Benchmark for Quantitative Evaluation of Assisted Object Segmentation Methods to Image Sequences

Franklin César Flores
Department of Informatics
State University of Maringá - UEM
Av. Colombo, 5790, zip 87020-900, Maringá - PR - BRAZIL
fcflores@din.uem.br

Roberto de Alencar Lotufo
School of Electrical and Computing Engineering
University of Campinas - UNICAMP
PO Box 6101, zip 13081-970, Campinas - SP - BRAZIL
lotufo@unicamp.br

Abstract

Evaluation of segmentation methods applied to image sequences consists in the analysis of such methods according to quantitative and/or qualitative criteria, usually driven to some application. Literature proposes several metrics for quantitative evaluation of object segmentation methods to image sequences, but it is still considered an open problem, since no one of the proposed metrics is considered the standard one. More, as the best of our knowledge, there is no method in literature that does computational quantitative evaluation of assisted methods to object segmentation in image sequence. This paper introduces a benchmark to do such quantitative evaluation. This evaluation is done according to several criteria such as the robustness of segmentation and the easiness to segment the objects through the sequence. Experimental results also evaluates the robustness of the watershed from propagated markers technique.

1 Introduction

An important issue about object segmentation in image sequences is the evaluation of the segmentation results according to one or more criteria. Such criteria can be qualitative (for instance, how the segmentation mask represents the objects of interest) or quantitative (for instance, the gain provided by the segmentation mask in the compression process or the measurement of work need to do an assisted segmentation). The evaluation can be driven to some application [5, 15] and, in some cases, not only the quality of segmentation is assessed but also the tracking of the objects.

Considering the availability of a ground truth (or golden standard) segmentation, the evaluation method may be classified as *relative* or *standalone* [10, 11]: it is relative when the evaluation is done by comparing the segmentation results to the ground truth segmentation. And it is standalone when it does not occur.

One of the most popular methods to evaluate object segmentation in image sequences is the visual inspection, following subjective criteria. However, the design of computational methods to do such evaluation has been strongly motivated because subjective evaluation is expensive and depends on a set of assessment conditions.

It may be found in literature several works that propose computational metrics for quantitative assessment of object segmentation in image sequences, such as performance measures [4, 5], contextual relevance [11], perceptual relevance [15], spatial accuracy and temporal coherence [16]. Usually, such evaluation methods do not consider if the segmentation method is automatic or assisted. They do only require the segmentation masks and the input sequences to do the measurements. Note, however, that the number of works about video segmentation evaluation is far below the number of works about video segmentation itself. Evaluation design is considered an open problem [10, 16]: there is no method considered standard to do object segmentation evaluation in image sequences [11].

Literature about automatic segmentation is very rich. And several works about automatic segmentation also present an evaluation about the method in order to illustrate its accuracy. Video coding performance is one of the most popular methods [9, 13, 12]. Unfortunately, it does not occur with assisted segmentation literature: besides the small number of works about assisted segmentation of objects in image sequences found in literature, as the best of our knowledge, there is no work about quantitative evaluation of assisted segmentation methods.

Considering that the most important feature of the assisted methods is the user intervention property, an evaluation method to assisted segmentation results should quantify, for instance, the time response for each intervention, the time needed to segment objects in a given frame and the amount of work needed to assure a good/correct segmentation, i.e., the number of interventions needed in each frame to achieve the desired segmentation of the object.

This paper proposes a benchmark to quantitative and relative evaluation of assisted segmentation results. It analysis the impact that user intervention has in the framework of such methods. A method is good if it provides a robust segmentation with a minimum amount of user intervention and a good time response. The benchmark consists in a set of correlated measurements that quantifies several factors such as the number of user interferences for each frame and the time spent in each frame to finish its segmentation. The segmentation error for each frame is given by the symmetrical difference between the segmentation mask computed by the evaluated method to the frame and its respective ground truth segmentation. The amount of motion between consecutive frames, which given a notion of difficult to segment the sequence, is also computed.

The paper is organized as follows: section 2 proposes the benchmark. Section 3 presents two experiments that illustrate the application of the benckmark. This section also presents an quantitative evaluation of the watershed from propagated markers technique [6, 8]. Finally, section 4 concludes the paper.

2 The benchmark

The proposed benchmark to evaluation of assisted methods consists in the following measures:

2.1 Motion information

Given the ground truth segmentation, the motion information of an object in the sequence is given by the symmetrical difference between the segmentation masks in consecutive frames of the ground truth sequence. The amount of motion information is given by a percentile error computed in function of the symmetrical difference cited above.

Let us consider a segmentation mask as a binary image valued 1 in the pixels that belong to the segmented object (and 0, otherwise). Let gt_i be the ground truth segmentation mask for frame i. Let $\#(f)$ be the number of pixels valued 1 in a binary image f. Let ψ be the symmetrical difference between two binary images f and g, given by

$$\psi(f,g)(x) = \begin{cases} 1 & \text{if } |f(x) - g(x)| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The motion information I_i for a frame i is given by

$$I_i = \frac{\#(\psi(gt_{i-1}, gt_i))}{\#(gt_{i-1}) + \#(gt_i)}.$$

The motion information for the first frame is 0 ($I_1 = 0$).

2.2 Number of user interferences in each frame

This measure is done by counting how many times the user intervenes in the current segmentation result in a given frame. The result of this measurement depends on the kind of interface is available to the user, i.e., the options the user has to interfere in the segmentation result. In this paper, the interface used by the assessed methods consists in the addition of points or line segments as internal and/or external markers. Such objects are given by mouse clicks and drags. It is also possible to select a marker, with the mouse, and call for its deletion. In this interface, all additions and removals count as an intervention.

This measure also provides the total of interventions in the sequence and the mean of interferences for frame. Figure 3 (a) shows the amount of user intervention, for each frame, in the ground truth segmentation of *Foreman* sequence. It will be also discussed below.

2.3 Time spent in the edition of each frame

Given a frame from the image sequence, this measurement gives the time passed from the ending of segmentation in the previous frame (and the consequent start of the segmentation process to the current frame) to the ending of segmentation in the current frame. Many things occur in this time interval: the assessment of the initial segmentation of the frame (given some propagated information from the previous frame), the edition of the segmentation result by addition and removal of information, the segmentation algorithms themselves, the tracking algorithms, the visual inspection of the current segmentation in the frame, etc. The harder the task to segment a given object in the current frame, the greater the time needed to accomplish the task.

Time information is computed to each frame, but it is "global" to the frame, i.e., the time spent in a frame is the sum of all actions occurred until the object segmentation is completed. Time measurement also provides the total time spent to segment the object through the entire sequence, and the mean time spent in the edition of each frame. Figure 3 (b) shows the time spent in the ground truth segmentation of *Foreman*, for each frame. The time is given in seconds.

2.4 Segmentation error in each frame (Percentile. Compared to the Ground Truth)

The last measurement is given by the percentile error between the segmentation mask in the current frame, provided by the assessed method and its respective reference ground truth segmentation. This error is also measured in function of the number of pixels that belongs to the symmetrical difference between the segmentation masks.

Let seg_i be the segmentation mask provided by the application of the evaluated method to frame i, and let gt_i be the ground truth segmentation for frame i.

The segmentation error SE_i for a frame i is given by

$$SE_i = \frac{\#(\psi(seg_i, gt_i))}{\#(seg_i) + \#(gt_i)}.$$

A robust method usually provides a good segmentation result, with a few small segmentation errors along the object border (see examples below). However, the error information computed to a given frame is considered "global" to this frame: the evaluation does not consider the segmentation errors locally in the frame.

This measure may be done after the segmentation of the object is complete, through the sequence. It also may provide the percentile mean error for each frame. Of course, segmentation error of the ground truth is zero.

All four measurements work together in the evaluation of the assisted method. It is expected that, in the frames which have high motion information, the segmentation error is also high, except if this error is fixed by user intervention. And, if it occurs, it will have impact in the measurement of the number of interferences and in the time spent to do the corrections. In other words, it is expected that frames with high motion information presents high segmentation error or high rate of interference.

More, the robustness of the method (i.e., it works with small segmentation errors) is related to the number of user interferences: the lower the segmentation error rates, the lower the number of interferences (and, thus, the time spent in editions). Please compare the measurements shown in Fig. 3, Fig. 4 and Fig. 6.

Given the popularity and the cost of the manual segmentation of objects in image sequences, this benchmark may also be used to quantify the difficult to segment manually an image sequence. In this paper, it is also shown the difficult to do the manual ground truth segmentation for *Foreman* sequence.

3 Experimental Results

This section presents the application of the proposed benchmark in two experiments. In the first one, three methods were assessed and compared to: the manual one (that provides the ground truth segmentation) and two variations of the watershed from propagated markers technique [6, 8], supported by the binding of markers heuristics (which variations are related to the type of propagation of the markers). The second experiment evaluates the influence of parameter choices in the application of the watershed from propagated markers with binding of markers heuristics and Lucas-Kanade propagation (please see [6] for more details about this segmentation method).

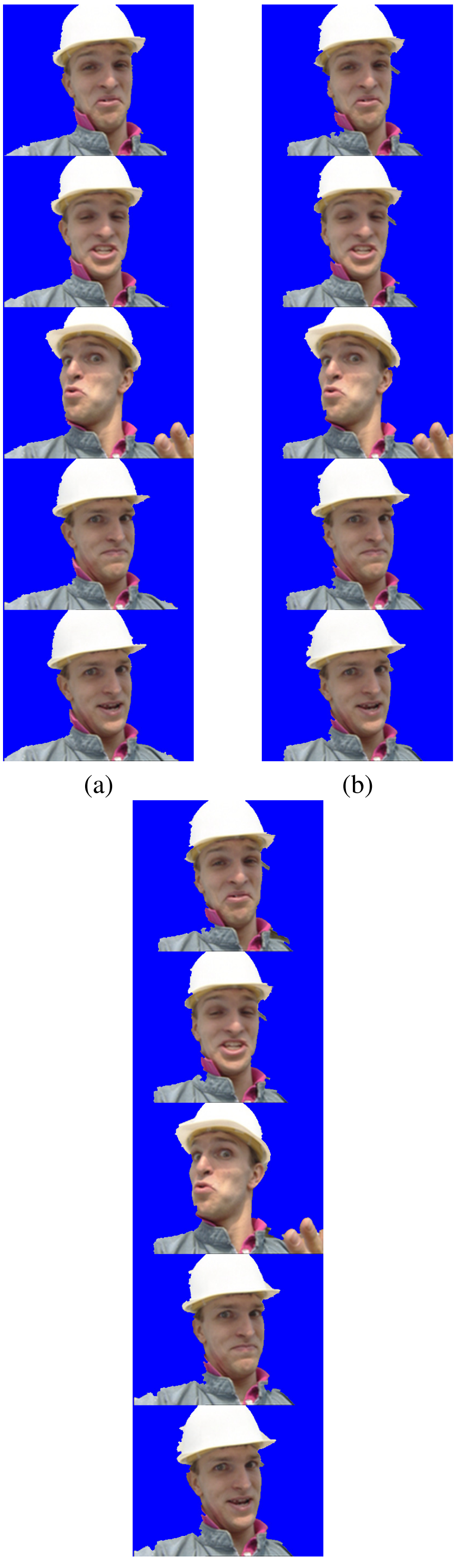
3.1 Evaluation among three assisted methods (applied to *Foreman* sequence)

Several authors use the *Foreman* sequence to demonstrate the performance of their methods, from sequence segmentation techniques to video coding [14, 13, 7]. Unfortunately, the great majority of such methods is automatic and it is hard to do a fair comparison among these methods using the proposed benchmark since it does not apply to them.

Foreman is a very interesting sequence because it presents many important test situations. The foreman himself, for instance, while he appears in the scene, moves his head in many directions (left, right and forward and backward, what provides a zooming sensation of the foreman face). Foreman head also sometimes rotates, what gives a kind of deformation of the object of interest. Camera is also moving in a smooth but uncontrolled way. If the foreman is the object of interest, note that he is composed by several objects that need to be segmented and tracked: the helmet, his face and jacket. More, the sequence presents several regions of low contrast such as the foreman shoulder, his ears and the region where helmet meets the white concrete background. *Foreman* sequence is not a trivial task to segment appropriately.

This experiment consists in to evaluate three methods applied to segment the worker in the first 150 frames of *Foreman*:

- The manual method;
- The watershed from propagated markers with binding of markers heuristic and **no propagation** (i.e., the markers are bound but are not propagated; they just stay where they are imposed).
- The watershed from propagated markers with binding of markers heuristic and **Lucas-Kanade motion estimator** [2];



(a) (b)

(c)

Figure 1. *Foreman* **segmentation (frames 70, 80, 90, 100 and 110) - (a) By manual segmentation (ground truth). (b) By watershed from propagated markers (no propagation). (c) By watershed from propagated markers (Lucas-Kanade propagation).**

The reason to apply the second method is simple: the experiment also aims to evaluate the importance of the marker propagation according to the motion of the border. The manual method was applied to create the ground truth, and it was also assessed in order to demonstrate the accuracy of the watershed from propagated markers. The interface is the same to all segmentation methods: the user may insert markers as points and/or line segments with the mouse, and select markers for removal as well. Each interference is followed by the application of classical watershed from markers [1, 3] in order to update the segmentation. Second and third experiments ran with parameters $\mathbf{m} = 10$ and $\mathbf{w} = 10$ (that are, respectively, the length of the markers and the distance between the bound markers [6]).

Figure 2 presents the motion information of *Foreman* sequence, given its ground truth segmentation. Foreman appears to move in a reasonable velocity in almost the entire sequence (note that the percentile difference between consecutive frames is about 1.5% in most frames. The percentile mean error between consecutive frames is 1.7652%). The exceptions are in the end of the sequence where the foreman appears to move quickly and in the frame **90**, where the hand of the foreman appears in the scene, disappearing in the next frame.

The ground truth is given by the manual segmentation of *Foreman* sequence (sampled in Fig. 1 (a)). It was necessary 3990 user interferences - mean of 26.6 interferences for frame - to do the manual segmentation. The total time spent to do all segmentation was 12266 seconds = 3.4073 hours. Note in Fig. 3 that the number of interferences and the time spent for frame is higher at the end of the sequence, due the foreman moves more quickly in that instant (see Fig. 2).

The watershed from propagated markers - *without propagation* - provided the segmentation results sampled in Fig. 1 (b). The application of this segmentation method required 98 interventions (mean of 0.6533 interventions for frame) an it was done in 851.72 seconds = 14.1953 minutes (mean of 5.6781 seconds each frame). Figure 4 (a) and (b), show, respectively, the number of interference and the segmentation time for each frame. Note in Fig. 4 (a) that there is a few intervals that did not need intervention. More, see the critical region at the end of both graphics where the foreman sped up. Note the amount of intervention needed in that interval and the time spent to do such intervention.

Figure 5 shows the segmentation error due the application of the watershed from propagated markers - *without propagation*. Besides bad segmentation may occur due the misplacement of the initial markers imposed to each frame, the segmentation of foreman is critical at some places. As stated above, there are several regions where the gradient is low and the watershed fails to segment correctly the object in that regions. Regions as the helmet, ears and shoulders require several interventions to be correctly segmented; it such interventions are minimal, as in both experiments with watershed from propagated markers shown in this section, the object is still segmented and tracked but segmentation error persists at some points. The percentile mean segmentation error was 2.0769% for each frame.

The last experiment was done with the application of the watershed from propagated markers - *Lucas-Kanade motion estimator* - which results are sampled in Fig. 1 (c). Figure 6 (a) and (b) show, respectively, the amount of interference required and the time spent to segment the foreman in each frame. The last experiment required 80 interventions (mean of 0.5333 interventions for frame) and it was accomplished in 1001.6 seconds = 16.6927 minutes (mean of 6.6771 seconds each frame). Again, note the intervals in the sequence where intervention was not needed (Fig. 6 (a)). Also note that the segmentation of the foreman at the end of the sequence is still critical and required a lot of intervention.

Segmentation error for each frame, due to the Lucas-Kanade motion estimation improvement, is shown in Fig. 7. Error due the misplacement of propagated markers was reduced and the segmentation of some critical regions was improved, but there still appeared some bad segmentation in several critical points that required intervention. Percentile mean segmentation error occurred due application of the watershed from propagated markers - *Lucas-Kanade motion estimator* - was 1.6434% for each frame.

Comparing the two approaches using watershed from propagated markers (both with parameters $\mathbf{m} = 10$ and $\mathbf{w} = 10$), the approach that did not propagated markers ran faster than the Lucas-Kanade propagated one (see both time graphics in Fig. 4 (b) and Fig. 6 (b)); the last approach still needs to compute the propagation when the segmentation of the current frame ends and it does not occur in the no-propagated version. However, the approach that applies the Lucas-Kanade estimation provided a lower number of interferences and a lower segmentation error for frame. Note that the segmentation error due the misplacement of the propagated markers is lower in the Lucas-Kanade supported method, since the markers are propagated in order to follow the motion of the border; a source of segmentation error in the no-propagated version is the fact that the borders usually move toward the static markers and it may make them misplaced according to their type of marker, internal or external one.

3.2 Parameters choice evaluation

The experiment in this subsection evaluates the influence of the choice of parameters to the watershed from propagated markers - supported by the binding of markers heuristics with based on Lucas-Kanade marker propagation [6]. *Foreman* sequence was segmented using several choices of

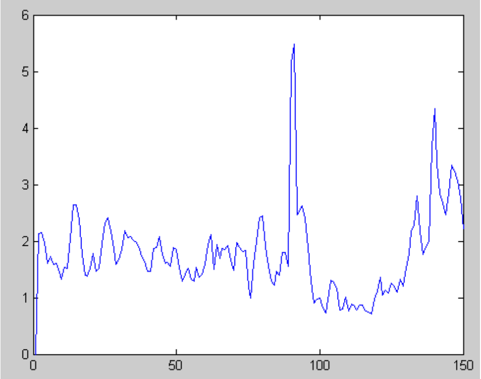
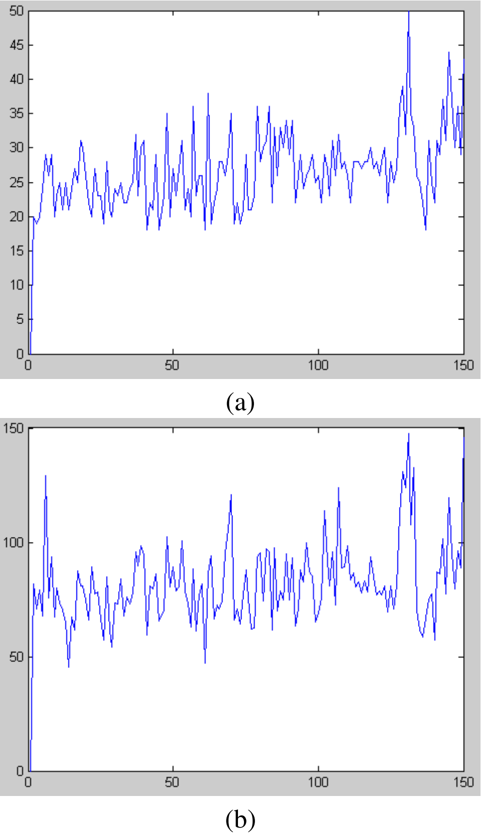


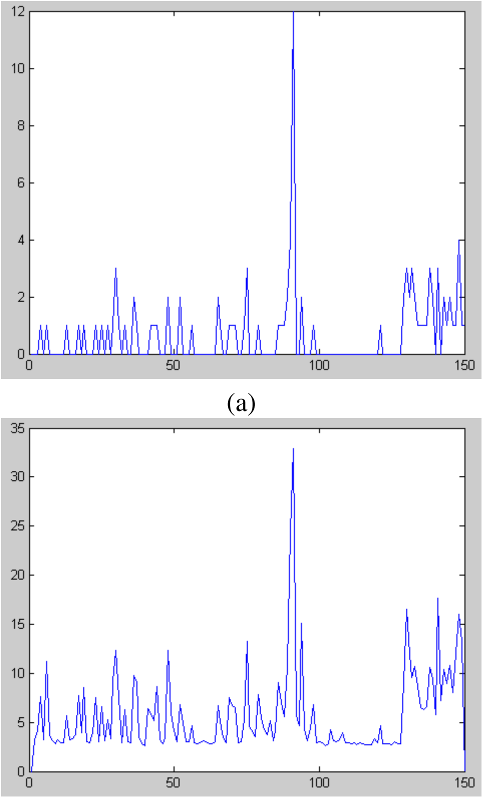
Figure 2. Motion information from *Foreman* sequence.



(a)

(b)

Figure 3. Analysis of the Ground Truth segmentation (manual one) of the *Foreman* sequence. (a) Number of user interferences in each frame. (b) Time spent in the edition of each frame (in seconds).



(a)

(b)

Figure 4. Analysis of the segmentation of *Foreman* **sequence by Watershed from Propagated Markers (No propagation). (a) Number of user interferences in each frame. (b) Time spent in the edition of each frame (in seconds).**

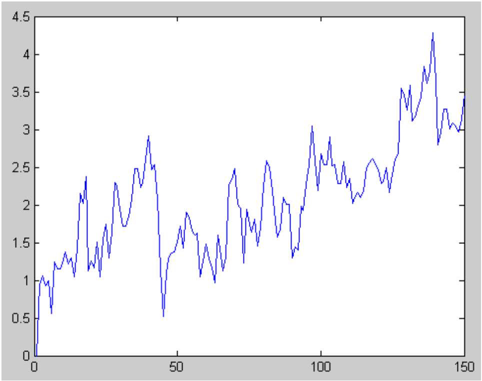


Figure 5. Analysis of the segmentation of *Foreman* **sequence by Watershed from Propagated Markers (No propagation): Segmentation Error in each frame (Compared to the Ground Truth.)**

m and **w**. Tables 1, 2 and 3 show the statistics taken from eight test cases. This experiment illustrates the trade-off in the choice of both parameters.

The suitable choices of the length of markers (**m**) and the distance between the markers in a pair (**w**) are important, since the pair of markers defines the area that will be applied to compute the displacement vector to that pair. More, for some images, the choice of both parameters is decisive when aiming a local improvement of the segmentation results.

Parameters	Total	Mean for Frame
m = 5 , **w** = 5	272	1.8133
m = 5 , **w** = 10	160	1.0667
m = 5 , **w** = 15	440	2.9333
m = 5 , **w** = 20	618	4.1200
m = 10 , **w** = 5	178	1.1867
m = 10 , **w** = 10	80	0.5333
m = 15 , **w** = 5	125	0.8333
m = 20 , **w** = 5	104	0.6933

Table 1. Quantitative Evaluation of the Choice of Parameters: Amount of Intervention.

The test cases and the collected statistics provided by the application of the benchmark illustrate some features of the watershed from propagated markers:

- Short markers require many interventions to be removed when bad propagation occurs in a given region;

- Short markers provided more regions to be propagated (i.e., more displacement vectors to be computed);

Parameters	Total (seconds)	Total (minutes)
m = 5 , **w** = 5	1817.3	30.2885
m = 5 , **w** = 10	1462.2	24.3703
m = 5 , **w** = 15	2112.9	35.2154
m = 5 , **w** = 20	2509.1	41.8190
m = 10 , **w** = 5	1251.4	20.8568
m = 10 , **w** = 10	1001.6	16.6927
m = 15 , **w** = 5	941.374	15.6896
m = 20 , **w** = 5	833.813	13.8969

Table 2. Quantitative Evaluation of the Choice of Parameters: Time Spent to Accomplish the Segmentation Task.

Parameters	Mean Time Spent	Error (Percentile)
m = 5 , **w** = 5	12.1154	1.2888
m = 5 , **w** = 10	9.7481	1.9046
m = 5 , **w** = 15	14.0862	1.8617
m = 5 , **w** = 20	16.7276	1.8736
m = 10 , **w** = 5	8.3427	1.6061
m = 10 , **w** = 10	6.6771	1.6434
m = 15 , **w** = 5	6.2758	1.7208
m = 20 , **w** = 5	5.5588	1.6712

Table 3. Quantitative Evaluation of the Choice of Parameters: Mean Time Spent for Frame (in seconds) and Mean Segmentation Error (percentile)

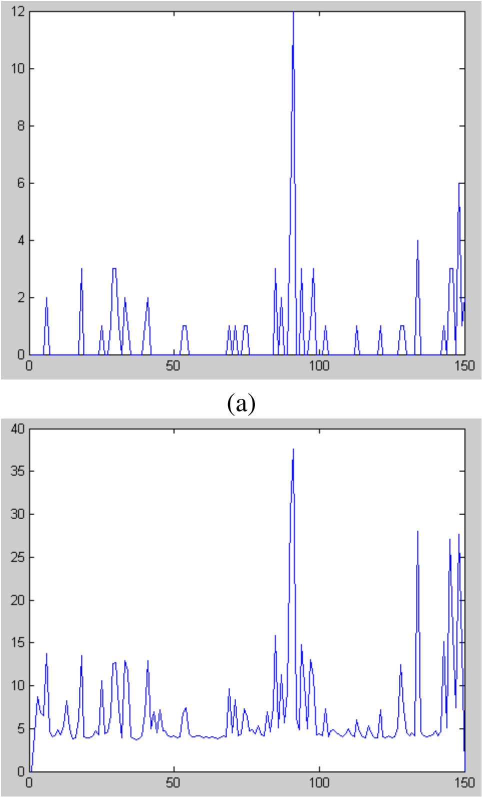
- Segmentation results require more intervention when markers are farther from the borders;

- Test cases that required more interference also took more time to be done;

- Combination of lengthy markers (lesser markers to be edited) and short distances (markers close to the border provide better segmentation) seems to be a good choice;

- Test case ($\mathbf{m} = 10$ and $\mathbf{w} = 10$) took more time to be finished that test cases ($\mathbf{m} = 15$ and $\mathbf{w} = 5$) and ($\mathbf{m} = 20$ and $\mathbf{w} = 5$). However, test case ($\mathbf{m} = 10$ and $\mathbf{w} = 10$) required less intervention and achieved a lower percentile segmentation error. Parameters ($\mathbf{m} = 10$ and $\mathbf{w} = 10$) are good choices for the segmentation of foreman in the homonymous sequence.

It could be appealing to choose lengthy markers and a short distance from them to the borders. However, there are situations when you may need greater distances or a set of short markers. Great distances from the border allow the tracking of objects with faster motion. Short markers are useful to segment deformable objects.



(a)

(b)

Figure 6. Analysis of the segmentation of *Foreman* **sequence by Watershed from Propagated Markers (Lucas-Kanade). (a) Number of user interferences in each frame. (b) Time spent in the edition of each frame (in seconds).**

4 Conclusion

This paper proposes a benchmark to evaluate assisted methods for object segmentation to image sequences. It consists in to quantify and to analyze several correlated criteria, such as the amount of intervention, the segmentation error and the time spent to complete the task. A method designed to segment objects interactively in image sequences is good if it is robust and requires a minor quantity of time and user effort.

Experiments were done in order to illustrate two applications of the benchmark: in the first one, three segmentation methods were assessed and compared to. In the second experiment, it was possible to do a quantified analysis about

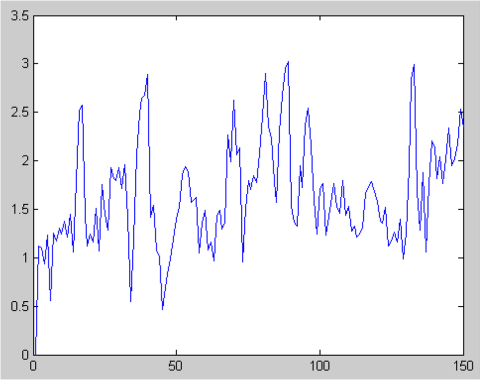


Figure 7. Analysis of the segmentation of *Foreman* sequence by Watershed from Propagated Markers (Lucas-Kanade): Segmentation Error in each frame (Compared to the Ground Truth.)

the performance of a given method - the watershed from propagated markers.

Besides the two proposed applications of the benchmark - comparison among methods and analysis of a method according its parameters - the experiments also illustrated the robustness of the watershed from propagated markers.

Future works include the analysis of additional features such as the *reproducibility* of segmentation results achieved by the assessed segmentation method.

References

[1] A. X. Falcão; J. Stolfi and R. A. Lotufo. The Image Foresting Transform: Theory, Algorithms and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):19–29, January 2004.

[2] B. Lucas and T. Kanade. An interative image registration technique with an application to stereo system. In *Proceedings of DARPA Image Understanding Workshop*, pages 121–130, 1981.

[3] S. Beucher and F. Meyer. *Mathematical Morphology in Image Processing*, chapter 12. The Morphological Approach to Segmentation: The Watershed Transformation, pages 433–481. Marcel Dekker, 1992.

[4] C. Erden; B. Sankur; A. Tekalp. Performance Measures for Video Object Segmentation and Tracking. *IEEE Transactions on Image Processing*, 13(7):937–951, July 2004.

[5] K. M. N.-P. D. D. Z. Ebrahimi. Evaluation of segmentation methods for surveillance applications. In *EUSIPCO 2000*, pages 1045–1048, Kobe, Japan, September 2000.

[6] F. C. Flores and R. A. Lotufo. Watershed from Propagated Markers Improved by a Marker Binding Heuristic. In J. B. G.J.F. Banon and U. Braga-Neto, editors, *Mathematical Morphology and its Applications to Image and Signal Processing*, Proc. ISMM'07, pages 313–323. MCT/INPE, 2007.

[7] F. Moscheni; Sushil Bhattacharjee and Murat Kunt. Spatiotemporal Segmentation Based on Region Merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9):897–915, September 1998.

[8] F. C. Flores and R. A. Lotufo. Object Segmentation in Image Sequences by Watershed from Markers: A Generic Approach. In *IEEE Proceedings of SIBGRAPI'2003*, pages 347–352, Sao Carlos, Brazil, October 2003.

[9] Z. He. Dynamic programming framework for automatic video object segmentation and vision-assisted video preprocessing. In *IEE Proceedings of Vision, Image and Signal Processing*, pages 597–603, October 2005.

[10] P. Correia and F. Pereira. Objective Evaluation of Video Segmentation Quality. *IEEE Transactions on Image Processing*, 12(2):186–200, February 2003.

[11] P. Correia and F. Pereira. Video Object Relevance Metrics for Overall Segmentation Quality Evaluation. *EURASIP Journal on Applied Signal Processing*, 2006:1–11, Article ID 82915 2006.

[12] P. Salembier and J. Ruiz. On Filters by Reconstruction for Size and Motion Simplification. In H. Talbot and R. Beare, editors, *Mathematical Morphology and its Applications to Image and Signal Processing*, Proc. ISMM'02, pages 425–434. CSIRO Publishing, 2002.

[13] P. Salembier; F. Marques; M. Pardas; J.R.Morros; I. Corset; S. Jeannin; L. Bouchard; F. Meyer; B. Marcotegui. Segmentation-Based Video Coding System Allowing the Manipulation of Objects. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(1):60–74, February 1997.

[14] P. Smith; T. Drummond and R. Cipolla. Layered Motion Segmentation and Depth Ordering by Tracking Edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(4):479–494, April 2004.

[15] P. Villegas and X. Marichal. Perceptually-weighted evaluation criteria for segmentation masks in video sequences. *IEEE Transactions on Image Processing*, 13(8):1092–1103, August 2004.

[16] R. Mech and F. Marques. Objective Evaluation Criteria for 2-D Shape Estimation Results of Moving Objects. *EURASIP Journal on Applied Signal Processing*, 2002:401–409, 2002.