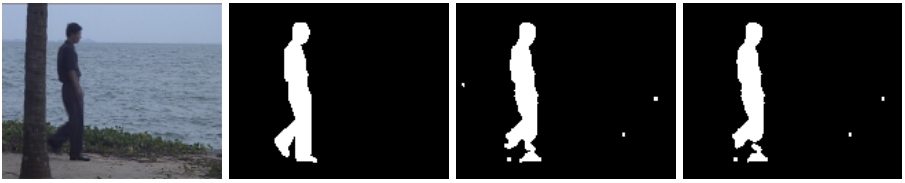
The original crisp SOBS algorithm is obtained if learning factors $\alpha_{i,j}(t)$ for the update step are chosen as in eq. (3), while the proposed multivalued algorithm, in the following denoted as MSOBS, is obtained if learning factors are chosen as in eq. (5). An alternative version of the multivalued algorithm, in the following denoted as MSOBS2, can be obtained if learning factors are chosen as in eq. (7). Results for all such algorithms will be compared in the following Section 3.

3 Experimental Results

Experimental results for moving object detection using the proposed approach have been produced for several image sequences. Here we describe two different publicly available sequences[1], that represent typical situations critical for moving object detection, and present qualitative and quantitative results obtained with the proposed method.

Sequence WS (*Water Surface*), where a person walks at a waterfront, has been chosen in order to test our method ability to cope with moving background (the

[1] http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html



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

Fig. 1. Segmentation of sequence WS: (a) test image; (b) ground truth; (c) MSOBS detection mask; (d) SOBS detection mask

water surface). The sequence contains 633 frames of 160×128 spatial resolution. One representative frame is reported in Fig. 1-(a) and the corresponding hand-segmented background in Fig. 1-(b).

The indoor scene of sequence MR (*Curtain*) consists in an initially empty meeting room, with a curtain slightly blowing in the wind, where a man comes in and starts making his presentation. The sequence consists of 2964 frames of 160×128 spatial resolution, and we consider the hand-segmented background mask available for frame 1773. The considered test image and the related binary ground truth are reported in Figs. 2-(a) and 2-(b), respectively.

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

Fig. 2. Segmentation of sequence MR: (a) test image; (b) ground truth; (c) MSOBS detection mask; (d) SOBS detection mask

The foreground masks computed by the proposed MSOBS algorithm are reported in Fig. 1-(c) for sequence WS and in Fig. 2-(c) for sequence MR, and those computed by SOBS algorithm are reported in Figs. 1-(d) and 2-(d), respectively. From such results it can be observed that both MSOBS and SOBS algorithms were quite successful in modeling the moving background (water surface and blowing curtain) and in detecting the moving person, both in the outdoor and in the indoor scene.

Results obtained by the proposed MSOBS algorithm have been compared with those obtained by the corresponding crisp SOBS algorithm in terms of different pixel-based metrics, namely *Precision, Recall*, and F_1.

Recall, also known as *detection rate*, gives the percentage of detected true positive pixels as compared to the total number of true positive pixels in the ground truth:

$$Recall = \frac{tp}{tp + fn},$$

where tp is the total number of *true positive* pixels, fn is the total number of *false negative* pixels, and $(tp + fn)$ indicates the total number of pixels present in the ground truth.

Recall alone is not enough to compare different methods, and is generally used in conjunction with *Precision*, also known as *positive prediction*, that gives the percentage of detected true positive pixels as compared to the total number of pixels detected by the method:

$$Precision = \frac{tp}{tp + fp},$$

where fp is the total number of *false positive* pixels and $(tp + fp)$ indicates the total number of detected pixels.

Using the above mentioned metrics, generally a method is considered *good* if it reaches high *Recall* values, without sacrificing *Precision*.

Moreover, we considered the F_1 metric, also known as *Figure of Merit* or *F-measure*, that is the weighted harmonic mean of *Precision* and *Recall*:

$$F_1 = \frac{2 * Recall * Precision}{Recall + Precision}.$$

Such measure allows to obtain a single measure that can be used to "rank" different methods.

All the above considered measures attain values in [0, 1], and the higher is the value, the better is the accuracy.

Accuracy values obtained by MSOBS and SOBS algorithms for sequences WS and MR are reported in Table 1. Here we can observe that both algorithms perform quite well, and that MSOBS performs slightly better than SOBS for both the sequences. More specifically, we can observe that the fuzzy approach achieves higher Recall values, but correspondingly lower Precision values. This is due to the fact that the fuzzy approach indeed allows to reinforce the contribution to the updating of the background model of pixels close to the model, thus leading to higher Recall values. At the same time, however, the fuzzy approach reinforces also the contribution of false positive pixels, thus reducing the Precision values. Nonetheless, results obtained by MSOBS algorithm are to be preferable to those obtained with the crisp approach, as shown by higher F_1 values.

Moreover, concerning different possible choices of learning factors for background updating, in Table 1 we also compare results obtained with MSOBS and with MSOBS2. Accuracy results attained are quite similar, and therefore we prefer to adopt learning factors defined as in eq. (5), since their computation is less computationally demanding.

Since we have already shown [9] that SOBS results are generally more accurate than those obtained by several state-of-the-art background subtraction algorithms (such as *BNN* [5], *Mixture of Gaussian* [13], and the method of Li et al. [8]), we can conclude that the proposed MSOBS algorithm compares favorably with them, too.

Table 1. Accuracy values for sequences WS and MR

	WS			MR		
	Recall	Precision	F_1	Recall	Precision	F_1
SOBS	0.8606	0.9684	0.9113	0.8751	0.9496	0.9108
MSOBS	0.8788	0.9571	0.9163	0.8653	0.9679	0.9138
MSOBS2	0.8844	0.9402	0.9114	0.8968	0.9288	0.9125

4 Conclusions

In this paper we propose to extend a previously proposed method for moving object detection [9] by introducing a fuzzy learning factor into the background model update procedure. The adopted method is based on self organization through artificial neural networks, and implements the idea of using visual attention mechanisms to help detecting objects that keep the user attention in accordance with a set of predefined scene features. Here we introduce a fuzzy function, computed pixel-by-pixel on the basis of the background subtraction phase. Such function is used in the background model update phase, providing an automatic and data dependent mechanism for further reinforcing into the background model the contribution of pixels that belong to it. It has been shown that the proposed fuzzy approach further improves the accuracy of the corresponding crisp moving object detection procedure, providing experimental results on real color video sequences that represent typical situations critical for moving object detection.

References

1. Baf, F.E., Bouwmans, T., Vachon, B.: Type-2 Fuzzy Mixture of Gaussians Model: Application to Background Modeling. In: 4th International Symposium on Visual Computing, Las Vegas, USA, pp. 772–781 (2008)
2. Baf, F.E., Bouwmans, T., Vachon, B.: Fuzzy Integral for Moving Object Detection. In: IEEE International Conference on Fuzzy Systems, Hong-Kong, China, pp. 1729–1736 (2008)
3. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of Optical Flow Techniques. International Journal of Computer Vision 12, 42–77 (1994)
4. Cheung, S.-C., Kamath, C.: Robust Techniques for Background Subtraction in Urban Traffic Video. In: SPIE Electronic Imaging - Video Communications and Image Processing, pp. 881–892 (2004)
5. Culibrk, D., Marques, O., Socek, D., Kalva, H., Furht, B.: Neural Network Approach to Background Modeling for Video Object Segmentation. IEEE Transactions on Neural Networks 18, 1614–1627 (2007)
6. Elhabian, S.Y., El-Sayed, K.M., Ahmed, S.H.: Moving Object Detection in Spatial Domain using Background Removal Techniques - State-of-Art. Recent Patents on Computer Science 1, 32–54 (2008)

7. Kohonen, T.: Self-Organization and Associative Memory, 2nd edn. Springer, Berlin (1988)
8. Li, L., Huang, W., Gu, I.Y.-H., Tian, Q.: Statistical Modeling of Complex Backgrounds for Foreground Object Detection. IEEE Transactions on Image Processing 13, 1459–1472 (2004)
9. Maddalena, L., Petrosino, A.: A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. IEEE Transactions on Image Processing 17, 1168–1177 (2008)
10. Piccardi, M.: Background Subtraction Techniques: a Review. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 3099–3104 (2004)
11. Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B.: Image Change Detection Algorithms: A Systematic Survey. IEEE Transactions on Image Processing 14, 294–307 (2005)
12. Sigari, M.H., Mozayani, N., Pourreza, H.R.: Fuzzy Running Average and Fuzzy Background Subtraction: Concepts and Application. International Journal of Computer Science and Network Security 8, 138–143 (2008)
13. Stauffer, C., Grimson, W.E.L.: Adaptive Background Mixture Models for Real-Time Tracking. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 246–252 (1999)
14. Toyama, K., Krumm, J., Brumitt, B., Meyers, B.: Wallflower: Principles and Practice of Background Maintenance. In: Seventh IEEE Conference on Computer Vision, pp. 255–261 (1999)
15. Zeng, J., Xie, L., Liu, Z.: Type-2 Fuzzy Gaussian Mixture Models. Pattern Recognition 41, 3636–3643 (2008)
16. Zhang, H., Xu, D.: Fusing Color and Texture Features for Background Model. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 887–893. Springer, Heidelberg (2006)