EURASIP Journal on Applied Signal Processing 2002:4, 390–402

Stand-Alone Objective Segmentation Quality Evaluation

Paulo Lobato Correia
Instituto Superior Técnico, Instituto de Telecomuniçaões, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
Email: paulo.correia@lx.it.pt

Fernando Pereira
Instituto Superior Técnico, Instituto de Telecomuniçaões, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
Email: fernando.pereira@lx.it.pt

Received 31 July 2001 and in revised form 4 January 2002

The identification of objects in video sequences, that is, video segmentation, plays a major role in emerging interactive multimedia services, such as those enabled by the ISO MPEG-4 and MPEG-7 standards. In this context, assessing the adequacy of the identified objects to the application targets, that is, evaluating the segmentation quality, assumes a crucial importance. Video segmentation technology has received considerable attention in the literature, with algorithms being proposed to address various types of applications. However, the segmentation quality performance evaluation of those algorithms is often ad hoc, and a well-established solution is not available. In fact, the field of objective segmentation quality evaluation is still maturing; recently, some more efforts have been made, mainly following the emergence of the MPEG object-based coding and description standards. This paper discusses the problem of objective segmentation quality evaluation in its most difficult scenario: stand-alone evaluation, that is, when a reference segmentation is not available for comparative evaluation. In particular, objective metrics are proposed for the evaluation of stand-alone segmentation quality for both individual objects and overall segmentation partitions.

Keywords and phrases: video segmentation, segmentation quality evaluation, objective segmentation quality, stand-alone segmentation quality evaluation.

1. INTRODUCTION

With the publication of the MPEG-4 standard in the Spring of 1999 [1], which allows to independently encode audiovisual objects, and the development of the MPEG-7 standard [2], allowing the content-based description of audiovisual material, the MPEG committee has given a significant contribution for the development of a new generation of interactive multimedia services. Innovative types of interaction are often based on the understanding of a video scene as composed by a set of video objects, to which it is possible to associate specific information as well as interactive "hooks" to deploy the desired application behaviour.

To enable such type of interactive services, an understanding of the scene semantics is required, notably in terms of the relevant objects that are present. It is in this context that video segmentation plays a determinant role. Segmentation may be automatically obtained at the video production stage, for example, by using chroma keying techniques, or it may have to be obtained from the images captured by a camera by using appropriate segmentation algorithms.

The evaluation of the adequacy of a segmentation algorithm, and its parameters' configuration, for a given application may be crucial to guarantee that the application interactive requirements can be fulfilled.

The current practice for segmentation quality evaluation mainly consists in subjective ad hoc assessment by a representative group of human viewers. This is a time-consuming and expensive process, whose subjectivity can be minimised by following strict evaluation conditions, with the video quality evaluation recommendations developed by ITU providing valuable guidelines [3, 4].

Subjective segmentation quality evaluation differs depending on the availability, or not, of a reference segmentation (often called the "ground truth" segmentation) to compare against the results of the segmentation algorithm under study. For both cases, the evaluation proceeds by analysing the segmentation quality of one object after another, with the human evaluators integrating the partial results and, finally, deciding on an overall segmentation quality score. It is worth noting that these current practice evaluation methodologies have not been formally presented, but they are regularly used in fora such as the COST 211quat European project [5]; some details on this evaluation procedure are available in [6].

Alternatively, objective segmentation quality evaluation methodologies can be used. Unfortunately, the amount of attention devoted to this issue in the past is not comparable to the investment made on the development of the segmentation algorithms themselves [7, 8, 9]. Some proposals for the objective evaluation of segmentation quality have been made

since the 1970s, mainly regarding the assessment of the performance of edge detectors—see reviews in [9, 10, 11]. More recently, the emergence of the MPEG-4 and MPEG-7 standards has given a new impulse, not only to the development of video segmentation technology, but also to the segmentation quality evaluation methodologies themselves—see for instance [12, 13]. However, the metrics available for segmentation quality evaluation typically perform well only for very constrained applications scenarios.

This paper discusses the objective evaluation of segmentation quality, in particular when no "ground truth" segmentation is available to use as a reference for comparison, this means, the so-called *stand-alone objective segmentation quality evaluation*.

The various types of stand-alone objective segmentation quality evaluation are discussed in Section 2. Metrics for individual object and overall segmentation quality evaluation are proposed in Sections 3 and 4, respectively. Results are presented in Section 5 and conclusions in Section 6.

2. TYPES OF STAND-ALONE SEGMENTATION QUALITY EVALUATION

Stand-alone segmentation quality evaluation is performed when no reference video segmentation is available. Therefore, the a priori information that may be available about the expected video segmentation results has a decisive impact on the type of evaluation procedure to be applied, so that meaningful results can be achieved. In particular, stand-alone evaluation of segmentation quality is not expected to provide as reliable results as the evaluation relative to a reference segmentation. A discussion on the relative evaluation of segmentation quality has been presented by the authors in [14].

When performing segmentation quality evaluation, two types of measurements can be targeted:

- *individual object segmentation quality evaluation:* each of the objects identified by the segmentation algorithm can be independently evaluated in terms of its video segmentation quality;
- *overall segmentation quality evaluation:* the set of objects identified by the segmentation algorithm can also be globally evaluated as the set of elements that compose the video sequence under analysis. Besides the individual object evaluation, it is important to assess if the appropriate objects have been detected. To produce a meaningful overall segmentation quality evaluation metric, also the relevance of each object present in the scene must be taken into account, since segmentation errors in the most important objects are more noticeable to a human viewer.

The need for individual object segmentation quality evaluation is motivated by the fact that each video object may be independently stored in a database, or reused in a different context, depending on the adequacy of its segmentation quality for the new purpose targeted. An overall segmentation evaluation is also of great importance as it determines, for instance, if the segmentation goals for a certain application have been globally met, and thus if a given segmentation algorithm is appropriate for a given type of application.

Objective segmentation quality evaluation uses automatic analysis tools and thus produces objective evaluation measures. The automatic tools operate on segmentation results obtained for a selected set of video sequences; if individual object evaluation is being performed, the object whose segmentation quality is to be assessed has first to be selected.

Both the individual object and the overall segmentation quality measures are typically computed for each time instant, requiring that some temporal processing of the instantaneous results is done to reflect the segmentation quality over the complete sequence or shot. For instance, a temporal mean or median may be computed.

Building on the existing knowledge on segmentation quality evaluation and also on some relevant aspects from the video quality evaluation field, a set of relevant features to be evaluated for performing the objective evaluation of stand-alone segmentation quality, as well as appropriate objective quality metrics for both individual object and overall partition segmentation quality evaluation are proposed in the following.

3. INDIVIDUAL OBJECT SEGMENTATION QUALITY EVALUATION

The stand-alone evaluation of segmentation quality is performed by applying the segmentation algorithms to the selected video sequences and then analysing the segmentation results produced. Since the evaluation is performed without using any reference segmentation for comparison, significant assessment results are only expected for well-constrained segmentation scenarios. These results will mainly provide the means for the ranking of partitions in terms of segmentation quality, that is, the results are expected to be more qualitative than quantitative.

The criteria to be applied in stand-alone objective segmentation quality evaluation may be generic, based on the human visual system (HVS) characteristics, or more adjusted to the specific application scenario targeted by considering the available a priori information. In the first case, all aspects considered important in terms of the HVS are included. Examples are the recognition that some types of shapes usually attract more the human viewer attention or the unequal treatment of the various image components with luminance receiving more attention. Additional assumptions, like a smooth temporal evolution implying limited changes in the object features for consecutive time instants, are usually more dependent on the specific application scenario. These assumptions can be clustered into the following classes: *shape regularity, spatial uniformity, temporal stability, and motion uniformity*, as discussed below.

The stand-alone evaluation of individual objects can rely on spatial and temporal features of the objects themselves (*intra-object homogeneity features*) as well as on the comparison of selected object features with neighbouring objects (*inter-object disparity features*). Intra-object features give an

indication about the internal homogeneity of the objects, while inter-object features indicate if the objects were correctly identified as separate entities.

The desired metrics for stand-alone segmentation quality evaluation can thus be established based on the following types of features.

Intra-object homogeneity features

Intra-object homogeneity regards the internal homogeneity of each object which can be evaluated by means of spatial and temporal object features.

(a) *Spatial features:* the stand-alone evaluation of an object's spatial features can be done by evaluating its shape regularity and spatial uniformity. However, the applicability and importance of shape regularity and spatial uniformity is different depending on the segmentation scenario considered.

Shape regularity: in some cases, the objects are expected to exhibit regular shapes, which can be evaluated by geometrical features such as the circularity, elongation, and compactness of the objects.

Spatial uniformity: in some circumstances, the texture of the object is expected to be reasonably uniform; features such as the spatial perceptual information [4] or texture variance can be used to measure the spatial uniformity.

(b) *Temporal features:* the importance of the temporal features for segmentation quality evaluation also differs depending on the segmentation scenario being considered. Stand-alone evaluation of temporal features relies on the assumption of a smooth temporal evolution or on the uniformity of the motion within the object area.

Temporal stability: assuming that the temporal evolution of the object features is smooth, the variation between consecutive time instants can be checked for evaluating their temporal stability. Significant variations in the temporal stability metrics, in scenarios where they are supposed to be small, indicate the presence of segmentation errors.

Motion uniformity: when objects are supposed to exhibit uniform motion, such properties as the variance of the object's motion vector values or the criticality [15] can provide valuable segmentation evaluation metrics since they are able to signal higher or lower segmentation qualities.

Inter-object disparity features

The comparison of an object's features against those of its neighbours can provide useful information for stand-alone evaluation: it is assumed that additional objects should be identified when they are sufficiently different from their neighbours. This comparison can be done locally, along the object boundaries, or it can be based on features computed for the entire objects.

(a) *Local contrast to neighbours:* one of the assumptions that holds in many circumstances is that there should be a significant contrast along the border between the inside and outside of an object. This can be evaluated by a local contrast metric.

(b) *Neighbouring objects features difference:* several features computed for the object area can be compared with the corresponding feature values for the neighbours, to check if they were correctly identified as separate entities. Examples are the shape regularity, spatial uniformity, temporal stability, or motion uniformity, whenever each of them is relevant for the target application.

Relevant metrics for each of these classes of features are presented below, followed by the proposal of composite metrics for two classes of content with different properties.

3.1. Elementary metrics for individual object evaluation

Metrics for individual object stand-alone segmentation quality evaluation can be established corresponding to the classes of features identified above.

In particular, *intra-object* homogeneity can be evaluated by means of spatial and temporal object features. The *spatial features* considered for individual object evaluation, and corresponding metrics, are as follows.

Shape regularity: regularity of shapes can be evaluated by geometrical metrics such as the compactness (compact), or a combination of the circularity and elongation (circ_elong) of the objects

$$\begin{aligned} \text{compact}(E) &= \max\left(\frac{\text{perimeter}^2(E)}{75\cdot\text{area}(E)}, 1\right),\\ \text{circ\_elong}(E) &= \max\left(\text{circ}(E), \max\left(\frac{\text{elong}(E)}{5}, 1\right)\right). \end{aligned} \tag{1}$$

With circularity and elongation defined by

$$\begin{aligned} \text{circ}(E) &= \frac{4\cdot\pi\cdot\text{area}(E)}{\text{perimeter}^2(E)},\\ \text{elong}(E) &= \frac{\text{area}(E)}{\left(2\cdot\text{thickness}(E)\right)^2}, \end{aligned} \tag{2}$$

where thickness(E) is the number of morphological erosion steps that can be applied to the object until it disappears [16]. The normalizing constants were empirically determined after an exhaustive set of tests.

Spatial uniformity: spatial uniformity can be evaluated by metrics such as spatial perceptual information (SI) [4] and texture variance (text_var)—see for instance [11]

$$\begin{aligned} \text{SI} &= \max_{\text{time}}\left(\text{SI}_{\text{stdev}}(I)\right),\\ \text{text\_var}(E) &= \frac{3\cdot\text{var}\,Y(E) + \text{var}\,U(E) + \text{var}\,V(E)}{5}, \end{aligned} \tag{3}$$

with

$$\text{SI}_{\text{stdev}}(I) = \sqrt{\frac{1}{N}\cdot\sum_i\sum_j\left(\text{Sobel}(I)\right)^2 - \left(\frac{1}{N}\cdot\sum_i\sum_j\left(\text{Sobel}(I)\right)\right)^2}. \tag{4}$$

The Sobel operator is specified, for instance, in Annex A of ITU-T Recommendation P.910 [4], and $\max_{\text{time}}(E)$ is the maximal value of E taken for all the temporal instants

considered. var $Y(E)$, var $U(E)$, and var $V(E)$, are the variances of the Y, U, and V components, respectively.

The metrics corresponding to the *temporal features* considered are as follows:

Temporal stability: a smooth temporal evolution of object features can be tested for checking temporal stability. These features may include: size, position, temporal perceptual information [4], criticality [15], texture variance, circularity, elongation, and compactness. The selected metrics for temporal stability evaluation are

$$\begin{aligned} \text{size}_{\text{diff}} &= \left| \text{area}\left(E_t\right) - \text{area}\left(E_{t-1}\right) \right|, \\ \text{elong}_{\text{diff}} &= \left| \text{elong}\left(E_t\right) - \text{elong}\left(E_{t-1}\right) \right|, \\ \text{crit}_{\text{diff}} &= \left| \text{crit}\left(E_t\right) - \text{crit}\left(E_{t-1}\right) \right|, \end{aligned} \tag{5}$$

with crit(E) being the criticality value as defined in [15]

$$\text{crit} = 4.68 - 0.54 \cdot p_1 - 0.46 \cdot p_2, \tag{6}$$

where

$$\begin{aligned} p_1 &= \log_{10}\left(\text{mean}_{\text{time}}\left(\text{SI}_{\text{rms}}(I) \cdot \text{TI}_{\text{rms}}(I)\right)\right), \\ p_2 &= \log_{10}\left(\text{max}_{\text{time}}\left(\text{abs}\left(\text{SI}_{\text{rms}}\left(I_t\right) - \text{SI}_{\text{rms}}\left(I_{t-1}\right)\right)\right)\right), \\ \text{SI}_{\text{rms}}(I) &= \sqrt{\frac{1}{N} \cdot \sum_i \sum_j \left(\text{Sobel}(I)\right)^2}, \\ \text{TI}_{\text{rms}}\left(I_t\right) &= \sqrt{\frac{1}{N} \cdot \sum_i \sum_j \left(I_t - I_{t-1}\right)^2}. \end{aligned} \tag{7}$$

Motion uniformity: the uniformity of motion can be evaluated by metrics such as the variance of the object's motion vector values (mot_var), or by criticality (crit) as defined above

$$\text{mot\_var}(E) = \text{varXvec}(E) + \text{varYvec}(E), \tag{8}$$

where varXvec(E) and varYvec(E) denote the variances for the x and y components of the motion vector field at a given time instant, respectively.

The above spatial and temporal features are not expected to be homogeneous for every segmented object; the applicability and importance of the corresponding metrics is conditioned by the type of application addressed.

Inter-object disparity: metrics can be computed either locally along the object boundaries, or for the complete object area. Again, these metrics are applicable only in some circumstances, such as when a significant contrast, or some other feature significant value difference between neighbouring objects is expected. The metrics considered are as follows.

Local contrast to neighbours: the following local contrast metric can be used for evaluating if a significant contrast between the inside and outside of an object, along the object border, exists

$$\begin{aligned} \text{contrast} = {} & \frac{1}{4 \cdot 255 \cdot N_b} \\ & \cdot \sum_{i,j} \left(2 \cdot \max\left(DY_{ij}\right) + \max\left(DU_{ij}\right) + \max\left(DV_{ij}\right)\right), \end{aligned} \tag{9}$$

where N_b is the number of border pixels for the object and DY_{ij}, DU_{ij}, and DV_{ij} are the differences between an object's border pixel Y, U, and V components, respectively, and its 4-neighbours.

Neighbouring objects features difference: several features, for which objects are expected to differ from their neighbours, can be tested. Examples are the shape regularity, spatial uniformity, temporal stability, and motion uniformity values, whenever each of them is relevant taking the application characteristics into account. In particular, a metric for the motion uniformity feature is considered of interest:

$$\text{mot\_unif}_{\text{neigh\_diff}} = \frac{1}{N} \cdot \sum_{j \in NS_i} \left| \text{mot\_unif}_j - \text{mot\_unif}_i \right|, \tag{10}$$

where i is the object under analysis, N and NS_i are, respectively, the number and the set of neighbours of object i, and the motion uniformity for each object is computed as

$$\text{mot\_unif}_i = \text{mot\_var}_i + \text{crit}_i. \tag{11}$$

Each of the elementary metrics considered for individual object segmentation quality evaluation is normalized to produce results in the interval $[0, 1]$, with the highest values associated to the best segmentation quality results.

3.2. Composite metrics for individual object stand-alone segmentation quality evaluation

The proposal of composite metrics for individual object stand-alone segmentation quality evaluation depends on the type of application (and thus content) being considered, since the adequate elementary metrics depend on the expected characteristics of the content. Therefore, a single general-purpose composite metric cannot be established. Instead, the approach taken here is to select two major classes of content differing in terms of their spatial and temporal characteristics, and propose different composite metrics for each of them.

The distinction between the two classes of content is mainly associated to the temporal characteristics; this fact is reflected in the names adopted for the two content classes defined.

Content class I: stable content: this class corresponds to content that is temporally stable and has reasonably regular shapes. Additionally, the contrast between objects is expected to be strong.

Content class II: moving content: this class corresponds to content with strongly moving objects, and thus temporal stability is less relevant. Often, the motion of the objects is uniform, and neighbouring objects may be less spatially

contrasted, while motion differences between neighbours are expected to be larger. Regular shapes are still expected, even if this characteristic assumes here a lower importance.

The proposed composite metrics for these two content classes are discussed below. Whenever the video content to analyse does not fit well into one of the two classes above, either the closest one is chosen and the results are interpreted with care, or a new combination of the various elementary metrics has to be selected to develop a more appropriate composite metric.

3.2.1 Composite metric for individual object evaluation of stable content

A composite metric to perform as reliably as possible individual object stand-alone segmentation quality evaluation for content class I is proposed below.

The composite metric includes some classes of elementary metrics and excludes some others, to reflect the fact that for this content class, object motion is expected to be weak and the objects are expected to have reasonably regular shapes. Among the excluded classes of metrics are the spatial uniformity (as defined for the elementary metrics proposed here), since arbitrary spatial patterns may be found in the expected objects, and the motion uniformity, as motion is not very relevant for this content class. Thus, the stand-alone evaluation of segmentation quality for this type of content includes the following classes of elementary metrics.

Shape regularity: the shape regularity class of metrics must be included in the composite metric, since shapes are expected to be reasonably regular. The two relevant elementary metrics, compact and circ_elong, are included in the composite metric with equal weights as they complement well each other.

Temporal stability: content in this class is expected to be stable. Therefore, the size, elongation, and criticality stability metrics are combined to represent this class of metrics, all equally weighted.

Local contrast to neighbours: in most cases, the type of content considered will exhibit a significant contrast between neighbouring objects. Assuming that this is the case, then the local contrast metric should be included in the composite metric.

The weights for each class of metrics within the composite metric have been adjusted according to their strength in capturing visual attention, and their ability to match the human subjective evaluation of the segmented sequences with the objective segmentation quality evaluation values. The final weight values were selected after verifying the above assumptions by testing several combinations of elementary metrics' weights.

The proposed composite metric for individual object stand-alone segmentation quality evaluation for this class of content (SQ_io_std_stable) is given by

$$\mathrm{SQ\_io\_std\_stable} = \frac{1}{N} \cdot \sum_{i=1}^{N} \mathrm{SQ\_io\_std\_stable}_i, \quad (12)$$

where N is the total number of images in the sequence whose segmentation is being evaluated, and the instantaneous values of $\mathrm{SQ\_io\_std\_stable}_i$ are given by

$$\mathrm{SQ\_io\_std\_stable}_i = \mathrm{intra}_i + \mathrm{inter}_i, \quad (13)$$

with

$$\begin{aligned} \mathrm{intra}_i &= 0.30 \cdot \mathrm{shape\_reg}_i + 0.33 \cdot \mathrm{temp\_stab}_i, \\ \mathrm{inter}_i &= 0.37 \cdot \mathrm{contrast}_i, \\ \mathrm{shape\_reg}_i &= 0.5 \cdot \mathrm{circ\_elong}_i + 0.5 \cdot \mathrm{compact}_i, \\ \mathrm{temp\_stab}_i &= 0.33 \cdot \mathrm{size}_{\mathrm{diff}_i} + 0.33 \cdot \mathrm{elong}_{\mathrm{diff}_i} + 0.33 \cdot \mathrm{crit}_{\mathrm{diff}_i}. \end{aligned} \quad (14)$$

3.2.2 Composite metric for individual object evaluation of moving content

A composite metric to perform as reliably as possible individual object stand-alone segmentation quality evaluation for content class II is proposed below.

Again, the composite metric only includes the relevant classes of elementary metrics, to adequately reflect the characteristics of this content class. In this case, the content is not expected to be temporally stable, but the objects should have rather uniform motion, and the neighbouring objects motion differences should be pronounced. The classes of metrics considered for the stand-alone evaluation of this type of content are as follows.

Shape regularity: the object shapes are expected to be regular in most of the applications envisioned, even if, due to the motion, this regularity may sometimes not be completely verified (for instance, a walking person will usually have a less regular shape than a person standing still). The compact and circ_elong elementary metrics are again used for the evaluation of shape regularity, with equal weights.

Motion uniformity: in this content class, objects are expected to exhibit reasonably uniform motion. This can be evaluated using the criticality elementary metric.

Local contrast to neighbours: in many cases, the various objects will exhibit a significant contrast to their neighbours. Contrast is not so important in terms of segmentation quality evaluation as for the case of stable content, but the local contrast metric is yet considered useful.

Neighbouring objects feature difference: neighbouring objects are expected to exhibit different motion characteristics. Therefore, the motion uniformity difference metric is here used for segmentation quality evaluation.

The proposed composite metric for individual object stand-alone segmentation quality evaluation for this class of content (SQ_io_std_moving) is given by

$$\mathrm{SQ\_io\_std\_moving} = \frac{1}{N} \cdot \sum_{i=1}^{N} \mathrm{SQ\_io\_std\_moving}_i, \quad (15)$$

where N is the total number of images in the sequence whose segmentation is being evaluated, and the instantaneous values of $\mathrm{SQ\_io\_std\_moving}_i$ are given by

$$\mathrm{SQ\_io\_std\_moving}_i = \mathrm{intra}_i + \mathrm{inter}_i, \quad (16)$$

with

$$\begin{aligned} \text{intra}_i &= 0.28 \cdot \text{shape\_reg}_i + 0.29 \cdot \text{mot\_unif}_i, \\ \text{inter}_i &= 0.19 \cdot \text{contrast}_i + 0.24 \cdot \text{mot\_unif}_{\text{neigh\_diff}_i}, \\ \text{shape\_reg}_i &= 0.5 \cdot \text{circ\_elong}_i + 0.5 \cdot \text{compact}_i, \\ \text{mot\_unif}_i &= \text{crit}_i . \end{aligned} \tag{17}$$

4. OVERALL SEGMENTATION QUALITY EVALUATION

The overall objective segmentation quality evaluation combines each individual object's segmentation quality evaluation mark, with the corresponding relevance in the scene and a factor reflecting the similarity between the sets of target and estimated objects.

Individual object evaluation has been specified in Section 3. The relevance of an object in the scene is evaluated using a metric called *Relative Contextual Relevance* (RC_rel), which has been previously proposed by the authors in [17]. This metric computes a relevance mark reflecting how much the human viewer attention is attracted by a given object, producing results in the [0, 1] range, with the restriction that the relevancies of all objects composing a partition at a given time instant have to sum one. A mark of one corresponds to the highest possible relevance.

The assessment of the similarity of objects for stand-alone segmentation quality evaluation, and a proposal for the overall segmentation quality metric are presented below.

4.1. Similarity of objects evaluation

The degree of correspondence between the objects found by a segmentation algorithm and those targeted by the application addressed must be taken into account by the overall segmentation quality metric. This is done in the *similarity of objects* evaluation step, by computing a metric called sim_obj_factor, which is a multiplicative factor to include in the computation of the overall segmentation quality evaluation.

For stand-alone segmentation quality evaluation, a first object similarity check can be done, if the target number of objects is known, by measuring a ratio between the target and estimated numbers of objects. The ratio proposed is defined by

$$\text{num\_obj\_comparison} = \frac{\min\left(\text{num\_est\_obj}, \text{num\_target\_obj}\right)}{\max\left(\text{num\_est\_obj}, \text{num\_target\_obj}\right)}, \tag{18}$$

where num_est_obj and num_target_obj refer to the estimated and the target number of objects, respectively. The num_obj_comparison metric takes value one when the estimated number of objects is equal to the target number, and smaller values as the two numbers become more different.

The metric above provides a limited amount of information about the correctness of the correspondence between estimated and target objects since it does not distinguish between too many or too few objects in the estimated segmentation.

To make the sim_obj_factor metric more informed, it is possible to consider also a measure of the partition stability, applicable to the cases where the evolution of the number of objects in a segmentation partition is assumed to be smooth. In this case, not many objects are expected to enter or leave the scene too frequently, and thus an additional, or alternative if the number of target objects is not known, metric can be defined, evaluating the number of label changes between consecutive time instants

$$\text{num\_obj\_stability} = \frac{\min\left(\text{num\_est\_obj}_{i-1}, \text{num\_est\_obj}_i\right)}{\max\left(\text{num\_est\_obj}_{i-1}, \text{num\_est\_obj}_i\right)}, \tag{19}$$

where num_obj_{i-1} and num_obj_i refer to the number of estimated objects in the previous and in the current time instants, respectively.

This num_obj_stability metric indicates if the number of objects in the partition has remained stable (value close to one) or not (metric value approaching zero).

The proposed sim_obj_factor metric for stand-alone segmentation quality evaluation is thus obtained by the multiplication of the two individual factors, num_obj_comparison and num_obj_stability, if both are available

$$\text{sim\_obj\_factor} = \text{num\_obj\_comparison} \cdot \text{num\_obj\_stability}. \tag{20}$$

Whenever one of the two factors above cannot be computed, for instance if the number of target objects present at each time instant is not known, or if the stability hypothesis is not applicable, only the other factor is considered in the sim_obj_factor. If none of the factors can be computed, then the sim_obj_factor cannot be taken into account for the final segmentation quality evaluation.

To obtain a sim_obj_factor representative of the complete sequence or shot, and since the two factors may vary as time evolves, a temporal integration of the instantaneous values can be done through their temporal average.

4.2. Metric for overall stand-alone segmentation quality evaluation

The proposal for the overall stand-alone segmentation quality evaluation metric combines the appropriate measures of individual object quality (depending on the type of content), the object's relevance and the similarity of objects factor. An initial proposal for an overall segmentation quality evaluation metric (SQ) is

$$\text{SQ} = \text{sim\_obj\_factor} \cdot \left(\sum_{j=1}^{\text{num\_objects}} \left(\text{SQ\_io}\left(E_j\right) \cdot \text{RC\_rel}\left(E_j\right)\right) \right), \tag{21}$$

where $\text{SQ\_io}(E_j)$ is the individual object segmentation

quality mark estimated for object j, $\mathrm{RC\_rel}(E_j)$ is the corresponding relative contextual relevance, and sim_obj_factor is the factor evaluating the degree of correspondence between the detected and target objects. The sum is performed for all the estimated objects in the scene segmented.

Alternatively, to more explicitly include the temporal dimension into the computation of the overall segmentation quality evaluation, instead of taking the temporally averaged marks for its various components and multiplying them together, the overall segmentation quality may be computed by weighting the instantaneous qualities of the various objects by their instantaneous relevance values. This alternative is justified by the fact that one object may have large variations in its quality or relevance marks along time. For instance, if an object has a bad segmentation quality during the short temporal period where it is very relevant, the overall segmentation quality metric should be more penalized than what is expressed using (21), where the object's low average relevance is multiplied by its average quality. Also the similarity of objects factor may have fluctuations along time that should be instantaneously acknowledged by the composite metric. Thus, the final proposal for the overall stand-alone segmentation quality evaluation metric computes the temporal average of the instantaneous values, as given by

$$\mathrm{SQ} = \frac{1}{N}\cdot\sum_{i=1}^{N}\left[\mathrm{sim\_obj\_factor}_i \cdot\left(\sum_{j=1}^{\mathrm{num\_objects}}\left(\mathrm{SQ\_io}_i(E_j)\cdot\mathrm{RC\_rel}_i(E_j)\right)\right)\right]. \tag{22}$$

This overall segmentation quality evaluation metric expresses the overall segmentation quality as a sum of the individual object segmentation quality marks weighted by the corresponding contextual relevance and affected by the similarity of objects factor, for each time instant. The higher the individual object quality is for the more relevant objects, the better is the overall segmentation quality, ensuring that the most relevant objects, which are the most visible to the human observers, have a larger impact on the overall segmentation quality result. Furthermore, the mismatch between the target objects and the estimated ones is expressed through an object similarity corrective factor, taking values between zero and one, and penalizing the overall segmentation quality if the target objects are incorrectly matched.

5. STAND-ALONE SEGMENTATION QUALITY EVALUATION RESULTS

This section presents and discusses the results obtained using the two composite metrics proposed for the stand-alone segmentation quality evaluation of individual objects and of entire segmentation partitions. Since the two proposed stand-alone metrics are applicable only under certain circumstances, each of the stand-alone composite metrics is tested with the appropriate content.

The test sequences and the corresponding segmentation partitions used are described below, before presenting the segmentation quality evaluation results obtained with the proposed composite metrics.

5.1. Test sequences and segmentation partitions

Several test sequences, mainly from the MPEG-4 test set, showing different spatial complexity and temporal activity characteristics have been used to test the proposed segmentation quality evaluation metrics. For each sequence, several segmentation partitions with different segmentation qualities were considered.

Three subsets of the test sequences, each with 30 representative images of the desired objects' behaviour and characteristics, were used to illustrate the results obtained. These subsequences were the following.

Akiyo, images 0 to 29. This is a sequence with low temporal activity and not very complex texture. It contains two objects of interest: the woman, and the background.

News, images 90 to 119. This is a sequence with low temporal activity and not very complex texture. It contains three objects of interest: the man, the woman, and the background.

Stefan, images 30 to 59. This is a sequence with high temporal activity and relatively complex texture. It contains two objects of interest: the tennis player, and the background.

Samples of the original images and of the segmentation partitions are shown in Figures 1, 2, and 3, respectively, for the sequences *Akiyo*, *News*, and *Stefan*. The segmentation partitions labelled as *reference* are those made available by the MPEG group; the other partitions were created with different segmentation quality levels, ranging from a close match with the reference to more objectionable segmentations. Notice that for the sequence *News* the reference segmentation provided by the MPEG group is not used, as the objects of interest here are not the same as those considered by MPEG.

5.2. Results and analysis

Stand-alone segmentation quality evaluation metrics are applicable only in certain circumstances, and thus the two metrics proposed have been tested with the appropriate contents. Results for these metrics, considering both the *individual object* and the *overall* evaluation cases are included below.

A set of preliminary experiments showed that similar segmentation quality evaluation results are produced independently of the input format, for example, CIF and QCIF, and thus the QCIF resolution was used to limit the algorithm execution time.

The results presented below include, for each test sequence, a graph, representing the temporal evolution of the overall segmentation quality, and a table, containing the temporal average of the instantaneous results computed both for individual object and for overall segmentation quality evaluation.

Content class I corresponds to video sequences which have relatively simple shapes, and present a limited amount of motion. To evaluate this type of content, the *Akiyo* and

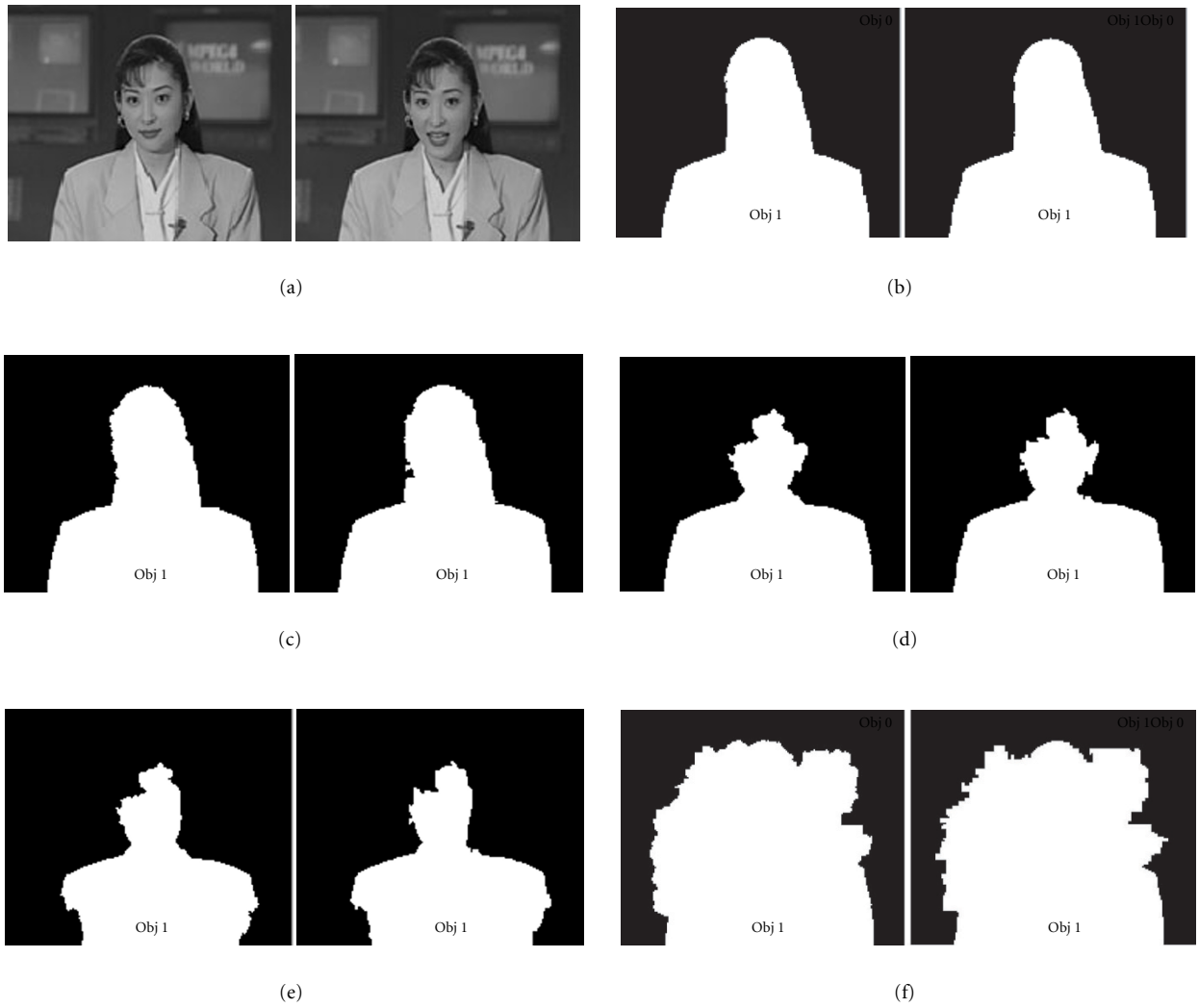

Figure 1: Sample original images (a) and segmentation partitions: *reference* (b), *seg1* (c), *seg2* (d), *seg3* (e), and *seg4* (f) for the images number 0 and 29 of the sequence *Akiyo*.

the *News* test sequences and the corresponding segmentation partitions were used.

For a human observer, the ranking of the segmentation partitions provided for the sequence *Akiyo* would most likely list the reference and segmentation 1 as having the best quality, followed by segmentation 2, then segmentation 3, and, finally, segmentation 4 would be considered the worst segmentation.

The results of the proposed objective evaluation algorithms, included in Figure 4 and in Table 1, show three segmentation quality groups for the *woman* object: the best quality is achieved by the reference, segmentation 1, and segmentation 2, then segmentation 3 achieves intermediate quality, and, finally, segmentation 4 gets the worst results. In this case, the reference segmentation does not get the best evaluation result since a part of the woman's hair is intensely illuminated, and when included as part of the woman it leads to a lower contrast to the background than when it is omitted, as it happens with segmentations 1 and 2. Segmentation 4, for which the *woman* object captures a significant part of the *background*, is clearly identified as the worst segmentation. Table 1 also shows that the individual object stand-alone segmentation quality results for the *background* object are less discriminative than for the *woman* object, but still clearly distinguish segmentation 4, for which the woman object captures a significant part of the *background*, as being worst than the other segmentations. The overall segmentation quality results also show the same three quality groups as for the individual object results, following the same ordering, and matching well the subjective ranking performed by human viewers.

The behaviour of the stand-alone segmentation quality evaluation metric for stable content, for sequences with more than two objects, is illustrated using the sequence *News*.

From a human observer point of view, the ranking of the segmentation partitions provided for the sequence *News* in

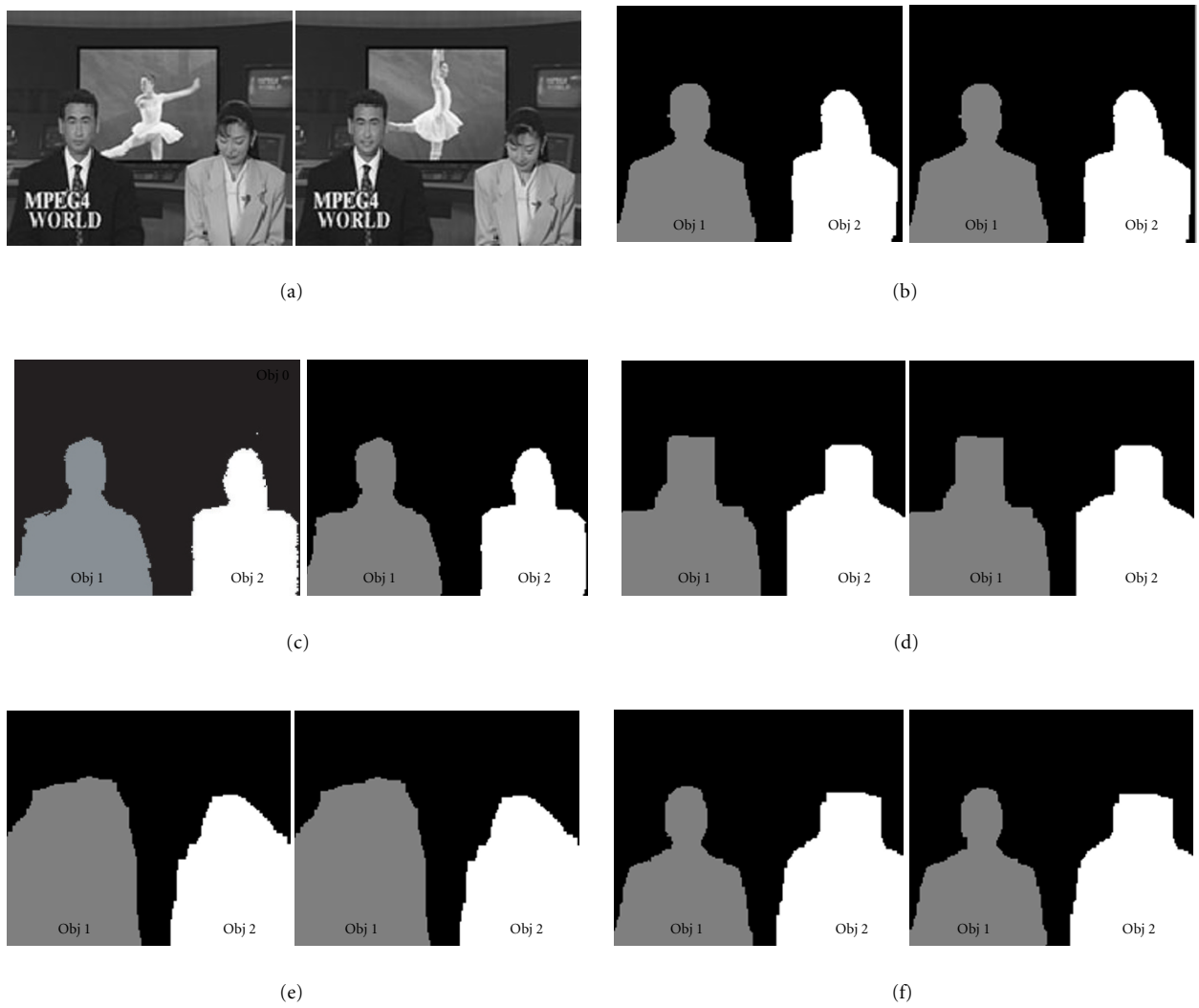

Figure 2: Sample original images (a) and segmentation partitions: *seg1* (b), *seg2* (c), *seg3* (d), *seg4* (e), and *seg5* (f) for the images number 90 and 119 of the sequence *News*.

terms of their segmentation quality would be in the order of their numbering. In fact, segmentation 1 has object contours very close to their correct positions, thus corresponding to the best quality. Segmentation 2 includes some small errors in the object contours, being the second best segmentation. Then, segmentation 3 has incorrect contours, but the shapes resemble the newscasters' objects. Segmentation 4 also has incorrect contours, but since the shapes of the newscasters are less similar to the desired shapes it would, very likely, be considered as the worst segmentation. In segmentation 5, the *man* is as well segmented as for segmentation 1, while the segmentation for the *woman* is somewhat worst than for segmentation 3; therefore, the subjective quality result would probably be some intermediate mark between those of segmentations 1 and 3.

The objective segmentation quality evaluation results for the sequence *News* are presented in Figure 5 and in Table 2. The overall results identify three levels of quality: the best quality is achieved by segmentation 1, then segmentation 2 achieves intermediate quality, and, finally, segmentations 3 and 4 get the worst values. As expected, segmentation 5 gets an intermediate overall segmentation quality value between those of segmentations 1 and 3. The main difference regarding the subjective evaluation ranking mentioned above is that the automatic algorithm did not distinguish between the qualities of segmentations 3 and 4. This is explained by the type of segmentation errors observed in these two segmentation partitions, which are accounted by the objective metrics in a similar manner: they both add part of the background (which is relatively homogeneous in texture) to the newscasters' objects; moreover, none of the considered object shapes is very irregular.

In terms of individual object segmentation quality results, the marks obtained for segmentation 5 show that the automatic evaluation algorithm is capable of distinguishing the quality of the different objects: the *man* object achieves

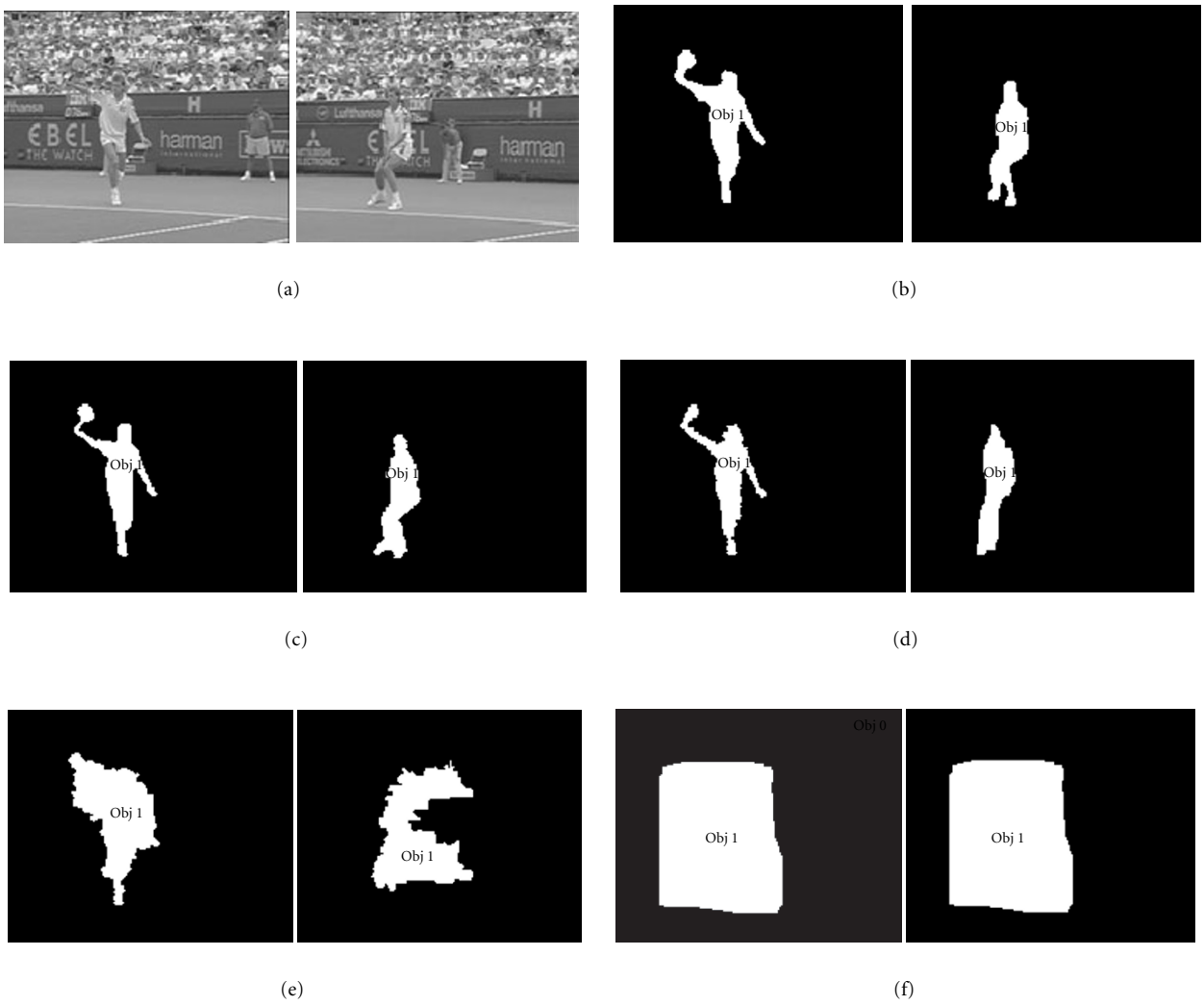

Figure 3: Sample original images (a) and segmentation partitions: *reference* (b), *seg1* (c), *seg2* (d), *seg3* (e), and *seg4* (f) for the images number 30 and 59 of the sequence *Stefan*.

the highest average individual object quality, together with segmentations 1 and 2, while the *woman* object gets the lowest average mark, together with segmentation 3, and the remaining *background* object gets an intermediate quality mark, as expected.

As shown by the two examples above, the stand-alone segmentation quality evaluation algorithm reveals itself capable of ranking the qualities of the various segmentation partitions, but the results should be interpreted in a more qualitative and relative way (e.g., for ranking purposes or for mutual comparison), rather than in a quantitative and absolute manner.

Content class II corresponds to more complex video content than that for the previous case. Object shapes may not be so simple, and motion should be more important. The sequence *Stefan* and the corresponding segmentation partitions were used to evaluate the metric proposal made in this paper for this type of content.

For the sequence *Stefan*, a human observer would most likely rank the segmentation partitions provided in the following order: the reference segmentation and segmentation 1 as having the best quality, closely followed by segmentation 2, then segmentation 3, and, finally, segmentation 4.

The results of the objective evaluation algorithm, included in Figure 6 and Table 3, show that segmentation 1 gets the best overall segmentation quality result followed by a group formed by the reference and segmentations 2 and 3. Segmentation 4 gets the worst result. These results can be explained as follows: segmentation 1 is in fact more precise than the reference partition, as the reference is smoother and sometimes includes fragments of the *background* as belonging to the *player* object; the reference and segmentation 2 are correctly classified as the next quality group, while segmentation 3 receives a higher ranking than expected due to the fact that it always includes the moving *player* object, which is not very contrasted to the surrounding *background* area.

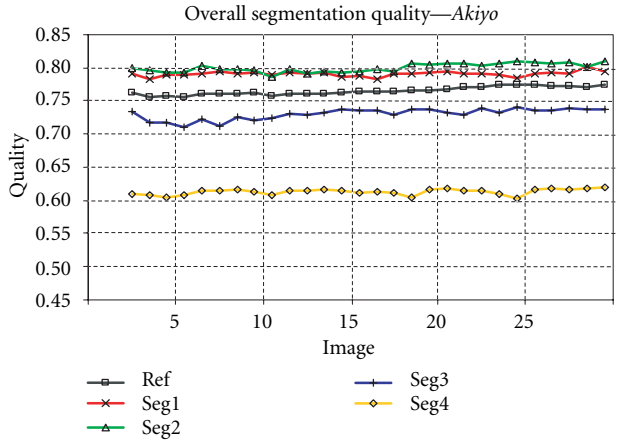

Figure 4: Stand-alone overall and individual object quality evaluation results for the sequence *Akiyo*.

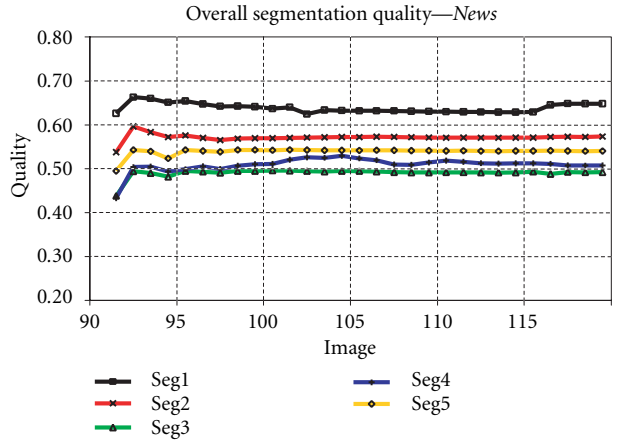

Figure 5: Stand-alone overall and individual object quality evaluation results for the sequence *News*.

Table 1: Stand-alone overall and individual object quality evaluation results for the sequence *Akiyo*.

	Average segmentation quality		
	Background	*Woman*	Overall
Ref	0.76	0.77	0.77
Seg1	0.79	0.79	0.79
Seg2	0.79	0.80	0.80
Seg3	0.73	0.73	0.73
Seg4	0.65	0.56	0.60

Table 2: Stand-alone overall and individual object quality evaluation results for the sequence *News*.

	Average segmentation quality			
	Background	*Man*	*Woman*	Overall
Seg1	0.63	0.57	0.69	0.64
Seg2	0.57	0.57	0.56	0.57
Seg3	0.47	0.49	0.49	0.49
Seg4	0.47	0.51	0.50	0.50
Seg5	0.53	0.57	0.49	0.54

Finally, segmentation 4 is correctly ranked as the worst, since the detected object mask is static in time, including a large amount of *background* as part of the *player* object. For this case, the overall segmentation quality marks are always in the lower half of the segmentation quality scale since the objective evaluation metrics do not find the objects to be very homogeneous either in texture or in motion, and thus cannot conclude that the best segmentations are rather good for a human observer (at least in the context of the assumptions made).

The results obtained show that the stand-alone segmentation quality evaluation algorithms proposed are capable of ranking the quality of the various segmentation partitions, but the results must be interpreted in a rather qualitative and relative way (e.g., for ranking purposes). Stand-alone evaluation results are not expected to be as reliable as those obtained with relative evaluation when a ground truth segmentation is available, but they can still be very useful for identifying the segmentation quality classes among the various tested segmentations/algorithms which is a major problem in the context of emerging interactive multimedia applications.

6. CONCLUSIONS

Video segmentation quality evaluation is a key element whenever the identification of a set of objects in a video sequence is required since it allows the assessment of the performance of segmentation algorithms in view of a given application targets. However, a satisfying solution for objective segmentation quality evaluation is not yet available.

This paper discusses the objective segmentation quality evaluation problem, in particular when a reference segmentation playing the role of "ground truth" is not available—stand-alone evaluation, and proposes metrics for both individual object and for overall stand-alone segmentation quality evaluation.

As expected, stand-alone evaluation revealed itself sensitive to the type of application/content considered. The various classes of elementary metrics available are not universally applicable, but when carefully selected metrics are employed for given classes of content then very useful segmentation quality evaluation results can be obtained. Two such metrics are proposed in this paper: for stable content and for moving content.

It is recognised that stand-alone objective segmentation quality evaluation is not as powerful as relative evaluation, but stand-alone evaluation results allow the comparative analysis of segmentation results and thus of segmentation algorithms, which is an important functionality for the adequate design of video segmentation enabled systems.

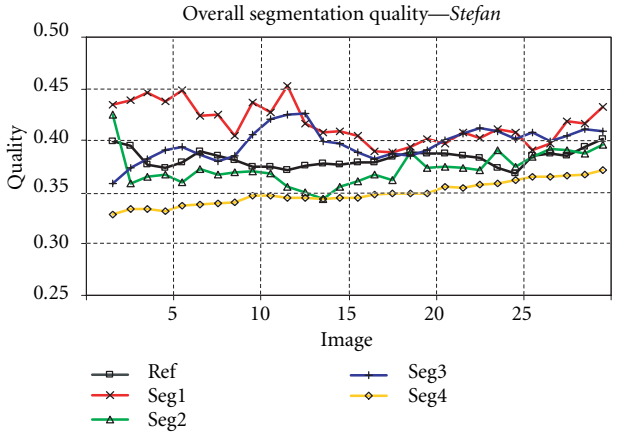

Figure 6: Stand-alone overall and individual object quality evaluation results for the sequence *Stefan*.

Table 3: Stand-alone overall and individual object quality evaluation results for the sequence *Stefan*.

	Average segmentation quality		
	Background	*Player*	Overall
Ref	0.33	0.43	0.38
Seg1	0.34	0.49	0.42
Seg2	0.32	0.43	0.38
Seg3	0.34	0.45	0.39
Seg4	0.32	0.37	0.34

REFERENCES

[1] ISO/IEC 14496, "Information technology—coding of audio-visual objects," 1999.

[2] MPEG Requirements Group, "MPEG-7 overview," Doc. ISO/IEC JTC1/SC29/WG11 N4031, March 2001, Singapore MPEG Meeting.

[3] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Recommendation BT.500-7, 1995.

[4] ITU-T, "Subjective video quality assessment methods for multimedia applications," Recommendation P.910, August 1996.

[5] COST 211quat, "Redundancy reduction techniques and content analysis for multimedia services," COST project, http://www.iva.cs.tut.fi/COST211/.

[6] COST 211quat, "Call for AM comparisons—compare your segmentation algorithm to the COST 211quat analysis model," COST project, available at http://www.iva.cs.tut.fi/COST211/Call/Call.htm.

[7] G. Rees and P. Greenway, "Metrics for image segmentation," in *Workshop on Performance Characterisation and Benchmarking of Vision Systems*, pp. 20–37, Essex, UK, January 1999.

[8] Y. Zhang and J. Gerbrands, "Objective and quantitative segmentation evaluation and comparison," *Signal Processing*, vol. 39, no. 1–2, pp. 43–54, 1994.

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

[10] M. Heath, S. Sarkar, T. Sanocki, and K. Bowyer, "A robust visual method for assessing the relative performance of edge-detection algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 12, pp. 1338–1359, 1997.

[11] M. Levine and A. Nazif, "Dynamic measurement of computer generated image segmentations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 155–164, 1985.

[12] P. Villegas, X. Marichal, and A. Salcedo, "Objective evaluation of segmentation masks in video sequences," in *WIAMIS' 99*, pp. 85–88, Germany, 31 May–1 June 1999.

[13] M. Wollborn and R. Mech, "Refined procedure for objective evaluation of video object generation algorithms," Doc. ISO/IEC JTC1/SC29/WG11 M3448, March 1998.

[14] 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.

[15] S. Wolf and A. Webster, "Subjective and objective measures of scene criticality," in *ITU Meeting on Subjective and Objective Audiovisual Quality Assessment Methods*, Turin, Italy, October 1997.

[16] J. Serra, *Image Analysis and Mathematical Morphology*, vol. 1, Academic Press, San Diego, Calif, USA, 1988.

[17] P. Correia and F. Pereira, "Estimation of video object's relevance," in *EUSIPCO' 2000*, pp. 925–928, Finland, September 2000.

Paulo Lobato Correia graduated as an Engineer and obtained an M.S. in electrical and computers engineering from Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Portugal, in 1989 and 1993, respectively. He is currently working towards a Ph.D. in the area of image analysis for coding and indexing. Since 1990 he is a Teaching Assistant at the Electrical and Computers Department of IST, and since 1994 he is a researcher at the Image Communication Group of IST. His current research interests are in the area of video analysis and processing, including video segmentation, objective video segmentation quality evaluation, and content-based video description and representation.

Fernando Pereira was born in Vermelha, Portugal in October 1962. He was graduated in Electrical and Computers Engineering by Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Portugal, in 1985. He received the M.S. and Ph.D. degrees in Electrical and Computers Engineering from IST, in 1988 and 1991, respectively. He is currently Professor at the Electrical and Computers Engineering Department of IST. He is responsible for the participation of IST in many national and international research projects. He is a member of the Editorial Board and Area Editor on Image/Video Compression of the Signal Processing: Image Communication Journal and an Associate Editor of IEEE Transactions of Circuits and Systems for Video Technology. He is a member of the Scientific Committee of several international conferences. He has contributed more than one hundred papers. He won the 1990 Portuguese IBM Award and an ISO Award for Outstanding Technical Contribution for his participation in the development of the MPEG-4 Visual standard, in October 1998. He has been participating in the work of ISO/MPEG for many years, notably as the head of the Portuguese delegation, and chairing many Ad Hoc Groups related to the MPEG-4 and MPEG-7 standards. His current areas of interest are video analysis, processing, coding and description, and multimedia interactive services.