Objective Evaluation Criteria for 2D-Shape Estimation Results of Moving Objects

Roland Mech

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover,
Appelstrasse 9A, 30167 Hannover, Germany
Email: mech@tnt.uni-hannover.de

Ferran Marqués

Universitat Politècnica de Catalunya, Campus Nord—Mòdul D5, C/ Jordi Girona 1-3, Barcelona 08034, Spain
Email: ferran@gps.tsc.upc.es

The objective evaluation of 2D-shape estimation results for moving objects in a video sequence is still an open problem. First approaches in the literature evaluate the spatial accuracy and the temporal coherency of the estimated 2D object shape. Thereby, it is not distinguished between several estimation errors located around the object contour and a few, but larger, estimation errors. Both cases would lead to similar evaluation results, although the 2D-shapes would be visually very different. To overcome this problem, in this paper, a new evaluation approach is proposed. In it, the evaluation of the spatial accuracy and the temporal coherency is based on the mean and the standard deviation of the 2D-shape estimation errors.

Keywords and phrases: shape evaluation, objective evaluation, shape estimation, segmentation, video object, MPEG.

1. INTRODUCTION

One major problem in the development of algorithms for 2D-shape estimation of moving objects, is to assess the quality of the estimation results. Up to now, mainly subjective evaluation, that is, tape viewing, has been used in order to decide upon the quality of a certain algorithm. Although this is very helpful and gives already some indication of the resulting quality, this procedure very much depends on the subjective conditions, that is, the attending people, the time of viewing, the used video equipment, and so forth. In the sequel, since we are only dealing with 2D-shape, the term "shape" will be used.

In the literature, first approaches for objective evaluation of shape estimation results can be found [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. During the standardization work of ISO/MPEG-4 [13], within the core-experiment on automatic segmentation of moving objects it became necessary to compare the results of different proposed shape estimators, not only by subjective evaluation, but also by objective evaluation. The proposal for objective evaluation [9], which was agreed by the working group, uses an a priori known shape to evaluate the estimation result. This shape is denoted as *reference* shape, and has to be created once in an appropriate way, for example, by manual segmentation of each frame, by color-keying, or using synthetic image sequences, where

shapes are known. The shape of a moving object can be represented by a binary mask, where a pel has *object* label if it is inside the object and *background* label if it is outside the object. In [9], such a mask is called *object mask*. There are two objective evaluation criteria defined:

(i) the first criterion evaluates the spatial accuracy of an estimated shape. The algorithm obtains the amount of pels that have different labels in the estimated and the reference object masks. Then, this value is normalized by the size of the object in the reference object mask;

(ii) the most subjectively disturbing effect is the temporal incoherence of an estimated sequence of object masks. This is evaluated by the second criterion. The number of pels with opposite label between two successive frames is calculated for the reference and the estimated sequence of object masks. For each frame, the difference of these two values is computed and normalized by the size of the object. A large resulting value hints to a large difference in activity between the reference and the estimated shapes.

Beside the ISO/MPEG-4 core-experiment, this objective evaluation approach was used by the European projects COST 211 [14] and ACTS/MoMuSys [15]. However, the approach has the following shortcomings:

(1) the criterion for spatial accuracy does not distinguish among several small deviations between the estimated and the reference masks (case 1) and a few, but larger, deviations (case 2). Both cases can lead to the same value for spatial accuracy, although they are visually very different;

(2) the same problem appears for the temporal coherency criterion, where several areas of small contour activity and a few ones of larger activity may lead to similar results;

(3) the temporal coherency evaluation may lead to a second type of problems in the case of camera or object motion. Then, changes in the object mask between two consecutive frames can be either caused by movement or by contour activity, which is not distinguished by the criterion.

Within the project COST 211 the above approach has been further developed [6, 8]:

- for evaluation of the spatial accuracy, it is distinguished between pels that have object-label in the estimated object mask, but not in the reference object mask, and vice versa; that is, if the estimated shape is too large or too small. Furthermore, the impact of a misclassified pel on the criterion for spatial accuracy depends on its distance to the object contour. By these improvements, the evaluation of shape estimation results can be adapted to specific applications;

- for evaluating the temporal coherency, two criteria are used. The first one analyzes local instabilities by comparing the variation of the spatial accuracy criterion between successive frames. The second one assumes that the shape is correctly estimated, but oscillates around the reference shape. For this case, the distance between the gravity center of the object in the estimated and in the reference object masks is analyzed for succeeding frames.

The use of the variation of the spatial accuracy criterion for evaluating the temporal coherency allows solving the third problem. However, these criteria do not solve the first and second ones and neither does the approach in [3]. There, additional geometric features such as the size and the position of an object as well as the average color within an object area are evaluated based on the estimated and the reference object masks.

In this paper, a simple approach [16] for objective evaluation of results from a 2D-shape estimation is proposed, which tackles the three mentioned problems. As in previous approaches, the spatial accuracy and the temporal coherency of an estimated shape are evaluated by comparing it with the corresponding reference shape. It is assumed that the reference shape does not contain any holes. In the case that the reference object consists of several components, each component is evaluated separately. The estimation error is defined as the spatial distance between the reference and the estimated shapes. In order to measure the distance, shapes are not represented as binary object masks, but as object con-

tours. An object contour is the set of pels that have *object* label in the corresponding object mask, and at least one of the four neighboured pels has *background* label. The evaluation approach is mainly based on calculating the mean and the standard deviation of the shape estimation errors.

The paper is organized as follows: in Section 2, the proposed evaluation method is described. The criteria for spatial accuracy and temporal coherency are explained. After that, it is discussed how these criteria can be used to evaluate shape estimation results with respect to a given application. In Section 3, results of the proposed evaluation method are presented, and it is demonstrated that, in addition to the third problem also the first two problems are solved. Section 4 summarizes the paper and gives conclusions.

2. OBJECTIVE EVALUATION CRITERIA

2.1. Spatial accuracy

The spatial accuracy of an estimated shape of a moving object can be defined by the spatial distance between the reference shape and the estimated one. In this paper, this distance is determined based on a given set of N_m measure points on the reference object contour. This means that for each measure point i its distance d_i to the estimated object contour is measured. Here, the Euclidean distance is used. For the measured distance values the mean and the standard deviation are calculated, which are then normalized by the maximal expansion $\emptyset_{\max}$ of the object, resulting in the normalized mean m_d and normalized standard deviation σ_d:

$$
\begin{aligned}
m_d &= \frac{1}{\emptyset_{\max}} \cdot \frac{1}{N_m} \sum_{i=1}^{N_m} d_i, \\
\sigma_d &= \frac{1}{\emptyset_{\max}} \cdot \sqrt{\frac{1}{N_m - 1} \sum_{j=1}^{N_m} \left[d_j - \frac{1}{N_m} \sum_{i=1}^{N_m} d_i \right]^2}.
\end{aligned}
\tag{1}
$$

The maximal expansion of an object $\emptyset_{\max}$ is defined as the length of the longest straight line segment between two pels of the reference object contour. Due to the normalization by $\emptyset_{\max}$, the mean and the standard deviation become independent from the object size. While the normalized mean m_d is a measure for the average distance between the reference and the estimated object contour, the normalized standard deviation σ_d represents how different the measured distances d_i are.

The algorithm for measuring the distance values d_i between the two object contours consists of two steps, which are shown in Figure 1: in the first step, the reference and the estimated object contours are split into parts that are assigned to each other. This is done by determining for each pel on the reference object contour the straight line, which is perpendicular to the tangent line for that pel. The tangent line is estimated based on the two neighbour contour pels on each side of the considered pel. The intersection point between the perpendicular line and the estimated object contour defines the corresponding pel on the estimated object contour. The corresponding pels associated to two succeeding pels on
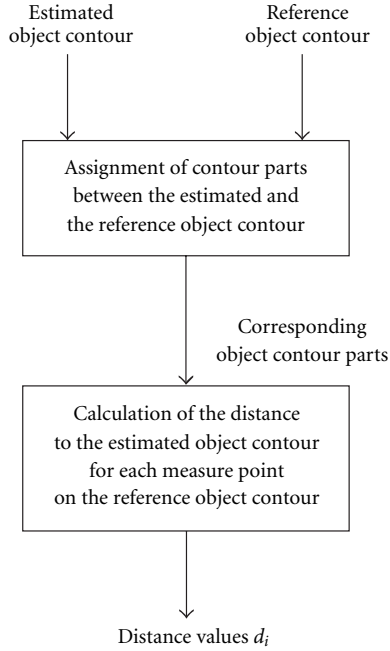
Estimated
object contour

Reference
object contour

Assignment of contour parts
between the estimated and
the reference object contour

Corresponding
object contour parts

Calculation of the distance
to the estimated object contour
for each measure point
on the reference object contour

Distance values d_i

Figure 1: Block diagram of the algorithm for measuring the distance values d_i between the reference and the estimated object contour.

the reference object contour define the corresponding segment in the estimated object contour (see zoomed area of Figure 2).

In the special cases that the reference object contour is intersected first, or the intersection point belongs to an already assigned part of the estimated object contour, the assignment is invalid (dotted arrows in Figure 2), and therefore the next pel on the reference object contour is processed. This is continued until the reference object contour is not intersected as first and the resulting intersection point is not yet assigned. The segment of the estimated object contour which is surrounded by this intersection point and the previous intersection point for which the assignment was valid (solid arrows in Figure 2) is assigned to the segment of the reference object contour, which is surrounded by the latest processed pel and the preceding pel for which the assignment was valid.

In the second step in Figure 1, the distance between corresponding measure points on the reference and on the estimated object contours is calculated. A measure point is defined as the point on the reference or estimated object contour in the middle of two succeeding contour pels. For each measure point on the reference object contour (rhombs in Figure 3), the average distance to all measure points within the corresponding part of the estimated object contour (circles in Figure 3) is calculated. In the example, which is shown in the zoomed area of Figure 3, there are two measure points on the estimated object contour assigned to measure point 12 on the reference object contour and, therefore, two distances are calculated. These two distances are averaged, resulting in the distance value d_{12} for the investigated measure point 12 on the reference object contour. If there is more than one measure point in the same part of the reference object con-

tour, as in the case of measure points 14 to 29 in Figure 3, the calculation is done for each of them, separately.

2.2. Temporal coherency

The temporal coherency of an estimated shape sequence is evaluated by the temporal variation of the two criteria for spatial accuracy between succeeding frames:

$$\Delta m_{d,t} = \left| m_{d,t} - m_{d,t-1} \right|,$$
$$\Delta \sigma_{d,t} = \left| \sigma_{d,t} - \sigma_{d,t-1} \right|, \tag{2}$$

where $m_{d,t}$ is the normalized mean m_d and $\sigma_{d,t}$ is the normalized standard deviation σ_d for the frame at time instance t. If the normalized mean m_d and the normalized standard deviation σ_d of the distance values d_i between the reference and the estimated object contour are similar for succeeding frames, their respective temporal variation $\Delta m_{d,t}$ and $\Delta \sigma_{d,t}$ are small. In this case the temporal coherency is judged as good.

However, these two parameters do not analyze whether the measured distances keep the same value in succeeding frames, while changing their spatial position. In order to detect such cases, a third criterion for evaluation of the temporal coherency is used, which is proposed in [8] (see Figure 4):

$$\Delta g_t = \left| \frac{1}{\emptyset_t} \left(\mathbf{g}_t^{\text{ref}} - \mathbf{g}_t^{\text{est}} \right) - \frac{1}{\emptyset_{t-1}} \left(\mathbf{g}_{t-1}^{\text{ref}} - \mathbf{g}_{t-1}^{\text{est}} \right) \right|. \tag{3}$$

The vectors $\mathbf{g}_t^{\text{ref}}$ and $\mathbf{g}_t^{\text{est}}$ are the gravity centers of the evaluated object in the reference and the estimated object mask at time instance t, respectively. Δg_t is the amount of variation from time instance $t - 1$ to t of the difference between the gravity centers in the reference and estimated object mask normalized by the maximal object expansion $\emptyset_{\text{max}}$. For this third criterion, it is assumed that changes on the position of the estimation errors are not symmetrically distributed with respect to the gravity center.

2.3. Interpretation of results of the objective evaluation criteria

In Sections 2.1 and 2.2, criteria for evaluating the spatial accuracy and the temporal coherency of an estimated object shape are proposed. Furthermore, it is described how these criteria are measured. In this subsection it is discussed how these criteria can be interpreted.

For evaluation of the spatial accuracy of an estimated shape two criteria are used, the normalized mean m_d and the normalized standard deviation σ_d of the measured distances d_i. For a specific application both criteria should be lower than given thresholds in order to meet the demanded accuracy. For example, one class of applications, which would contain MPEG-4 [13] and MPEG-7 [17] content generation tools, demands a high spatial accuracy, which means low values for m_d and σ_d. Another class of applications could allow a few shape errors, but overall the shape should be well estimated. This class, which could include tools for scene interpretation, demands mainly a small mean value m_d. A third class of applications could allow larger shape errors, but the errors should be of constant amplitude, which is an
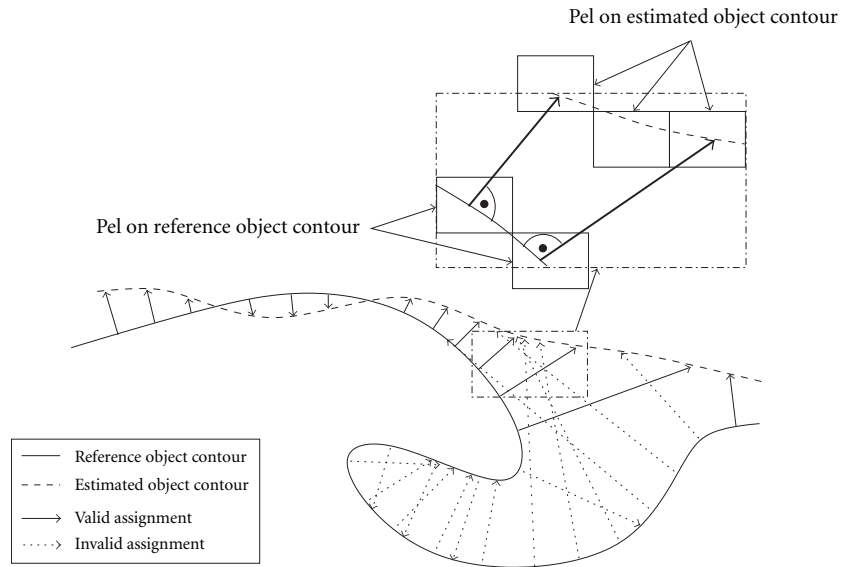
FIGURE 2: Example for the assignment of contour parts between the estimated and the reference object contour.
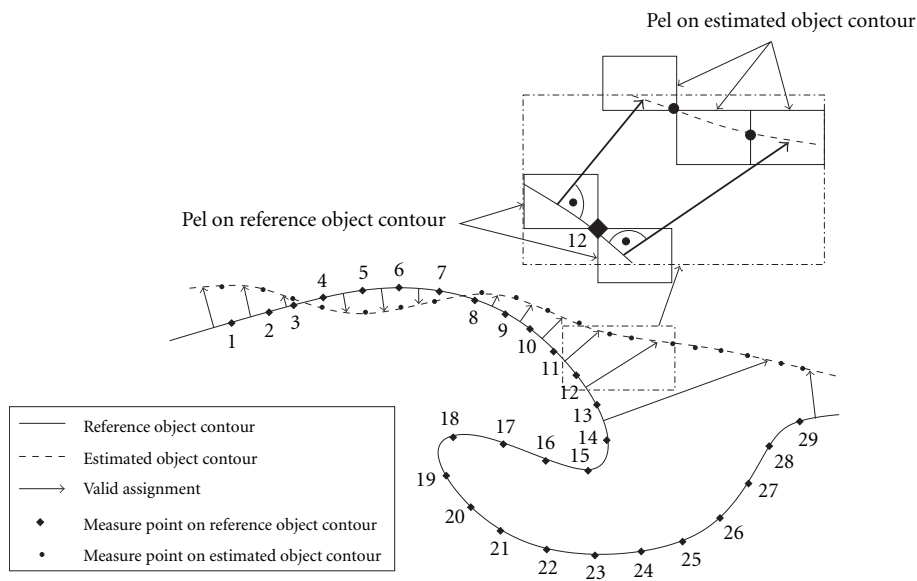

FIGURE 3: Example for the calculation of the distance to the estimated object contour for each measure point on the reference object contour.

advantage, if the shape has to be coded. For this class, m_d can be larger, but σ_d must be small.

Additionally, for the case that a human observer should not be disturbed by the spatial inaccuracy of an estimated shape, thresholds for m_d and σ_d can be found. This means that it is possible to represent the impression that a human observer gets from a shape estimation result by the two proposed criteria for spatial accuracy, opening the door to replacing subjective evaluation by objective criteria.

In an analogous way, the above statements are valid for the temporal evaluation criteria. For a given application, thresholds for $\Delta m_{d,t}$, $\Delta \sigma_{d,t}$, and Δg_t have to be fixed. Then,

it can be decided if the temporal behavior of the shape estimation errors is good enough for a specific application.

3. EXPERIMENTAL RESULTS

The proposed evaluation method has been applied to shape estimation results for several test sequences. Thereby, a good correspondence with the visual impression of the results was established.

With the results in Figure 5, it is shown that the first two problems of previous approaches are solved by the proposed evaluation method. Figure 5a shows the reference frame of
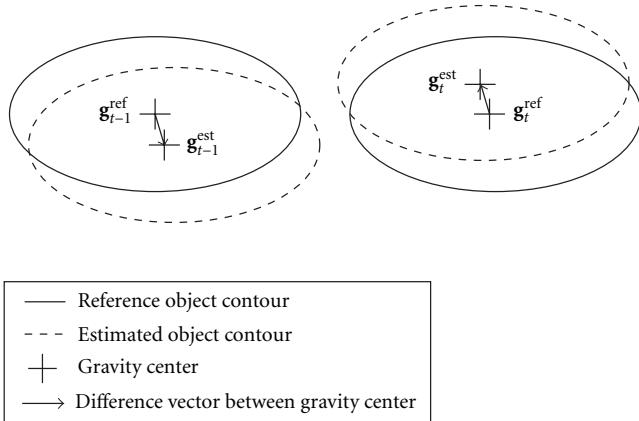
FIGURE 4: Evaluation of the temporal coherency by investigating the temporal variation of the gravity center difference $(\mathbf{g}_t^{\text{ref}} - \mathbf{g}_t^{\text{est}})$ between succeeding frames at time instances $t - 1$ and t.
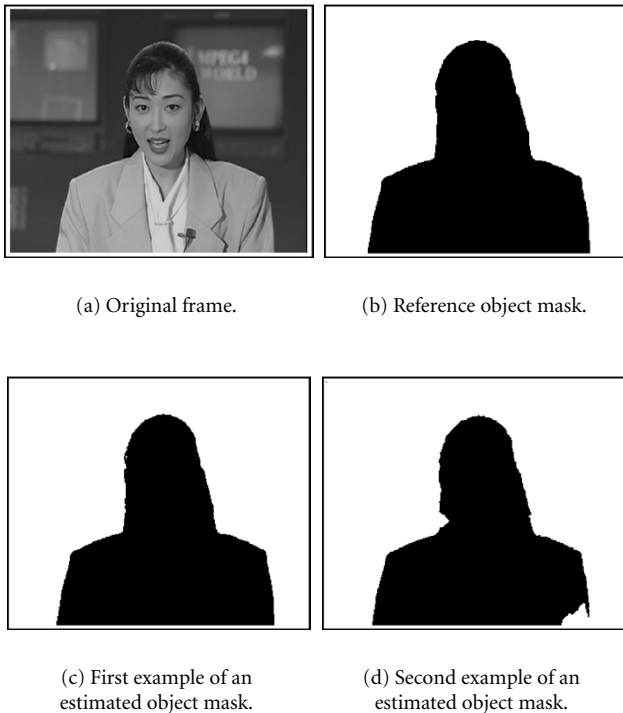

(a) Original frame.

(b) Reference object mask.

(c) First example of an estimated object mask.

(d) Second example of an estimated object mask.

FIGURE 5: Examples for 2D-shape estimation results for frame 30 of the MPEG-4 test sequence *Akiyo*.

the MPEG-4 test sequence *Akiyo*. The corresponding reference shape represented by an object mask is shown in Figure 5b. Figures 5c and 5d present two examples for shape estimation results. The first one (Figure 5c) is the reference object mask after dilation [18], which therefore has various small estimation errors around the object contour. This corresponds to case 1 in the introduction. In the second one (Figure 5d), which corresponds to case 2 in the introduction, a part of the left arm and a part of the hair of the person are missing. Thus, there are large estimation errors mainly

TABLE 1

Estimation result	$m_d\,[\%]$	$\sigma_d\,[\%]$
Object mask in Figure 5c	0.856	0.437
Object mask in Figure 5d	0.824	1.485

TABLE 2

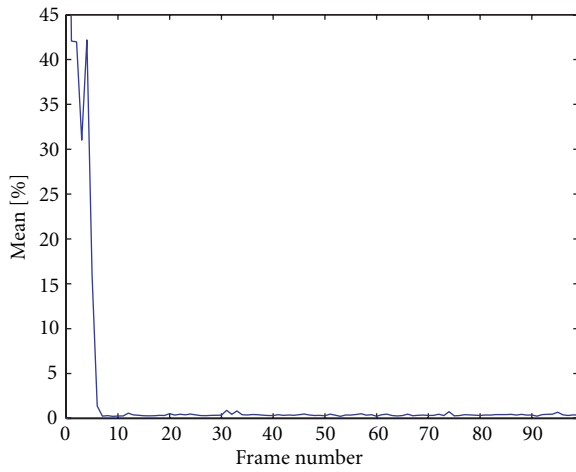| Estimation result	$\Delta m_{d,30}\,[\%]$	$\Delta \sigma_{d,30}\,[\%]$
Object mask in Figure 5c	0.856	0.437
Object mask in Figure 5d	0.824	1.485

at two positions of the object contour. Although both shapes look very different, they would give similar values for the spatial accuracy, if evaluated by an approach from the literature, for example, [6]. Using the evaluation method proposed in this paper, the two criteria for evaluating the spatial accuracy have the values (given as percentage) in Table 1.

The normalized mean m_d of the estimation errors of both shapes is nearly equal. However, their normalized standard deviation σ_d is quite different. Therefore, the spatial accuracy of both results is judged different if using the proposed evaluation method.
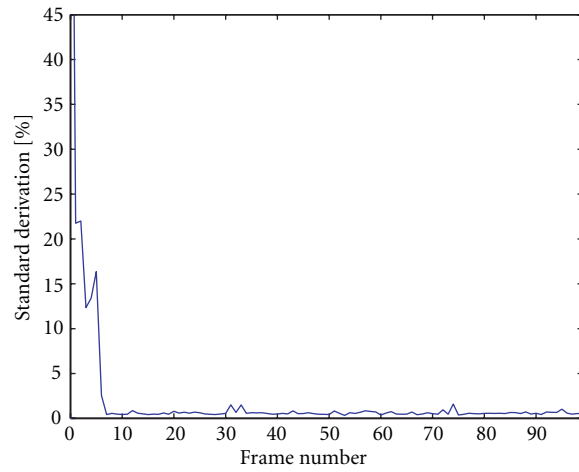
Assuming that the two estimation results in Figure 5 have been perfect for the preceding frame 29, both, m_d and σ_d would have been zero. Then, the temporal coherency criteria for frame 30, $\Delta m_{d,30}$ and $\Delta \sigma_{d,30}$, would be as in Table 2.

The temporal variation of the normalized mean $\Delta m_{d,30}$ is nearly the same for both estimation results, because in the case of the mask in Figure 5c there is small temporal shape activity around the whole object, while in case of the mask in Figure 5d the temporal shape activity is much higher, but mainly at two positions of the object. However, caused by this, the temporal variation of the normalized standard deviation $\Delta \sigma_{d,30}$ is small for the mask in Figure 5c and much larger for the mask in Figure 5d. This shows that the case of several areas of small contour activity can be distinguished from the case of only a few areas, but of larger activity. Therefore, also the second problem from the introduction is solved by the proposed evaluation method.
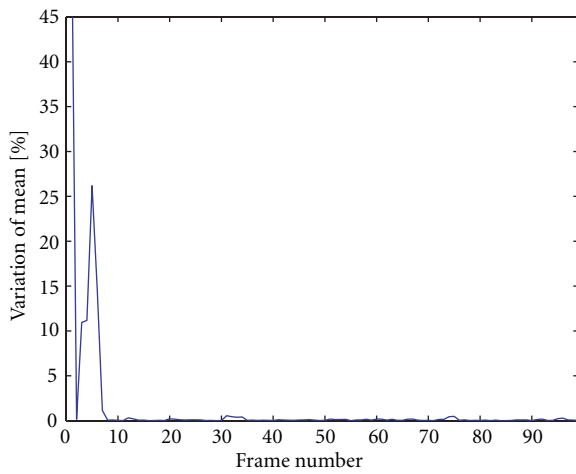
In Figures 6 and 7, the results for all criteria of the proposed evaluation method are shown to estimate the results of the MPEG-4 test sequences *Akiyo* and *Hall-monitor* generated by the COST 211 Analysis Model (Version 5.1) [14, 19]. In the results for *Akiyo*, it is visible that the two spatial criteria (Figures 6a and 6b) and the three temporal criteria (Figures 6c, 6d, and 6e) have quite large values for the first seven frames. In these first frames, the estimated shape tends to complete Akiyo's silhouette. Therefore, the shape is not correctly estimated, and it changes rapidly between these frames. For all following frames, Akiyo's shape is correctly estimated, which results in low values for the spatial criteria and also for the temporal criteria. At frames 31, 33, and 95, the estimated shape presents small estimation errors in the head area and so does at frame 74 at the right arm. Such errors explain the small peaks in Figure 6. Only the criterion for variation of
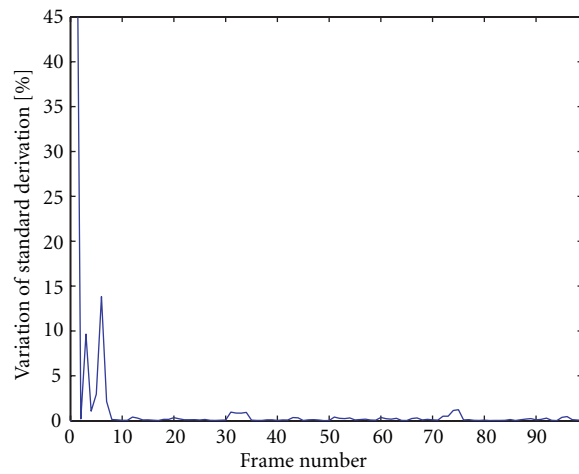
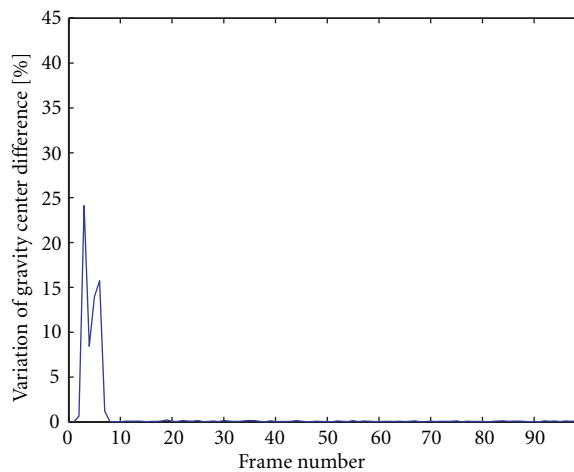(a) Normalized mean of distances.

(b) Normalized standard deviation of distances.

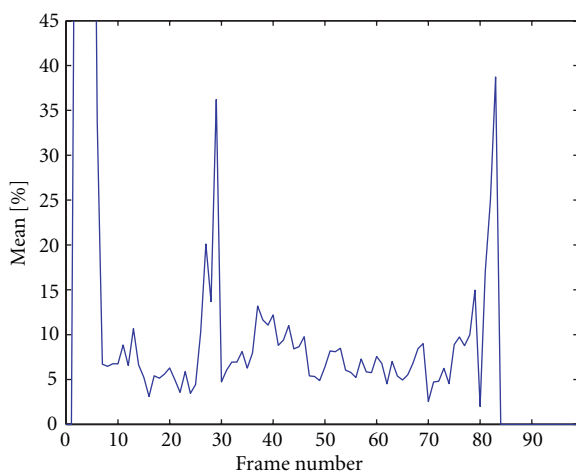(c) Variation of normalized mean of distances.

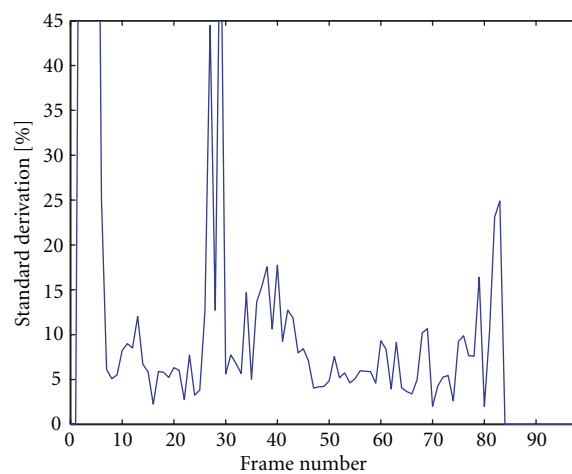(d) Variation of normalized standard deviation of distances.

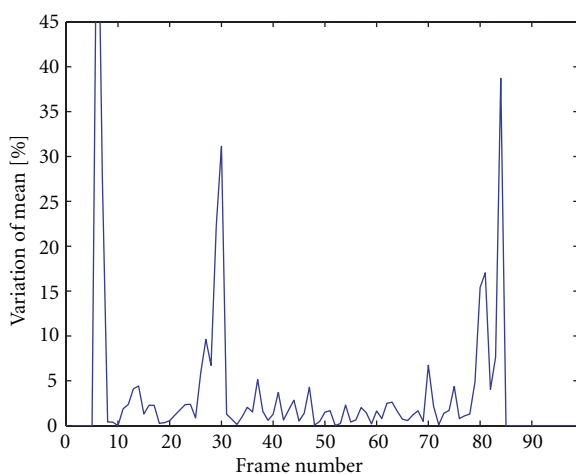(e) Variation of normalized gravity center difference.

FIGURE 6: Evaluation of 2D-shape estimation results for the MPEG-4 test sequence *Akiyo* (10 Hz) generated by the COST 211 Analysis Model (Version 5.1) using the proposed evaluation method.
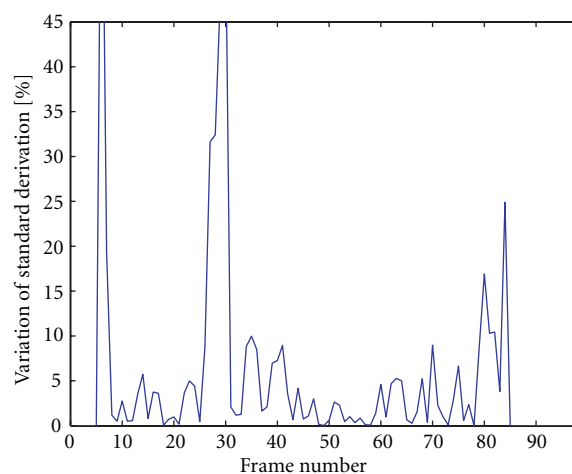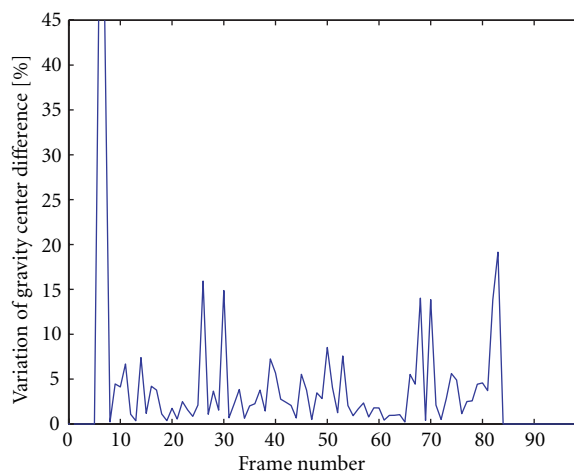
(a) Normalized mean of distances.

(b) Normalized standard deviation of distances.

(c) Variation of normalized mean of distances.

(d) Variation of normalized standard deviation of distances.

(e) Variation of normalized gravity center difference.

FIGURE 7: Evaluation of 2D-shape estimation results for the MPEG-4 test sequence *Hall-monitor* (10 Hz) generated by the COST 211 Analysis Model (Version 5.1) using the proposed evaluation method.

(a) Original frame 29.

(b) Estimated object mask for frame 29.

(c) Original frame 84.
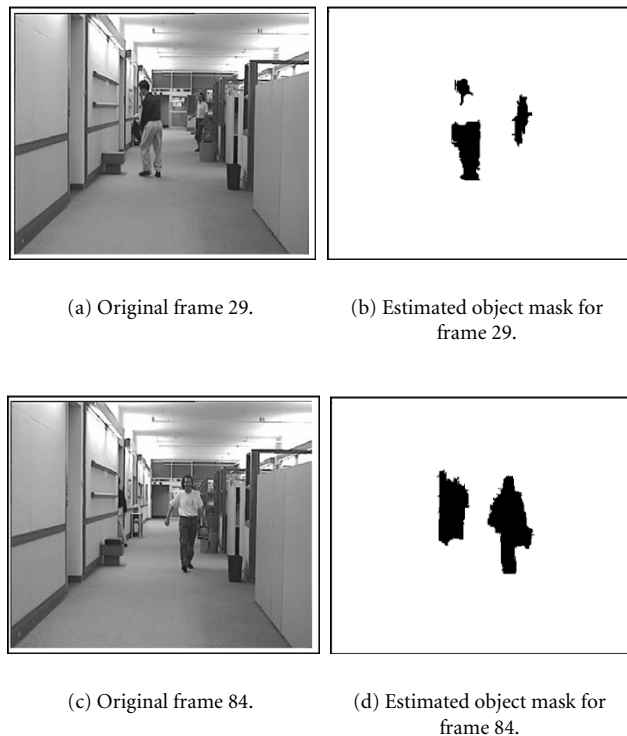

(d) Estimated object mask for frame 84.

FIGURE 8: 2D-shape estimation results of the COST 211 Analysis Model (Version 5.1) for the MPEG-4 test sequence *Hall-monitor* (10 Hz).

the difference between the gravity centers (Figure 6e) is not much affected by these estimation errors, because they are quite small.

Figure 7 shows the evaluation results for the test sequence *Hall-monitor*. Here, only the shape of the person on the left side of the image is evaluated. This person becomes visible in the second frame, but appears in the estimation result of frame 6 for the first time. Therefore, the two spatial criteria are zero for frame 0 and very large for frames 1 to 5 (Figures 7a and 7b). In the following frames the mean of the estimation errors lies between 2 and 15% of the object expansion. There are only two exceptions: the first one is between frames 26 and 29, where half of the body of the person is missing (see Figures 8a and 8b). The second one is between frames 80 and 84, where the person leaves the scene so that he is not visible after frame 84 (see Figures 8c and 8d). Because of the memory usage in the COST 211 Analysis Model and some shadow effects in the scene, the disappearance of the person is not detected. This results in a growing estimation error, which is visible in Figure 7a. Figure 7b presents the normalized standard deviation of the distances. Its value is large especially between frames 26 and 29, where half of the body is missing. This is reasonable, because in the missing part of the body the estimation errors are much larger than in the other part. Thus, the estimation errors are quite different. Of course for these frames also the temporal variation of the gravity center difference is large, as it can be seen in Figure 7e.

4. CONCLUSIONS

In this paper a method for objective evaluation of 2D-shape estimation results is proposed. The estimation error of an estimated object shape is defined as the distance between the reference and the estimated object contour, which is measured for several points of the reference object contour. For evaluating the spatial accuracy, the mean and the standard deviation of the measured distances are calculated.

It is shown that the normalized mean of the measured deviations between the estimated and the corresponding reference shape is a useful criterion to evaluate the spatial accuracy. Furthermore, by the normalized standard deviation it can be distinguished if an estimated shape has several small estimation errors or if it has only a few, but larger, estimation errors.

For evaluating the temporal coherency, the temporal variation between succeeding frames of the normalized mean and of the normalized standard deviation is investigated. It is shown that by these two criteria it can be assessed if there are various small contour activity areas around the object contour between succeeding frames, or if there is a higher contour activity, but only at a few positions of the object contour.

A third criterion is applied to detect changes of the spatial position of estimation errors. It evaluates the temporal variation of the difference between the gravity centers of the reference and the estimated shape.

The approach has been tested with shape estimation results for several test sequences. Thereby, a good correspondence with the visual impression of the results was established. This have lead to use the evaluation approach within the project COST 211.

Finally, it is explained that the evaluation method can be adapted to a specific application by definition of thresholds for the spatial and temporal criteria. Specifically, thresholds can be found to model a human observer's impression on estimation errors. Furthermore, it is possible to combine the proposed evaluation method with the ideas from [6, 8], where positive and negative distances are distinguished.

REFERENCES

[1] M. Borsotti, P. Campadelli, and R. Schettini, "Quantitative evaluation of color image segmentation results," *Pattern Recognition Lett.*, vol. 19, no. 8, pp. 741–747, 1998.

[2] P. Correia and F. Pereira, "Estimation of video object's relevance," in *European Conference on Signal Processing (EUSIPCO '2000)*, Tampere, Finland, September 2000.

[3] P. Correia and F. Pereira, "Objective evaluation of relative segmentation quality," in *Int. Conference on Image Processing (ICIP)*, pp. 308–311, Vancouver, Canada, September 2000.

[4] C. E. Eroglu and B. Sankur, "Performance evaluation metrics for object-based video segmentation," in *10th European*

Signal Processing Conference (EUSIPCO '2000), pp. 917–920, Tampere, Finland, September 2000.

[5] M. D. Levine and A. M. Nazif, "Dynamic measurement of computer generated image segmentations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 155–165, 1985.

[6] X. Marichal and P. Villegas, "Objective evaluation of segmentation masks in video sequences," in *European Conference on Signal Processing (EUSIPCO '2000)*, vol. 4, pp. 2193–2196, Tampere, Finland, September 2000.

[7] K. McKoen, R. Navarro-Prieto, B. Duc, E. Durucan, F. Ziliani, and T. Ebrahimi, "Evaluation of video segmentation methods for surveillance applications," in *Proc. European Signal Processing Conference 2000*, Tampere, Finland, September 2000.

[8] P. Villegas, X. Marichal, and A. Salcedo, "Objective evaluation of segmentation masks in video sequences," in *Proc. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '99)*, pp. 85–88, Berlin, Germany, May/June 1999.

[9] M. Wollborn and R. Mech, "Refined procedure for objective evaluation of VOP generation algorithms," Doc. ISO/IEC JTC1/SC29/WG11 MPEG98/3448, Fribourg, Switzerland, October 1997.

[10] W. A. Yasnoff, J. K. Mui, and J. W. Bacus, "Error measures for scene segmentation," *Pattern Recognition*, vol. 9, no. 4, pp. 217–231, 1977.

[11] Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recognition*, vol. 29, no. 8, pp. 1335–1346, 1996.

[12] Y. J. Zhang, "Evaluation and comparison of different segmentation algorithms," *Pattern Recognition Lett.*, vol. 18, no. 10, pp. 963–974, 1997.

[13] MPEG-4: Doc. ISO/IEC JTC1/SC29/WG11 N2502, "Information Technology—Generic Coding of Audiovisual Objects, Part 2: Visual, Final Draft of International Standard," October 1998.

[14] M. Gabbouj, G. Morrison, F. Alaya-Cheikh, and R. Mech, "Redundancy reduction techniques and content analysis for multimedia services—The European COST 211quat action," in *Proc. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '99)*, pp. 69–72, Berlin, Germany, 31 May–11 June 1999.

[15] B. Marcotegui, P. Correia, F. Marques, et al., "A video object generation tool allowing friendly user interaction," in *International Conference on Image Processing (ICIP '99)*, Kobe, Japan, October 1999.

[16] R. Mech and F. Marques, "Objective evaluation criteria for 2D-shape estimation results of moving objects," in *Proc. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '01)*, Tampere, Finland, May 2001.

[17] MPEG-7: Doc. ISO/IEC JTC1/SC29/WG11 N2822, "Visual Part of Experimentation Model Version 2.0," Vancouver, Canada, July 1999.

[18] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, Chapman & Hall Computing, London, UK, 1993.

[19] A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, and T. Sikora, "Image sequence analysis for emerging interactive multimedia services—The European COST 211 framework," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 8, no. 7, pp. 802–813, 1998.

Roland Mech received the Diplom-Informatiker degree in computer science from the University of Dortmund, Dortmund, Germany, in 1995. Since 1995 he is with the "Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung" at the University of Hannover, Germany, where he works in the areas of image sequence analysis and image sequence coding. He is a member of the European project COST 211. Furthermore, he was a member of the finished European project ACTS-MoMuSys and contributed actively to the ISO/MPEG-4 standardization activities. His present research interests cover image sequence analysis, especially 2D-shape estimation of moving objects, and the application of object-based image sequence coding.

Ferran Marqués received the Electrical Engineering degree from the Polytechnic University of Catalunya (UPC), Barcelona, Spain, in 1988. From 1989 to June 1990, he worked in the Swiss Federal Institute of Technology in Lausanne (EPFL) in the group of "Digital image sequence processing and coding." In June 1990, he joined the Department of Signal Theory and Communications of the Polytechnic University of Catalunya (UPC). From June 1991 to September 1991, he was with the Signal and Image Processing Institute at USC in Los Angeles, California. He received the Ph.D. degree from the UPC in December 1992 and the Spanish Best Ph.D. thesis on Electrical Engineering Award-1992. Since 1995, he is Associate Professor at UPC, having served as Associate Dean for International Relations of the Telecommunication School (ETSETB) at UPC (1997–2000). He is lecturing on the area of digital signal and image processing. His current research interests include still image and sequence analysis, still image and sequence segmentation, image sequence coding, motion estimation and compensation, mathematical morphology and biomedical applications. In the area of image coding and representation, he has been a very active partner in the MPEG-4 standard process, mainly through the European project MoMuSys. In MoMuSys, he acted as Work Package Leader of the Video Algorithms work package that, among other tasks, implemented the MPEG4 VM. He has served as Officer responsible for the Membership Development (1994–1998) for the EURASIP AdCom, as elected member responsible of the Member Services (1998–2000) and currently he is serving as Secretary and Treasurer. He was Associate Editor of the Journal of Electronic Imaging (SPIE) in the area of Image Communications (1996–2000) and he serves in the Editorial Board of the EURASIP Journal on Applied Signal Processing. He is author or co-author of more than 50 publications that have appeared as journal papers and proceeding articles, 4 book chapters, and 4 international patents.