Proceedings of the 8th International Symposium on Mathematical Morphology,
Rio de Janeiro, Brazil, Oct. 10–13, 2007, MCT/INPE, v. 2, p. 71–72.
http://urlib.net/dpi.inpe.br/ismm@80/2007/06.15.16.20

A morphological gradient-based method to motion segmentation

Arnaldo Câmara Lara and
Roberto Hirata Jr.

Universidade de São Paulo (USP), Brazil
{alara,hirata}@ime.usp.br

1. Introduction

Motion segmentation is one of the most important steps in any motion-related applications because it delimits the regions of interest to the next phases. It have been studied extensively in the last years but it remains a difficult problem [1]. There are some basic approaches to deal with the problem as the comparison of consecutive frames or the modeling of a background model and compute the symmetrical difference between this model and the current frame. Optical flow can also be used to segment the moving areas [2]. In this work, we propose a novel algorithm to motion segmentation based on the background subtraction approach [3]. We estimate the background model of a video sequence using the contours of the objects in the frames instead of intensity of the pixels. Our technique has a good detection rate and it is more robust to false-positives when compared to other classical algorithms.

2. The method

The algorithm receives as input a gray-scale video sequence and saves an output binary video sequence (containing the foreground-background mask). For each frame of the input video sequence, the following steps are executed:

Background modeling: the countours of the current frame are calculated using the external morphological gradient. The background model is estimated applying the median of the contours of the last m frames. Assuming that $I = [i-m+1, i]$ when $i-m+1 \geq 1$ and $I = [1, i]$ when $i \leq m$, f_I is the set of frames in the interval I, bck_i is the background estimated in the frame i, we can define the background model as: $\mathrm{bck}_i = \mathrm{Median}(grad_r^{ext}(f_I))$, where $grad_r^{ext}(f_I)$ is the external morphological gradient operator using the structuring element r applied in the set of frames f_I.

Moving objects detection: we can calculate the contours of objects in the current frame by applying the external morphological gradient operator. Objects that moved in the video sequence have different contours from those in the background scene. Subtracting the contours that appear in the background model from the complete set of contours of the current frame, we obtain the contours of the target objects: $fe_i = grad_r^{ext}(f_i) - \mathrm{bck}_i$, where f_i is the current frame, fe_i is the contours image of the current moving objects and the sign $-$ is the morphological subtraction.

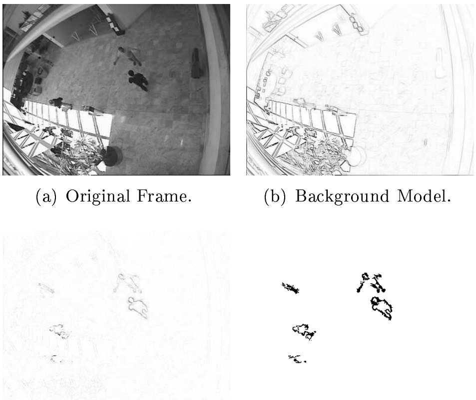
Noise filtering: the result of the last step is filtered using morphological operators to eliminate noise. Firstly, a volume filter that eliminates basins or peaks with volume less than a specified parameter is applied. Then, a contrast volume, that eliminates basins or peaks with contrast less than a specified parameter is applied. Finally, an opening by reconstruction top-hat operator is applied to extract just the peaks areas of the image. All morphological operators used are connected. Connected operators are suitable to eliminate unimportant contours in the image while preserving the significant ones and without creating new ones. This step can be formalized as: $fr_i = \mathrm{MF}(fe_i)$, where fr_i is the result image and MF corresponds to the morphological filters applied in this step.

Temporal coherence heuristic: the filtered image is converted to a binary using a trivial threshold ($h = 1$). In low-quality videos, some edges that do not belong to objects could be in this result. To keep just the edges of the target objects in the results, we apply a temporal coherence heuristic. Contours are usually in the same position in two consecutive frames. So, dilating the edges of the binary image in the last 2 frames and intersecting this dilated edges with the actual result, just edges that are approximately in the same position in the last frames are kept in the final result: $ft_i = bw_i \wedge (bw_{i-1} \oplus c) \wedge (bw_{i-2} \oplus c)$, where bw_i is the binary image of the frame i, $\wedge$ is the intersection operator, $\oplus$ is the dilation operator, c is the structuring element used in the dilation, a flat, circle with radius 3 and ft_i is the final result added to the output video sequence.

The steps of the algorithm can be followed in Figure 1.

7 3 4

Proceedings of the 8th International Symposium on Mathematical Morphology,
Rio de Janeiro, Brazil, Oct. 10–13, 2007, MCT/INPE, v. 2, p. 71–72.
http://urlib.net/dpi.inpe.br/ismm@80/2007/06.15.16.20



(a) Original Frame. (b) Background Model.

(c) Contours of the Target Objects. (d) Final result.

Figure 1. Steps of the algorithm, frame 215 of the sequence *Fight_OneManDown* of the CAVIAR Project (http://homepages.inf.ed.ac.uk/rbf/CAVIAR).

3. Results

The current work is implemented in MATLAB and it uses the MMACH Morphology Toolbox (http://www.mmorph.com). To compare our results, some classical motion segmentation algorithms were implemented in the same platform. The algorithms were implemented according to [4] and they are: 2 (DIF2) and 3-frames temporal differentiation (DIF3), background subtraction using the mean (BSME), median (BSMD), median of the last frames (BSMDF), a gaussian model (GS) and W^4 (W4). All the tested videos have a ground-truth for all frames delimiting the targets by bounding boxes. We draw a bounding box in each connected component of the results. The bounding boxes of the results are compared to the bounding boxes of the ground-truth. Each pixel can be classified as true positive (TP) - belongs to the target bounding boxes in the ground-truth and in the results; true negatives (TN) - belongs to the background in the ground-truth and in the results; false positive (FP) - belongs to the background in the ground-truth and in the target in the results and false negative (FN) - belongs to the target in the ground-truth and in the background in the results.

The *Percentage of Correct Classification* (PCC) can be defined using these measurements ($\frac{TP+TN}{TP+FP+TN+FN}$)[5]. Another performance number is the *Detection Rate* (DR) that measures the percentual of target detection in the results. If the overlap between the ground-truth bounding box and the results bounding box is over 20%, the target is considered detected. Table 1 shows the mean of these two performance numbers calculated in each frame of 45 video sequences of CAVIAR Project (excluding the training period).

Table 1. Performance numbers of motion segmentation algorithms.

Algorithm	PCC	DR
DIF2	97.04%	40.75%
DIF3	97.03%	37.91%
BSME	96.15%	95.06%
BSMD	96.92%	94.74%
BSMDF	98.26%	88.26%
GM	98.00%	80.50%
W4	96.94%	86.78%
Morph.Gradient-Based	98.36%	93.45%

4. Conclusion and future work

In this work, we present some results of a morphological gradient-based motion segmentation algorithm. The algorithm were tested in many videos of a public dataset and it showed a good overall performance. We are planning to obtain the total shape of the target objects (instead of only the contours) by using some post-processing steps.

References

[1] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, *Wallflower: Principles and Practice of Background Maintenance*, ICCV (1), 1999, pp. 255–261.

[2] W. Hu, T. Tan, L. Wang, and S. Maybank, *A Survey on Visual Surveillance of Object Motion and Behaviors*, IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews **34** (2004), no. 3, 334–352.

[3] A. Lara and R. Hirata Jr., *Motion segmentation using mathematical morphology*, SIBGRAPI, 2006, pp. 315–322.

[4] D. Hall, J. Nascimento, P. Ribeiro, E. Andrade, P. Moreno, S. Pesnel, T. List, R. Emonet, R. B. Fisher, J. Santos-Victor, and J. L. Crowley, *Comparison of target detection algorithms using adaptive background models*, Proc. 2nd Joint IEEE Int. Workshop, 2005, pp. 113–120.

[5] R. Radke, S. Andra, O. Al-Kofabi, and B. Roysam, *Image Change Detection Algorithms: A Systematic Survey*, IEEE Transactions on Image Processing **14** (2005), no. 3, 294–307.