Motion Segmentation using Mathematical Morphology

Arnaldo Câmara Lara
University of Sao Paulo
Instituto de Matemática e Estatística
Rua do Matão, 1010, 05508-090, São Paulo, Brazil
alara@ime.usp.br

Roberto Hirata Jr.
hirata@ime.usp.br

Abstract

Motion segmentation is one of the first steps in image sequence applications. The problem has been extensively studied and approached in several ways but it remains a difficult task even for sequences acquired by stationary cameras. In this paper we propose a novel algorithm that is based on Mathematical Morphology. The algorithm is simple to be implemented, less subject to lighting changes, is more robust to false positives, and it has been tested in security cameras' image sequences showing promising results. We have also implemented some algorithms described in the literature and compared their results to ours.

1. Introduction

Motion segmentation is vital to several video applications because it will delimit regions of interest to analysis phases. The problem has been approached extensively in the last two decades but it remains a difficult task [23] even when the videos are acquired by stationary cameras.

There are basically four approaches to treat this problem [24]: 1) background modelling and subtraction, 2) statistical point analysis, 3) temporal differentiation and 4) optical flow.

1. In videos generated by stationary cameras, motion segmentation can be achieved by analyzing each new frame in relation to a model of the scene background. Usually, the symmetric difference operator [3] is used to show the different pixels between the two images because it detects pixels that changed in the current frame in relation to the background model. For instance, the pixels of a chair that was in the model but is right shifted in the current frame. Some advantages of this approach are its simplicity and the good shape recovery of the moving objects. Some problems are the susceptibility to lighting changes and the computational cost to create and keep the background model [23]. A trivial solution to estimate the background is to film the scenery without moving objects or to set a specific color to the background as in the Chroma keying technique [7]. Both solutions are impossible in uncontrolled scenes. A more realistic solution is using a subset of the initial frames of the image sequence to estimate the background using some statistics like mean or median. The mean is a simple and efficient statistics, but the most susceptible to discrepant values. The median is a more robust statistics [24], however, median implies in more computational costs and memory usage. Adaptive mean has already been tried [4] to reestimate the background at each frame, becoming more robust to lighting changes.

2. In the statistical point analysis approach, statistics are computed and maintained for each pixel based on an interval of time of the video. The pixels are classified as background or foreground according to arbitrary thresholds or statistical hypothesis tests. One example of this technique is the use of the Wiener filter in the Wallflower algorithm [23]. Techniques that use this approach are very robust and popular nowadays.

3. In temporal differentiation approach, the most used technique is to compute the difference between two consecutive frames. It is simple and robust to lighting changes, but it recovers a very poor shape of the foreground and it is very susceptible to noise. One way to reduce noise is to use the information of three consecutive frames, as in [4]. The complete shape of the target objects can be obtained by using a connected operator [5].

4. Optical Flow is defined as a 2D distribution of apparent displacement vectors among pixels of consecutive frames [21] and it is used to describe a coherent movement of points or features in an image sequence. The optical-flow-based segmentation uses features of

the vectors to find regions of similar displacement. A problem in this approach is the computational cost that inhibits its use in a real-time application without specialized hardware.

Combinations of the four approaches above have been tried, for instance, in [4], the authors use a hybrid approach mixing temporal differentiation and background subtraction obtaining good results; in [1], the authors use temporal differentiation between the gradient [9] of two consecutive frames and a 3D human model is fitted into the result (to obtain a real-time performance, specialized hardware has been designed and used).

In this paper, the problem of motion segmentation is treated by a novel algorithm based on the background modelling and subtraction approach. The algorithm uses mainly the geometrical information of the scene extracted by operators defined by Mathematical Morphology (MM). The algorithm has been tested on two different public video sets of the CAVIAR project [10] in more than 38 video samples showing good preliminaries results. The proposed approach is simple (it uses very simple MM operations and operators), fast, robust to noise and most important, the results are less susceptible to lighting changes.

After this Introduction, we review some basic Mathematical Morphology operations and operators used in this work in Section 2. In Section 3, we present the proposed algorithm. In Section 4, we present the experimental results and in Section 5, we present some conclusions of this work.

2. Mathematical Morphology

Mathematical Morphology (MM) is a branch of nonlinear image processing and analysis [5] developed initially by Matheron [15] and Serra [20]. It is also a solid algebraic theory to study transformations between complete lattices [2, 3]. In this section, we review the most important MM concepts used in the proposed algorithm.

Let Z be the set of integers, $k \in Z, k > 0$, E be a subset of Z^2 and K be the interval $[0, k-1] \subset Z$. A gray-scale image f can be represented as an application from E to K, $f : E \longrightarrow K$. The set of all images from E to K is also denoted by K^E. An image sequence $\mathcal{V}$ is an ordered list of images, $(f_1, f_2, ..., f_n)$, where each item $f_i \in K^E$ is the i^{th} frame in the image sequence.

Let B be a subset of E and b a function from B to Z. We will call B a structuring element and b a structuring function. When the structuring function is defined from B to $\{-\infty, 0\}$ it is called flat structuring element (in contrast to non-flat structuring element) and it is denoted simply by B. Structuring elements and functions play an important role in the definition of dilations and erosions, two MM elementary operators.

To define dilation and erosion by a structuring function b, one needs to define the reflection, the dot sum and the dot subtraction [3]. The reflection of b, denoted $\breve{b}$, is given by $\breve{b}(x) = b(-x)$. Let $\dot{+}$ be an operation defined in $K \times Z$ to Z, by, any $t \in K$ and $v \in Z$:

$$t\dot{+}v = \begin{cases} 0 & \text{if } t = 0, \\ 0 & \text{if } t > 0 \text{ and } t + v \leq 0, \\ t + v & \text{if } t > 0 \text{ and } 0 \leq t + v \leq k, \\ k & \text{if } t > 0 \text{ and } t + v > k. \end{cases} \quad (1)$$

Similarly, we can define the $\dot{-}$ operation by:

$$t\dot{-}v = \begin{cases} 0 & \text{if } t < k \text{ and } t - v \leq 0, \\ t - v & \text{if } t < k \text{ and } 0 \leq t - v \leq k, \\ k & \text{if } t < k \text{ and } t - v > k, \\ k & \text{if } t = k. \end{cases} \quad (2)$$

The dilation of a gray-scale image f by the structuring function b is the function $f \oplus b$ defined for any $x \in E$ as:

$$(f \oplus b)(x) = max\{f(y)\dot{+}b(x - y) : y \in (\breve{B} + x) \cap (E)\} \quad (3)$$

Similarly, the erosion of a function f by the structuring function b is the function $f \ominus b$ defined for any $x \in E$:

$$(f \ominus b)(x) = min\{f(y)\dot{-}b(x - y) : y \in (B + x) \cap (E)\} \quad (4)$$

Erosions and dilations can be combined in different manners producing new operators [3]. Closings and openings are the result of a trivial composition of a dilation and an erosion by the same structuring function:

$$f \circ b = (f \ominus b) \oplus b \text{ (opening)} \quad (5)$$

$$f \bullet b = (f \oplus b) \ominus b \text{ (closing)} \quad (6)$$

Conditional erosion of a function f by a structuring element b_c subject to a marker $g \in K^E$ is the function $f \ominus_g b_c$ defined as:

$$f \ominus_g b_c = (f \ominus b_c) \vee g, \quad (7)$$

where $\vee$ is the union operation. Conditional dilation of f by the structuring element b_c conditioned to the marker g is the function $f \oplus_g b_c$ defined as:

$$f \oplus_g b_c = (f \oplus b_c) \wedge g \quad (8)$$

where $\wedge$ is the intersection operation. The structuring element b_c defines the connectivity to be used in the operation: a 3×3 cross defines a 4-connectivity and a 3×3 box defines a 8-connectivity [5]. A sequence of n conditional erosions and dilations are defined respectively as:

$$(f \ominus_g b_c)^n = (((f \ominus_g b_c) \ominus_g b_c) \ominus_g ... \ominus_g b_c), (n \text{ times}) \quad (9)$$

$$(f \oplus_g b_c)^n = (((f \oplus_g b_c) \oplus_g b_c) \oplus_g ... \oplus_g b_c), (n \text{ times}) \quad (10)$$

The sup-reconstruction $f\nabla_{b_c}g$ and the inf-reconstruction $f\triangle_{b_c}g$ operators of a image f by a marker g using connectivity defined by b_c are conditional erosions and dilations applied until the stability or the result of the i^{th} application is equal the previous result.

$$f\nabla_{b_c}g = (g \ominus_f b_c)^{\infty} \text{ (sup-reconstruction)} \quad (11)$$

$$f\triangle_{b_c}g = (g \oplus_f b_c)^{\infty} \text{ (inf-reconstruction)} \quad (12)$$

The closing and opening by reconstruction are defined utilizing sup- and inf-reconstructions in their implementation.

$$f \bullet_{b_c} g = (f \bullet g)\nabla_{b_c} f \text{ (closing by reconstruction)} \quad (13)$$

$$f \circ_{b_c} g = (f \circ g)\triangle_{b_c} f \text{ (opening by reconstruction)} \quad (14)$$

The h-Min and h-Max operators eliminates basins and peaks with contrast less then a specified parameter h. A structuring element b_c is used to specify the connectivity to be used. These operators are also implemented using the previous sup- and inf-reconstructions operators.

$$h\text{-Min}_{h,b_c}(f) = (f + h)\nabla_{b_c} f \quad (15)$$

$$h\text{-Max}_{h,b_c}(f) = (f - h)\triangle_{b_c} f \quad (16)$$

Analogously, the v-Min and v-Max operators eliminates basins and peaks with volumes less then a specified parameter v. They are also implemented using the previous sup- and inf-reconstructions operators.

The edges of an object can be detected by morphological operators. External gradient $grad^{ext}_{g_e}$ produces a external edges and is defined as:

$$grad^{ext}_{g_e}(f) = (f \oplus g_e) - f \quad (17)$$

The internal gradient produces internal edges:

$$grad^{int}_{g_i}(f) = f - (f \ominus g_i) \quad (18)$$

And the morphological gradient is:

$$grad_{g_e,g_i}(f) = (f \oplus g_e) - (f \ominus g_i) \quad (19)$$

The opening top-hat operator extracts the peaks of an image. It is defined by the morphological subtraction between the image f and the opening of same f by the structuring element b,

$$f\hat{\circ}b = f - (f \circ b). \quad (20)$$

The opening by reconstruction top-hat operator also extracts peaks of an image, but keeping the original contours. It is defined by the morphological subtraction between the image f and the opening by reconstruction of the same image f by the structuring element b_c,

$$f\hat{\circ}_{b_c}b = f - (f \circ_{b_c} b), \quad (21)$$

where b_c is a structuring element that specifies the used connectivity.

A connected operator is an operator that preserves the edges in the result image, or all the edges in the result image was in the original image. It don't create new edges in the original image. A flat-zone is a connected region in a image with a constant gray-level. This kind of operator always decrease the flat-zone's number in an image. This is useful to the segmentation process.

3. The Proposed Algorithm

Algorithm 1 Motion Segmentation using MM.

Input: $\mathcal{V}$, Background's Statistics
Output: $\mathcal{V}'$
for all $f_i \in \mathcal{V}$ **do**
  $grad^{ext}_r(f_i)$ is computed.
  if Background's statistics is mean **then**
    $bck_i(f_i)$ is computed. {using mean}
  else if Background's statistics is median **then**
    $bck_i(f_i)$ is computed. {using median}
  end if
  $fe_i(f_i)$ is computed.{Targets' edges}
  $vres_i(f_i)$ is computed. {Volume filters are applied to reduce noise}
  $hres_i(f_i)$ is computed. {Contrast filters are applied to reduce noise}
  $thatres_i(f_i)$ is computed. {Peaks' edges are enhanced}
  $bw_i(f_i)$ is computed. {Thresholding}
  $qt_i(f_i)$ is computed. {Temporal coherence filter}
  $qt_i(f_i)$ is added to $\mathcal{V}'$
end for

The proposed algorithm is based on the background modeling and subtraction approach. In algorithm 1 is shown a high-level description of the proposed algorithm. It receives as input a gray-scale image sequence $\mathcal{V}$ and saves an output image sequence $\mathcal{V}'$ containing the foreground objects of the input image sequence. The idea and its implementation are straightforward and can be followed in the diagram of the solution (Figure 1). For each frame f_i of $\mathcal{V}$, the morphological gradient of the current frame is computed, the background model is estimated and a subtraction is performed to get an approximation of the foreground. After that, the resulting image is filtered using morphological filters, mainly connected operators are applied to reduce noise. Connected operators do not create new edges in the resulting image, i.e, all the edges in the resulting images are in the original image. Therefore, they are important to preserve the most significant edges while filtering the irrelevant ones. Peaks are enhanced and a binary image of the

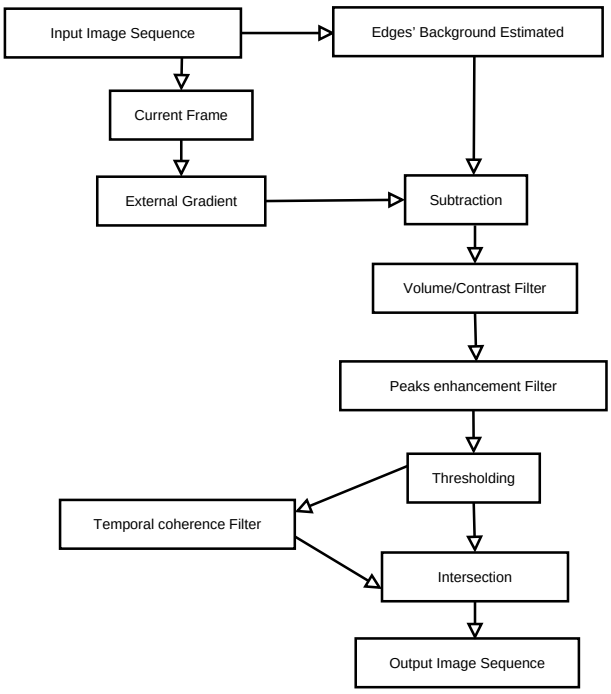


Figure 1. Diagram of the solution.

last result is produced and, finally, regions that do not show a temporal coherence are eliminated. The result reveals an approximation of the foreground's edges.

Let $f_i \in \mathcal{V}$ be an image of the sequence and let r the non-flat structuring element represented in table 1 that is a rough approximation of a semi-sphere. If one sees the image as a topographical surface, dilating it by r will change the image in all directions uniformly. So, the calculation of the external gradient using r produces a better model of the moving borders than when using a flat-structuring element. Denoting by bck_i the background model based on the morphological gradient, we compute the mean, or the median, of all frames up to frame i (a variation that is being tested computes the mean or the median from frame i_0 to i, where $1 \leq i_0 < i$). Formally, let I be the interval $[1, i]$ and $\mid I \mid$, the number of frames in the interval (in this case i),

$$\text{bck}_i^m = \frac{1}{\mid I \mid} \sum_{j=1}^{i} grad_r^{ext}(f_i) \tag{22}$$

$$\text{bck}_i^M = \text{Median}(grad_r^{ext}(f_I)), \tag{23}$$

where $grad_r^{ext}(f_I)$ means we consider all the gradient results of the sequence up to frame i. Notice that we have introduced a letter m or a letter M to the notation to distinguish mean and median, respectively.

The median computation is optimized by using a special data structure that holds the histogram of each pixel. The histogram is updated to the frame i for each pixel. The advantage is that, for a long video, is not necessary to keep all the previous frames to extract the median.

1	2	1
2	3	2
1	2	1

Table 1. Non-flat structuring element.

To have an estimate of the moving objects, we apply the morphological subtraction of the background model from the gradient of the current frame, which is here denoted by fe_i.

$$fe_i = grad_r^{ext}(f_i) - bck_i^s \tag{24}$$

where s can be the letter m or M.

To reduce noise, we first compose the morphological filters v-Min and v-Max to eliminate, respectively, basins and peaks with a volume less than a specified argument V_{min}.

$$vres_i = v\text{-Max}_{V_{min},b_c}(v\text{-Min}_{V_{min},b_c}(fe_i)) \tag{25}$$

where the structuring element b_c is a cross (3×3) meaning that the operator uses a 4-connectivity.

Next, we compose the morphological filters h-Min and h-Max to eliminate, respectively, basins and peaks with contrast less than a specified parameter H_{min}. The same cross b_c is used as the structuring element.

$$hres_i = h\text{-Max}_{H_{min},b_c}(h\text{-Min}_{H_{min},b_c}(vres_i)) \tag{26}$$

To enhanced the regions of the moving edges of the image, an opening by reconstruction top-hat operator is used to obtain the peaks of $hres_i$. The structuring element used, d, is a flat disk with radius 3 and the same structuring element b_c is used specifying a 4-connectivity.

$$thatres_i = hres_i \hat{\circ}_{b_c} d \tag{27}$$

The result is finally converted to binary using a trivial threshold 1.

$$bw_i = Th(thatres_i, 1) \tag{28}$$

where Th represents the threshold operator [9].

For very low-quality videos, some edges that do not belong to the moving objects will be in bw_i, for some i. A possible solution to this problem is to use a heuristics based on time coherence that helps to keep just the edges that were approximately in the same position for some previous frames. Because of the small displacement of the moving objects between a frame and the previous, bw_{i-1} and bw_{i-2}

are dilated by the structuring element c a flat disk with radius 3 and used as a previous information to keep the edges of bw_i. A flat disk is used as the structuring element because the direction of the edges' displacements are unknown a priori, so the edges are equally dilated in all the directions with the flat disk. The following formula formalizes this operator,

$$qt_i = bw_i \wedge (bw_{i-1} \oplus c) \wedge (bw_{i-2} \oplus c), \qquad (29)$$

where $\wedge$ is the intersection operation [5]. The final image result qt_i is added to the video output $\mathcal{V}'$.

Figure 2 shows some steps of the proposed algorithm. In 2(a) it is the original frame f_i, in 2(b) the background's model bck_i^s, in 2(c) the current frame's edges $grad_i^{ext}(f_i)$, in 2(d) the subtraction's result fe_i and, in 2(e), the final result qt_i.

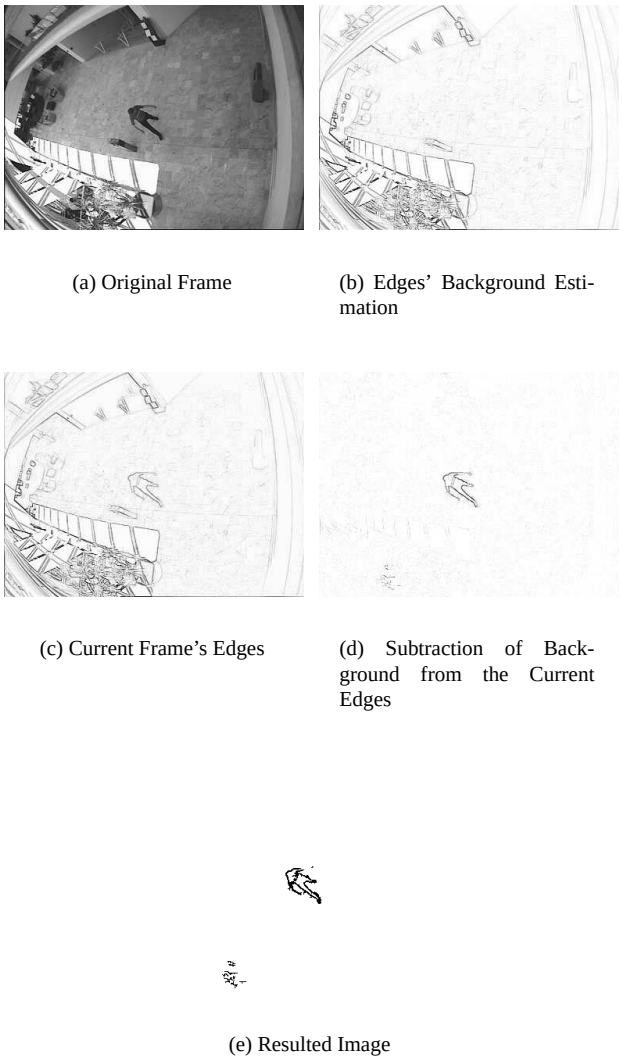
4. Experimental Results

The implementation has been done in Matlab with the SDC Morphology Toolbox for Matlab [5]. The algorithm was tested using the videos of the CAVIAR project [8]. These video sequences are divided in 2 sets. The first set has been acquired with a wide angle camera in the entrance lobby of the INRIA Labs (France) in July, 2003. The second set has been acquired in a mall in Portugal using a wide angle camera. All the video sequences have a resolution of 384 x 288 pixels at 25 frames per second. The image sequences have from 500 to 1400 frames in length and are in MPEG format.

The first set shows actors walking alone, walking together in a group, meeting another person, a group of people being split, simulating a fight, leaving a package in the scene and falling in the ground. There are some challenging features in the videos like a person behind plants in difficult position to be seen and illuminated and dark regions. These videos have low-quality with lots of noise.

The videos in the second set show real customers in a mall. There are frontal and lateral views of a corridor and the sequences show the same time frame in both views. In these videos, people are walking in the mall, entering and exiting shops, meeting another group, walking and stopping suddenly, passing out and changing direction.

No information other than the sequences has been used in the proposed algorithm. There are, of course, some basic assumptions like the sequence being acquired by a single and static camera implying that the background is static, no zoom effects is being applied and etc. Our algorithm returns almost complete edges of the moving objects in all videos. The noise was almost completely eliminated in all videos, even in the low-quality videos. Our algorithm is completely operational since the beginning of the image sequence and



(a) Original Frame

(b) Edges' Background Estimation

(c) Current Frame's Edges

(d) Subtraction of Background from the Current Edges

(e) Resulted Image

Figure 2. The steps of the algorithm.

it has a 3-frames delay to detect a new object in the scene due to the temporal coherence used in the algorithm.

To initially compare our results to the literature, we have implemented three different algorithms:

1. **Two frames differentiation:** this is an algorithm based on the temporal differentiation algorithm approach. It computes the symmetrical difference among the corresponding pixel values of two consecutive frames and then applies a threshold. Pixels (values) greater than the threshold belongs to the foreground.

2. **Two frames differentiation plus a morphological filter:** this is the same of the algorithm 1 plus a connected morphological filter that eliminates connected components with area less than a specified threshold.

3. **Background subtraction:** This is an implementation of the background model subtraction approach. Our implementation can estimate the model using mean, median, mode or adaptive mean, as described in [4]. The foreground segmentation is the result of a threshold of the symmetrical difference between the current frame and the estimated model.

4. **Hybrid algorithm:** it uses a hybrid approach [4]. The algorithm implemented uses a three-frame differentiation, an estimation of the moving areas and, finally a subtraction of a background model, as proposed in [4]. The three-frame differentiation implements the formula, $Th(\mid f_i - f_{i-1} \mid, h_1) \wedge Th(\mid f_i - f_{i-2} \mid, h_1)$, where h_1 is a threshold empirically specified. The estimation of the moving areas is done by computing, for each connected component of the three-frame differentiation result, the smallest rectangle that contains it. Finally, the symmetrical difference between this area and the background model is computed. The background model is similar to the one used before but now it is adaptive and uses a learning parameter α that is defined heuristically. Formally, it implements the formula $bck_{i+1} = \alpha * f_i + (1 - \alpha) * bck_i$, where α is usually $0.05 \leq \alpha \leq 0.15$. Connected components inside the area of interest that have pixel values similar to the model are eliminated. The remaining pixels are considered foreground.

The overall results of algorithm 1 show poor shape lateral regions of the moving objects (due to the displacement between consecutive frames). The first set of sequences are very low-quality and lots of noise appear in the results. This algorithm has the best time performance, mainly due to its simplicity. The difference between algorithm 1 and 2 is the noise reduction due to the connected filter.

The overall results of algorithm 3 have good shape recovery. However, the approach is very sensible to the statistics used in the estimation of the background. Median or mode have shown to be more robust in the tested sequences. The results are also very sensible to shadows cast that is a real problem in the sequences taken in the mall.

The overall results of algorithm 4, the hybrid algorithm, have a good shape recovery too with less noise. However, cast shadows is still a problem in the results. The results in the first set of videos are worse due to bad quality of them. One disadvantage of this algorithm is to set a good value to the learning rate α. As some sequences tested shows many different patterns of movement, a single learning rate doesn't work well in some parts of the sequences.

The overall results of the proposed algorithm show almost all edges of the foreground objects. The noise is low, even in the low-quality videos. As the algorithm uses the edges of objects in the frames and not the pixels' intensity,

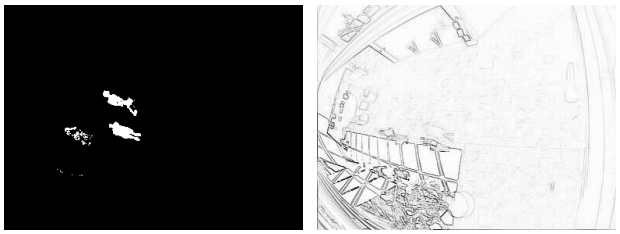
(a) Background Subtraction's Results (b) Our Results

Figure 3. Shadows in Result.

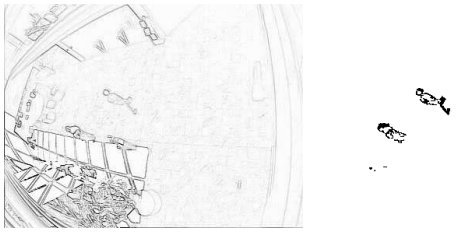
it also shows a robustness to lighting changes. In some situations, like in indoor sequences, this robustness implied in a desirable result: cast shadows do not appeared in the proposed algorithm's results. Figure 3 illustrates this situation showing cast shadows in the results of background's subtractions 3(a) compared to our results 3(b) in a indoor mall sequence. The parameters V_{min} and H_{min}, if they are too high, parts of the real borders are eliminated. But in the other side, if they are too low, noise is not adequately eliminated. The parameters used in the algorithm have been experimentally chosen. But the same values of the parameters were used in all the image sequence tested in both sets. The results are very similar in quality for both video sets. The algorithm is completely operational since the beginning of the image sequence and it has a 3-frames delay to detect a new object in the scene due to the temporal coherence used in the algorithm. When the mean is used for the background estimation, the time performance of the algorithm is about 40% better compared to median's but this is compensated by a better result with little noise. If a little noise is not a problem for the application, mean background estimation can be a plus.

Another interesting point of the proposed algorithm is the absence of false-positives in a specific situation. If an object remains immobile for a long period, it becomes part of the background model in all algorithms that uses a background model. When this object begins to move, a false-positive appears in the region of the scene that the object occupied for a long period. This situation happens because it is done a symmetric difference among the current frame and the background model that still shows the object in the model. The background model will spend some time to adapt to this change in the background. Until the adaptation completes, the false-positives will appear in the algorithms' results. The proposed algorithm has a different behavior. As it uses a simple subtraction among the edges of the current frame and the edges of the background model, it is not necessary to spend some time to adapt the background model

(a) Frame 101 original (b) Background estimated by median



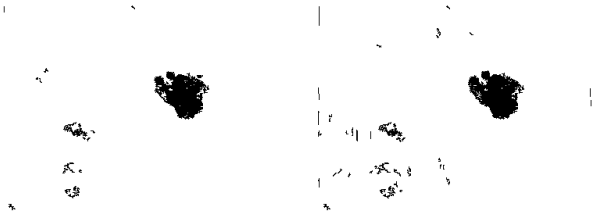
(c) Targets in background subtraction (d) Our background



(e) Edges of frame 101 (f) Our result

Figure 4. Showing the absence of false-positives in our result.

(a) Original Frame (b) 2-Frames Differentiation



(c) Background's Subtraction (d) Hybrid

(e) Our Result

Figure 5. Comparing algorithms.

to the changes in scenery. The results won't show the false-positive while the object that was immobile begins to move. The figure 4 shows the situation described above comparing a background (estimated by median) subtraction algorithm to the proposed algorithm. In 4(a) is shown the frame 101 of a tested sequence, 4(b) shows the background estimated by median of the same frame, 4(c) shows the result of the background subtraction algorithm, a false-positive is in the result in the same place that the person remained motionless for a long time, 4(d) shows the background used by our algorithm, 4(e) shows the edges of the same frame and 4(f) shows our result without the false-positive.

Figure 5 shows our results compared to another tested algorithms. In 5(a) it is the original frame, in 5(b) it is the algorithm 2's results, 5(c) shows results of the background's subtraction, 5(d) shows the hybrid's results and, finally, in 5(e) our results are shown.

All results reported here will be available at our website.

5. Conclusion

We have proposed a new algorithm based on Mathematical Morphology to segment moving objects in an image sequence. The algorithm is simple to implement and the overall performance results are promising. In terms of robustness, results are similar in both video sets tested [10], independently of the video's quality. The algorithm is also robust to lighting changes as it uses the edges of objects in the scene and not the pixels' intensity. In some situations, the results have almost no cast shadows. When compared to other algorithms, our algorithm has shown a good noise

reduction.

Other advantage of our algorithm is that there is no distinction among training and running set as some algorithms reported in the literature. The algorithm is operational since the firsts frames of the image sequence. That is, it is not necessary to have an initial period to train the algorithm. This turns the algorithm more robust and it can also be used in uncontrolled scenes where there is no way to have a sequence without target objects like in a train station or in a full-crowded corner (results shown in our website).

When compared to the background's subtraction algorithms, our algorithm shows an interesting result that is the absence of false foreground, a situation that can happen when a person or an object is motionless (becoming part of the background) and suddenly starts moving.

Using just the edges of the frames, the obtained results are encouraging. The next steps will be to implement a complete shape recovery of the foreground's objects. After that, we are planning to test our results in video sets with challenging situations as exemplified in [23] and to measure the error found in that situations.

An incomplete shape recovery is a lack of our algorithm. But a lot of applications in the Video Surveillance area don't need a perfect shape recovery.

6 Acknowledgment

The authors would like to thank for the public downloadable data that comes from the EC Funded CAVIAR project/IST 2001 37540, found at URL: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/. This first author acknowledges the CNPq, the Brazilian National Research Council, for its support via fellowship, process number 131407/2005-8.

References

[1] J. Amat, A. Casals, and M. Frigola. Stereoscopic system for human body tracking in natural scenes. In *Proc. IEEE Int. Work. on Modelling People*, 1999.

[2] G. J. F. Banon and J. Barrera. Decomposition of Mappings between Complete Lattices by Mathematical Morphology, Part I. General Lattices. *Signal Processing*, 30:299–327, 1993.

[3] J. Barrera, G. J. F. Banon, R. A. Lotufo, and R. Hirata Jr. Mmach: A mathematical morphology toolbox for khoros system. *Journal of Eletronic Imaging*, 7(1), January 1998.

[4] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson. A system for video surveillance and monitoring. Technical report, Robotics Inst., Carnegie-Mellon Univ., Pittsburgh, PA, 2000. Tech. Rep. CMU-RI-TR-00-12.

[5] E. R. Dougherty and R. A. Lotufo. *Hands-On Morphological Image Processing*. SPIE Press, Bellingham, WA, USA, 2003.

[6] J. M. Ferryman, S. J. Maybank, and A. D. Worral. Visual surveillance for moving vehicles. *Int. Journal of Computer Vision*, 37(2), June 2000.

[7] C. Finch. *Special Effects: Creating Movie Magic*. Abbeville Press, 1984.

[8] R. B. Fisher. Pets04 surveillance ground truth data set. In *Proc. Sith IEEE Int. Work. on Performance Evaluation of Tracking and Surveillance*, May 2004.

[9] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley Publishing Company, second edition, 2002.

[10] D. Hall, P. Ribeiro, E. Andrade, P. Moreno, S. Pesnel, T. List, R. Emonet, R. B. Fisher, J. S. Victor, and J. L. Crowley. Comparison of target detection algorithms using adaptative background models. In *Proc. 2nd Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, October 2005.

[11] I. Haritaoglu, D. Harwood, and L. S. Davis. W4: Who? when? where? what? a real time system for detecting and tracking people. In *3rd. Int. Conf. on Face & Gesture Recognition*, 1998.

[12] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Trans. on Systems, Man, and Cybernetics-Part C: Applic. and Reviews*, 34(3), August 2000.

[13] T. Kanade, R. T. Collins, A. J. Lipton, P. Burt, and L. Wixson. Adv. in cooperative multi-sensor video surveillance. In *Proc. DARPA Image Understanding Workshop*, 1998.

[14] Y. Lu, W. Ga, and F. Wu. Automatic video segmentation using a novel background model. In *IEEE Int. Symposium on Circuits and System*, volume 3, 2002.

[15] G. Matheron. *Random Sets and Integral Geometry*. John Wiley, New York, 1975.

[16] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi. Illumination normalization with time-dependent intrinsic images for video surveillance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(10), October 2004.

[17] T. B. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3), 2001.

[18] J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques. Motion segmentation for activity surveillance. In *1st Work. on Systems, Decision and Control Robotic Monitoring and Surveillance, Lisbon*, June 2005.

[19] R. J. Radke, S. Andra, O. Al-Kofabi, and B. Roysam. Image change detection algorithms: A systematic survey. *IEEE Trans. on Image Processing*, 14(3), March 2005.

[20] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1982.

[21] S. M. Smith and J. M. Brady. Asset-2: Real time motion segmentation and shape tracking. *IEEE Trans. on Pattern Analysis and Machine Inteligence*, 17(8), August 1995.

[22] E. Stringa. Morphological change detection algorithms for surveillance applications. In *Proc. British Machine Vision Conf.*, 2000.

[23] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *7th Int. Conf. on Computer Vision (ICCV99)*, volume 1, 1999.

[24] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36(3), 2003.

[25] C. R. Wren, A. Azerbayejani, T. Darrell, and A. P. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. on Pattern Analysis on Machine Intelligence*, 19(7), July 1997.