# A Method for Single-Stimulus Quality Assessment of Segmented Video

**R. Piroddi[1] and T. Vlachos[2]**

[1] Department of Electrical and Electronic Engineering, Imperial College London, Exhibition Road, London SW7 2AZ, UK
[2] Centre for Vision, Speech and Signal Processing (CVSSP), School of Electronics and Physical Sciences, University of Surrey, Guildford GU2 7XH, UK

We present a unified method for single-stimulus quality assessment of segmented video. This method takes into consideration colour and motion features of a moving sequence and monitors their changes across segment boundaries. Features are estimated using a local neighbourhood which preserves the topological integrity of segment boundaries. Furthermore the proposed method addresses the problem of unreliable and/or unavailable feature estimates by applying normalized differential convolution (NDC). Our experimental results suggest that the proposed method outperforms competing methods in terms of sensitivity as well as noise immunity for a variety of standard test sequences.

## 1. INTRODUCTION

Object-based descriptions of still images and moving sequences are becoming increasingly important for multimedia and broadcasting applications offering many well-documented advantages [1]. Such descriptions allow the authoring, manipulation, editing, and coding of digital imagery in a far more creative, intuitive, efficient, and user-friendly manner compared to conventional frame-based alternatives. A key tool towards the identification of objects or regions of interest is segmentation which has emerged as a very active area of research in the past 20 years. Segmentation has often been regarded as a first step towards automated image analysis with applications in scene interpretation, object recognition, and compression, especially in view of the fact that it was shown to be well tuned to the characteristics of human vision.

Despite its potential usefulness, segmentation is a fundamentally ill-posed problem and, as a consequence, generic non-application-specific solutions have remained elusive [2]. Additionally, a critical factor which has prevented any particular algorithm from gaining wider acceptance has been the lack of a unified method for the quality assessment of segmented imagery. While such assessment has traditionally relied on subjective means, it is self-evident that the development of an objective evaluation methodology holds the key to further advances in the field.

In Figure 1, a classification of quality assessment methods for video object-based segmentation is shown. *Reference* methods require ground-truth information as opposed to *no-reference* methods, which have no such requirement. No-reference methods can be further subdivided to *interframe*, where the temporal consistency of segmentation from one frame to another is taken into consideration, and *intraframe*, where this is not an issue.

In relation to the assessment of still segmented images, although there have been a number of noteworthy attempts such as [3] for grey-level imagery and [4] for colour imagery, a commonly accepted approach has not emerged. Other researchers have incorporated elements of human visual perception [5], especially in the field of image compression [6]. Nevertheless such efforts have been moderately successful in establishing a credible relationship between human visual perception and an objective measurement of quality.

In the case of moving sequences, much less work has been reported despite the demand for a standardised objective evaluation methodology from the broadcasting and entertainment industry [7]. Given the lack of objective and automatic means for evaluation, the generic assessment standard is based on subjective evaluation [8, 9], which is cumbersome, difficult to organise, and requires dedicated infrastructure of a very high specification [10].

The straightforward application of metrics developed for the evaluation of video sequence segmentation has been attempted and proved ineffective [11]. Such metrics are in fact well suited to describe similarity or dissimilarity between homogeneous quantities, while video object segmentation
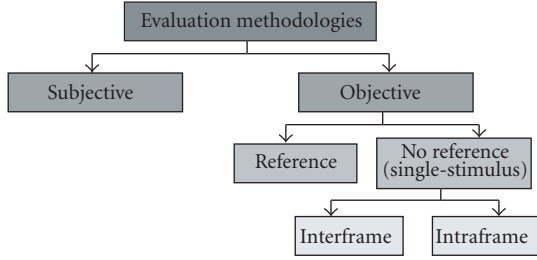
FIGURE 1: Methodologies for quality assessment of video object production.

often involves the complex interaction of inhomogeneous features [1] making the performance evaluation of *video object* segmentation even more difficult than the one of still image segmentation [12].

Most performance evaluation methods suitable for object-based video segmentation rely on the use of ground truth [14–16]. In [16, 17], a human visual system (HVS) driven approach is presented using a perceptually weighted set of evaluation metrics. The creation of suitable ground truth information typically involves the manual segmentation of moving objects of interest. Unfortunately this requires a formidable amount of operator effort, concentration, and experience and ultimately prevents any systematic experimentation beyond just a limited number of frames.

Taking into account the above difficulties, it is evident that methods that do not rely on ground-truth references (single stimulus) would be of significant practical value especially for the purpose of algorithmic performance comparisons involving sequences of a longer duration. With some notable exceptions [13, 18] this class of no-reference assessment methods is rather under-represented in the literature.

In this work, we formulate a single-stimulus, intraframe assessment method suitable for the evaluation of the performance of object-based segmentation algorithms. Some aspects of our approach are derived from the single-stimulus method described in [13]. An important element of our approach is the consideration of local spatial and temporal characteristics of an object of interest on a frame-by-frame basis. This diminishes the influence of object inhomogeneity on the overall result. On the other hand, the colour and motion boundary criteria used in [13] do not take into account that objects are coherent spatio-temporal entities.

The novelty of our approach lies additionally in the development of a unified method for dealing with both spatial and temporal data in the presence of noisy and uncertain data. This method relies on the concept of normalised differential convolution (NDC). The criteria for the localisation of correct spatial and temporal boundaries are enriched by the introduction of a requirement on the spatio-temporal consistency of the contrast information. The approach is independent of parameter definition and experimental results show an increased robustness to noise and increased sensitivity to local error with respect to the methods already proposed [13].

The proposed evaluation method is of great help not just in the performance evaluation of segmentation, but also in the correction of erroneous segmentations in all those areas requiring a high segmentation quality. Referring to the classification of application scenarios in [19], this methodology targets both off-line user-interactive and non-user-interactive applications and real-time user-interactive applications. Examples of the first category are all applications that need to produce semantic information, which may be reused: broadcasting and video production for database storage. Examples of the second category are videotelephony and videoconference.

This paper is structured as follows. In Section 2, the conceptual methodology for obtaining local accuracy measures without the use of ground truth is presented. In Section 3, the characteristics of the current local methods are described, improvements to the current methodology are suggested, and the improved methodology is embedded in a unified method for dealing with spatial and temporal data in presence of noise and uncertainty. In Section 4, the proposed method is compared to the previous methodology with the use of both automatic object segmentation and ground truth, obtained by manual segmentation, and its application to algorithmic performance comparison is demonstrated. Conclusions follow in Section 5.

## 2. METRICS USING COLOUR AND MOTION DISPARITIES

The proposed method relies on the computation of metrics which capture the disparity in terms of colour and motion between adjacent regions in a previously generated segmentation map. In that sense, our work has similarities with [20] and for the benefit of the reader, we briefly summarise some of the key notions.

### 2.1. Colour disparity metric

The colour values of pixels *just inside* and *just outside* of a segment boundary are considered. In order to define the *just outside* and *just inside*, normal lines of length $L$ are drawn from the boundary at equal intervals towards the outside and the inside of the segment as shown in Figure 2(a), obtaining $K$ sampling points on the boundary. The end points are marked as $p_O^i$ and $p_I^i$, for $i = 1, \ldots, K$. The colour disparity metric $d_C(t)$, of a segment in frame $t$ is defined in (1) and (2) below:

$$0 \leq d_C(t) = \frac{1}{K} \sum_{i=1}^{K} d_C(t, i) \leq 1, \qquad (1)$$

where

$$d_C(t, i) = \frac{||C_O^i(t) - C_I^i(t)||}{\sqrt{3 \times 255^2}} \qquad (2)$$

and $C_O^i(t)$ is the average colour calculated in an $M \times M$ neighbourhood of pixel $p_O^i(x, y, t)$. $C_I^i(t)$ is defined similarly. The colour metric for the whole sequence is

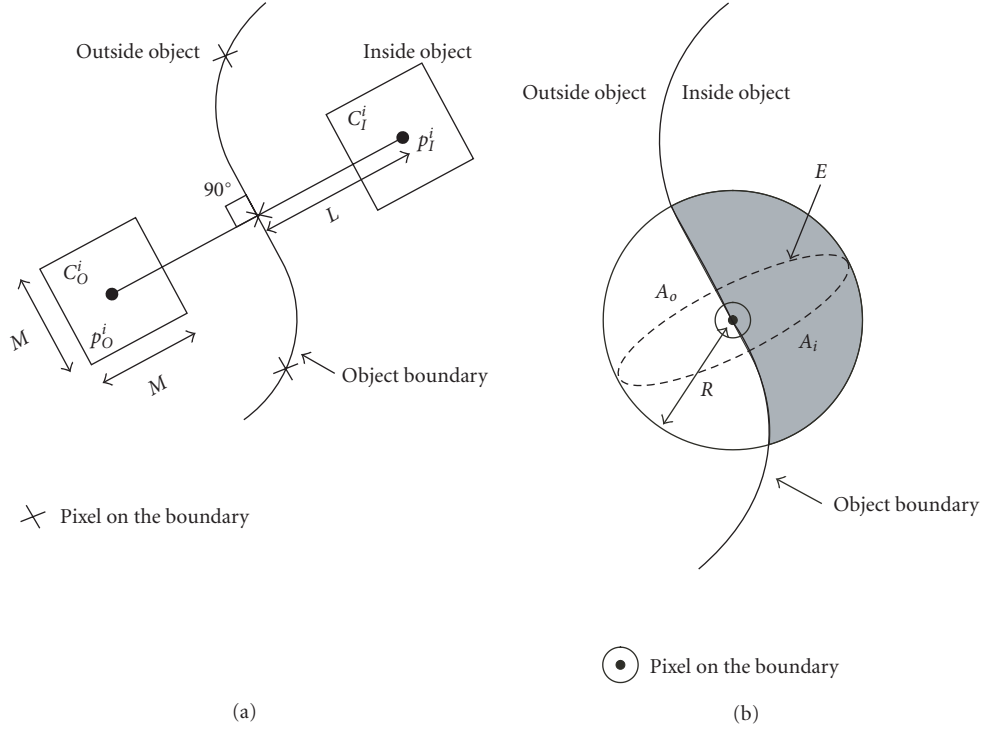$$0 \leq D_C = f(d_C(t), \, t = 1, \ldots, T), \qquad (3)$$

FIGURE 2: (a) Definition of *just inside* and *just outside* areas for the computation of contrast in [13] and (b) definition of the support area for the *applicability function* in the NC/NDC.

where $f(\cdot)$ denotes a linear function obtained by the contributions of $T$ colour disparities measures $d_C$ calculated for frames at instants $t = 1, \ldots, T$, and $\|\cdot\|$ denotes the Euclidean distance.

### 2.2. Motion disparity metric

The motion metric $d_M(t)$ for a frame $t$ is conceptually similar to the colour metric discussed above. Here, $v_O^i(t)$ and $v_I^i(t)$ denote the average motion vectors calculated in an $M \times M$ neighbourhood of pixels $p_O^i(x, y, t)$ and $p_I^i(x, y, t)$. Then, $d(v_O^i(t), v_I^i(t))$ denotes the distance between the two average motion vectors and is calculated according to the following:

$$0 \leq d\left(v_O^i(t), v_I^i(t)\right) = \frac{\left\|v_O^i(t) - v_I^i(t)\right\|}{\left\|v_O^i(t)\right\| + \left\|v_I^i(t)\right\|}. \tag{4}$$

Whenever possible, it is advisable to associate a reliability measure to the estimates of the motion vectors. In [20] the reliability measure is based on the motion and colour coherence in the prediction of the motion between frame $t$ and $t+1$. Let us denote $b^i(t+1)$ as the backward motion vector at location $p^i + v^i$ in frame $t+1$; $c(p^i, t)$ as the colour intensity; and parameters $\sigma_m$ and $\sigma_c$ as the standard deviations of the motion field and colour in frame $t$, respectively. The reliability measure $R(v^i(t))$ for a neighbourhood around pixel $i$ in

frame $t$ is defined as

$$\begin{aligned} R\left(v^i(t)\right) = {} & \exp\left(-\frac{\left\|v^i(t) - b^i(t+1)\right\|^2}{2\sigma_m^2}\right) \\ & \times \exp\left(-\frac{\left\|c(p^i, t) - c(p^i + v^i, t+1)\right\|^2}{2\sigma_c^2}\right). \end{aligned} \tag{5}$$

For each sample $i$ on the boundary of a segmented object, two motion averages $v_O^i(t)$ and $v_I^i(t)$ of a neighbourhood immediately outside and immediately inside the boundary location $i$ should be calculated. Therefore, the total reliability measure $w_i$ for the location $i$ is a combination of the reliability measures of $v_O^i(t)$ and $v_I^i(t)$:

$$0 \leq w_i = R\left(v_O^i(t)\right) \cdot R\left(v_I^i(t)\right) \leq 1. \tag{6}$$

The reliability measure may be used as a weight for the distance measure expressed by $d(v_O^i(t), v_I^i(t))$, defined in (4). This is necessary to reduce the influence of erroneous estimates in the calculation of the motion disparity metric. The weighted distance between the two average motion vectors is then defined as

$$d_M(t, i) = d\left(v_O^i(t), v_I^i(t)\right) \cdot w_i. \tag{7}$$

Finally, the overall motion metric $d_M(t)$ is obtained as the sum of the differences in corresponding motion vectors *just inside* and *just outside* the motion boundary (a sort of motion *contrast*) weighted by the reliability of the same motion

vectors and normalised by the sum of all the weights, for a number $K$ of boundary samples of the object in frame $t$. This is expressed by

$$0 \leq d_M(t) = 1 - \frac{\sum_{i=1}^{K} d_M(t,i)}{\sum_{i=1}^{K} w_i} \leq 1. \tag{8}$$

## 3. NEIGHBOURHOOD TOPOLOGY

The neighbourhood topology used in [20] is subject to the following limitations.

(i) Occasional unreliability due to the fact that the averages are calculated in an area further away from the boundary. In fact the closest pixel is at a distance $L - (M/2)$.

(ii) No adaptation to the local structure of the boundary. The neighbourhood used for the calculation of the averages does not follow the local curvature of the boundary, but its shape is fixed.

(iii) The distance from the boundary is not taken into account. All the pixels in the neighbourhood contribute in equal measure to the average, irrespective to their actual distance from the boundary, which can be up to $L + (M/2)$.

In response to the above we have redesigned the neighbourhood topology so that it follows closely the actual boundary between two segments and therefore provides an element of local adaptation.

In Figure 2(b) a schematic description of the proposed improvement is shown. Metrics are calculated for each point $p_b$ belonging to the boundary. The area for the calculation of the contrast is defined by a circle of radius $R$ centred in $p_b$. This area of support closely follows the object boundary and allows the collection of information from areas adjacent to the boundary inside, $A_i$, and outside, $A_o$, the moving object.

### 3.1. Treatment of unreliable and missing data

It should be noted that not all boundary elements contribute to the calculations, but an element of sampling is introduced in [20]. In this work, we avoid the sampling of the boundary when possible. However, especially when dealing with motion information, pixels along the boundary may convey noisy or incorrect information and may need to be excluded from the computation, introducing some irregular sampling. This may lead to further difficulties in the determination of the sampling points: if they are regularly spaced, it is possible that they ignore salient features of the contour. If they are irregularly spaced, there is the added complication of determining a suitable sampling criterion and a strategy needs to be developed for dealing with locations that do not contribute to the sampling operation, in which case data will be missing altogether. Additionally, if colour/intensity information inside the data collection neighbourhood is relatively homogeneous, the corresponding motion estimates are likely to be unreliable.

We reduce the influence of unreliable and missing data due to irregular sampling by employing the normalized differential convolution (NDC).

### 3.2. Normalized differential convolution

In [21], the problem of image analysis with irregularly sampled and uncertain data is addressed in a novel way. This involves the separation of both data and operator applied to the data in a signal part and a *certainty* part. Missing data in irregularly sampled fields are handled by setting the certainty of the data equal to zero.

In our work we consider the normalized differential convolution which is a variant of the above methodology [21–23]. In addition to the separation of the data into a signal part, which will be indicated as $f(x, y)$, and a *certainty* part, indicated as $c(x, y)$, the NDC requires the use of an *applicability* function $g(x, y)$ and its derivatives. The applicability function and its derivatives indicate what is the contribution of the data to the gradient according to their relative position. Additionally, they determine the extent of the influence of the neighbourhood to the measure.

Let us denote with $C$ the convolution of image $f(x, y)$, previously weighted by a reliability or certainty map $c(x, y)$, with a smoothing filter $g(x, y)$:

$$C(x, y) \equiv (f(x, y)c(x, y)) * g(x, y). \tag{9}$$

Let us further denote with NC the convolution of the certainty map $c(x, y)$ with the filter $g(x, y)$:

$$\text{NC}(x, y) \equiv c(x, y) * g(x, y). \tag{10}$$

Then the point-by-point division between the outputs of the two convolutions above is the *normalized convolution*. Among other applications, this has been used for image denoising and image reconstruction purposes when pixel values are occasionally unreliable or even totally unavailable within a given neighbourhood.

Dropping the explicit dependence of $C$ and NC on $(x, y)$, we now define the following:

$$\begin{aligned} C_x &\equiv (xg) * cf, \\ \text{NC}_x &\equiv (xg) * c, \\ C_y &\equiv (yg) * cf, \\ \text{NC}_y &\equiv (yg) * c, \end{aligned} \tag{11}$$

where $xg$ and $yg$ indicate the multiplication of filter $g$ with variables $x$ and $y$. As filter $g$ is a smoothing filter, filters $xg$ and $yg$ are edge enhancement filters. For the filter used in [22], $xg = x\cos^2(\pi\sqrt{x^2 + y^2}/8)$ and $yg = y\cos^2(\pi\sqrt{x^2 + y^2}/8)$ and those are shown in Figure 3.

We also define [24] vector $D_\Delta(x, y) \equiv [D_x, D_y]$, the components of which, $D_x$ and $D_y$, are calculated as follows:

$$\begin{aligned} D_x &\equiv \text{NC} \times C_x - \text{NC}_x \times C, \\ D_y &\equiv \text{NC} \times C_y - \text{NC}_y \times C. \end{aligned} \tag{12}$$

(a)


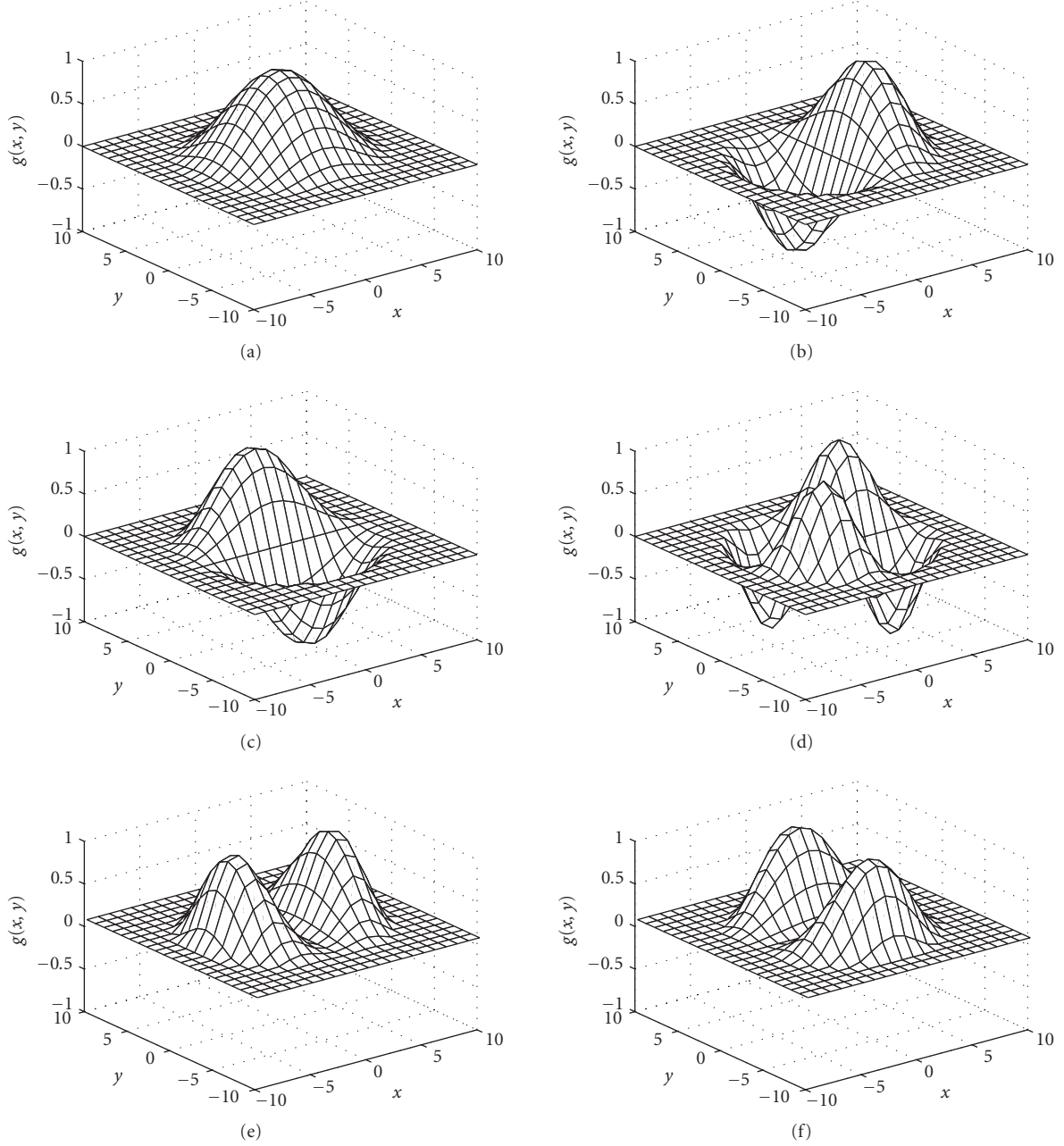
(b)



(c)



(d)



(e)



(f)

FIGURE 3: The product of filter $g(x, y)$ with variables $x$ and $y$ and some functions of them may produce some highpass filters shown here. These filters were normalised so that their maxima are equal to one, for visualisation purposes only. (a) $g(x, y)$. (b) $xg(x, y)$. (c) $yg(x, y)$. (d) $xyg(x, y)$. (e) $x^2g(x, y)$. (f) $y^2g(x, y)$.

Next we define the $2 \times 2$ matrix $N_\Delta$, as follows:

$$N_\Delta \equiv \begin{pmatrix} N_{xx} & N_{xy} \\ N_{yx} & N_{yy} \end{pmatrix}, \quad (13)$$

where

$$\begin{aligned} N_{xx} &\equiv \mathrm{NC} \times ((x^2 g) * c) - \mathrm{NC}_x^2, \\ N_{xy} &\equiv N_{yx} = \mathrm{NC} \times ((xyg) * c) - \mathrm{NC}_x \times \mathrm{NC}_y, \\ N_{yy} &\equiv \mathrm{NC} \times ((y^2 g) * c) - \mathrm{NC}_y^2. \end{aligned} \quad (14)$$

If filter $g = \cos^2(\pi\sqrt{x^2 + y^2}/8)$, then filters $x^2g$, $y^2g$, and $xyg$ are given by $x^2g = x^2\cos^2(\pi\sqrt{x^2 + y^2}/8)$, $y^2g = y^2\cos^2(\pi\sqrt{x^2 + y^2}/8)$, and $xyg = xy\cos^2(\pi\sqrt{x^2 + y^2}/8)$, and those are shown in Figure 3. The elements of matrix $N_\Delta$ depend only on the certainty of the data. $N_{xx}$ gives an estimate of the certainty of the data along the $x$ direction, $N_{yy}$ gives an estimate of the certainty of the data along the $y$ direction, and $N_{xy}$ gives an estimate of the certainty of the data along both the $x$ and $y$ directions.

The normalized differential convolution (NDC) $U_{N\Delta}$ is finally defined as

$$U_{N\Delta} \equiv N_\Delta^{-1} D_\Delta, \tag{15}$$

where $N_\Delta^{-1}$ is the inverse of the $2 \times 2$ matrix $N_\Delta$.

The effectiveness of the method towards dealing with irregularly sampled and incomplete data was demonstrated in [24, 25] for one-dimensional and two-dimensional signals, respectively. For a typical natural imagery, even if only 10% of the original pixels are known, the image gradient can be recovered to a satisfactory extent. It has also been shown, that the NC yields the best reconstruction results for reconstruction of irregularly sampled data for sampling ratios smaller than 5%. Additionally, NDC is the only method that allows the direct calculation of gradients of irregularly and sparsely sampled data [24].

### 3.3. Adaptation to local topology

As shown in Figure 3, the applicability function used in [21] and its derivatives are symmetrical and fixed in size. However, it was shown [26] that an element of adaptation to the local topology can yield performance gains relative to the performance obtained using a nonadaptive filter function [21].

For our purposes, it would be advantageous to use a smoothing function which can have variable size and orientation so that it can adapt to the local curvature of the segment boundary. This can be achieved by using a Gaussian type of function whose variance can be adjusted to provide the desired adaptation. Since our topology is inherently two-dimensional, we use a two-dimensional Gaussian function with parameters $\sigma_u$ in the horizontal direction and $\sigma_v$ in the vertical direction.

The local curvature is estimated using the *regularized gradient structure tensor* $T$ [27] defined as

$$\bar{T} = \overline{\nabla I \nabla I^T} = \lambda_u \vec{u} \vec{u}^T + \lambda_v \vec{v} \vec{v}^T, \tag{16}$$

where $I$ is the intensity of the grey-level image, $\vec{u}$ is the eigenvector of the largest eigenvalue $\lambda_u$, which determines the local orientation, the over-lining indicates the averaging of the elements over a local neighbourhood, and the superscript $T$ indicates the transpose of the vector. Defining as $A$ the local anisotropy, $A = (\lambda_u - \lambda_v)/(\lambda_u + \lambda_v)$, the scales are finally calculated as

$$\sigma_u = (1+A)\sigma_a, \qquad \sigma_v = (1-A)\sigma_a. \tag{17}$$

Using the above, the applicability function reflects the curvature of the boundary so that, for example, elongation can be induced in the direction of the normal to that boundary, as shown by the elliptical area of support $E$ in Figure 2(b). At the same time, this provides a mechanism for a reliability weighting of pixels according to their distance from the boundary.

### 3.4. Computation of metrics using the NDC

The NDC provides a way of obtaining dense contrast information on a multiplicity of different features, using sparse and/or irregular and uncertain estimates of such features. The flowchart in Figure 4 explains the method of computation of the disparity metrics with the use of the NDC. In this figure, the boundaries of the object, the segmentation of which is evaluated, are denoted collectively by $b$. However, $b$ is the union of all points $p_b$ belonging to the boundary of the object. Colour description of any frame is given by three colour channels $c_1 = R$, $c_2 = G$, and $c_3 = B$. In general, any three-dimensional colour space other than Red-Green-Blue may be employed. The motion description is given by the optic flow, which consists of two components, the horizontal $u$ and the vertical $v$ component.

To summarise, the NDC is a function of a feature $f$ calculated on a location $p$, in our case of a two-dimensional regular grid, that is, $\text{NDC} \equiv \text{NDC}(f, p)$. In the application here considered, the NDC is calculated on the location of an object boundary, indicated as $p_b$. The features considered are colour $c$, which consists of three colour planes $c_1$, $c_2$, and $c_3$, and motion $m$, which consists of the horizontal and vertical estimates of the optic flow, indicated as $u$ and $v$, respectively. The colour and motion metrics, CM and MM, respectively, are therefore calculated as

$$
\begin{aligned}
\text{CM}(p_b) &= \text{NDC}(c, p_b) \\
&= \frac{\text{NDC}(c_1, p_b) + \text{NDC}(c_2, p_b) + \text{NDC}(c_3, p_b)}{3}, \\
\text{MM}(p_b) &= \text{NDC}(m, p_b) = \frac{\text{NDC}(u, p_b) + \text{NDC}(v, p_b)}{2}.
\end{aligned}
\tag{18}
$$

The *applicability* function is adapted to the shape and the local orientation of the boundary. It provides *weighting* with regards to the distance from the boundary on the location $p_b$. It also provides *averaging* of the information on a kernel centred in $p_b$, which gives robustness to noise.

The *certainty* function provides extra *robustness* to noise as noisy data can be discarded or weighted negatively. Therefore the information is reconstructed on the basis of more certain data. Additionally, in a novel element of modelling, a part of the certainty function is used to provide an indication of the *spatio-temporal coherence* of the boundary of the objects.

This requires further explanation. In the motion metric, one may use the certainty function to model both the spatio-temporal coherence and the uncertainty of the motion estimates. In this method, the *certainty* function $c(x, y)$ is composed of three elements.

The first element is motion certainty, $mc$, a function reflecting motion estimation reliability. In our approach, a robust motion estimator has been employed [28]. Robust methods exclude, in the estimate of the motion, the points that do not comply with the model used for the estimation, that is, the outliers. We use outlier information coded into a binary map $mc$, which makes the distinction between a point being an outlier or not. Outliers are then ignored in the calculation of the NDC.

Additionally, motion estimation is more reliable in textured areas and vice versa. Thus a measure of texture activity has been incorporated as the second element of our certainty
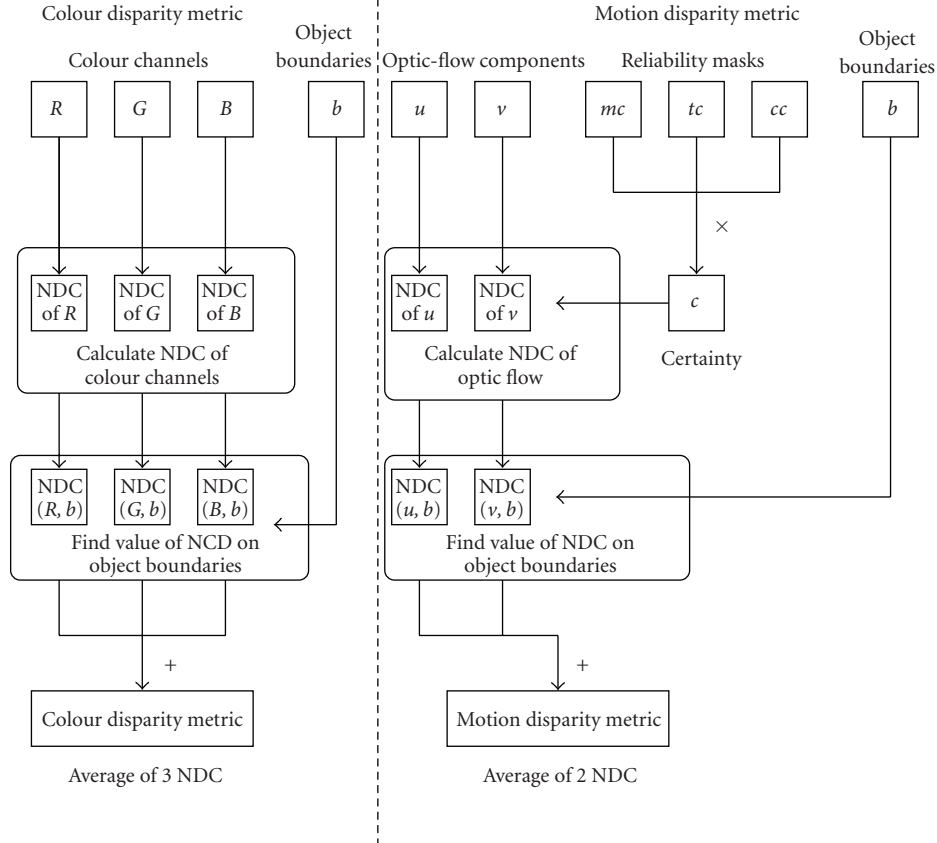
FIGURE 4: Flowchart of proposed method of calculation of disparity metrics with the use of NDC.

map, indicated as $tc$. The texture activity is expressed taking into consideration the following fact. The more distant a point is from an edge, the more difficult it becomes for the motion estimator to find a good match. We therefore calculate an edge map of a given frame and associate to each pixel the Euclidean distance between its own location and the closest edge to it [29]. This matrix, scaled in the range from 0–1, provides what the required texture certainty measure $tc$.

Even in highly textured areas, errors are concentrated in the vicinity of motion boundaries, due to so-called *smoothness constraints* frequently used in motion estimation methodologies. To account for that, a measure of error along motion boundaries can be obtained by assuming that the motion boundary of an object coincides with spatial boundaries. This is a spatio-temporal coherence consideration and it is reflected by the third element of our certainty, denoted as $cc$. In order to calculate the matrix $cc$, we calculate the motion boundaries corresponding to the object to be evaluated using an edge detector on the component of the optic flow. We then calculate the distance between each motion boundary location and the closest colour edge. The colour edges have already been used to produce the $tc$. If the distance at a location of the boundary is bigger than a given threshold $d_T$, then such a location is set to zero in $cc$ and ignored in the calculation of the NDC. All the other motion boundary locations are set to one in $cc$.

The overall certainty map contains a measure of motion reliability, a measure of spatial reliability, and a measure of spatio-temporal reliability. The three elements are combined into a single certainty map $c$ to be used for the calculation of the NDC:

$$c(x, y) = mc(x, y) \cdot tc(x, y) \cdot cc(x, y), \qquad (19)$$

where the operator $\cdot$ indicates point-by-point multiplication. The coherence map $cc$ may be used also to enforce spatio-temporal coherence in the calculation of the colour metric CM.

## 4. EXPERIMENTAL WORK

The results shown in this section were obtained using six standard MPEG test sequences called *Renata*, *Mobile and Calendar*, *Garden*, *Mother and Daughter*, *Foreman*, and *Stefan* [30]. To avoid complications due to interlacing, only even-parity field data were retained.

Renata is a head-and-shoulders sequence, showing a person moving in front of a complex-textured background. The background consists of synthetic textures both in luminance and colour. The sequence presents very low-contrast and very similarly textured areas between background and foreground in some frames. A field from test sequence Renata is shown in **Figure 5(a)**, showing the boundaries of the moving
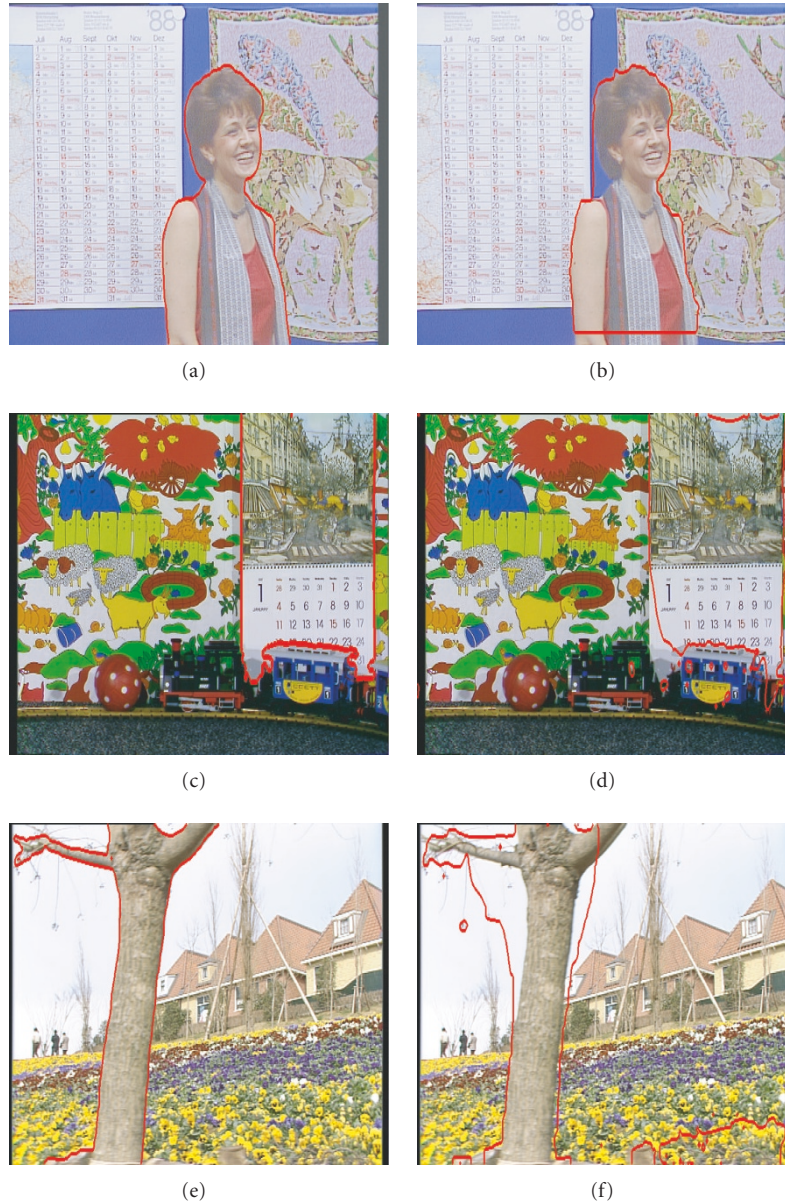
FIGURE 5: (a), (c), (e) Boundaries of manual segmentation of moving object superimposed to the original field and (b), (d), (f) boundaries of erroneous segmentation of the same moving object for test sequences Renata, Mobile and Calendar, and Garden, respectively.

object, manually segmented. In Figure 5(b), an incorrectly segmented video object corresponding to the foreground object is shown.

Mobile and Calendar is a synthetic sequence rich in colour and textures. It presents three main moving objects. In this work, we present only data from the calendar object. The calendar is moving behind the train and in the upper part of the frame, following a roughly vertical direction. There is slight camera panning. A field from test sequence Mobile and Calendar is shown in Figure 5(c), showing the boundaries of the moving object, manually segmented. In Figure 5(d), an incorrectly segmented video object corresponding to the foreground object is shown.

Garden (flower garden) is a natural image rich in texture. Strictly speaking, there is no major object in motion, the movement is apparent, and it depends on the panning of the camera and scene depth. A tree appears to move from the right to the left at a higher speed than the objects further away from the observer. This sequence does not have a high contrast and has very similar textures in parts of the tree trunk and parts of the wooden fences of the surrounding gardens. A field from test sequence Garden is shown in Figure 5(e), showing the boundaries of the moving object, manually segmented. In Figure 5(f), an incorrectly segmented video object corresponding to the foreground object is shown.
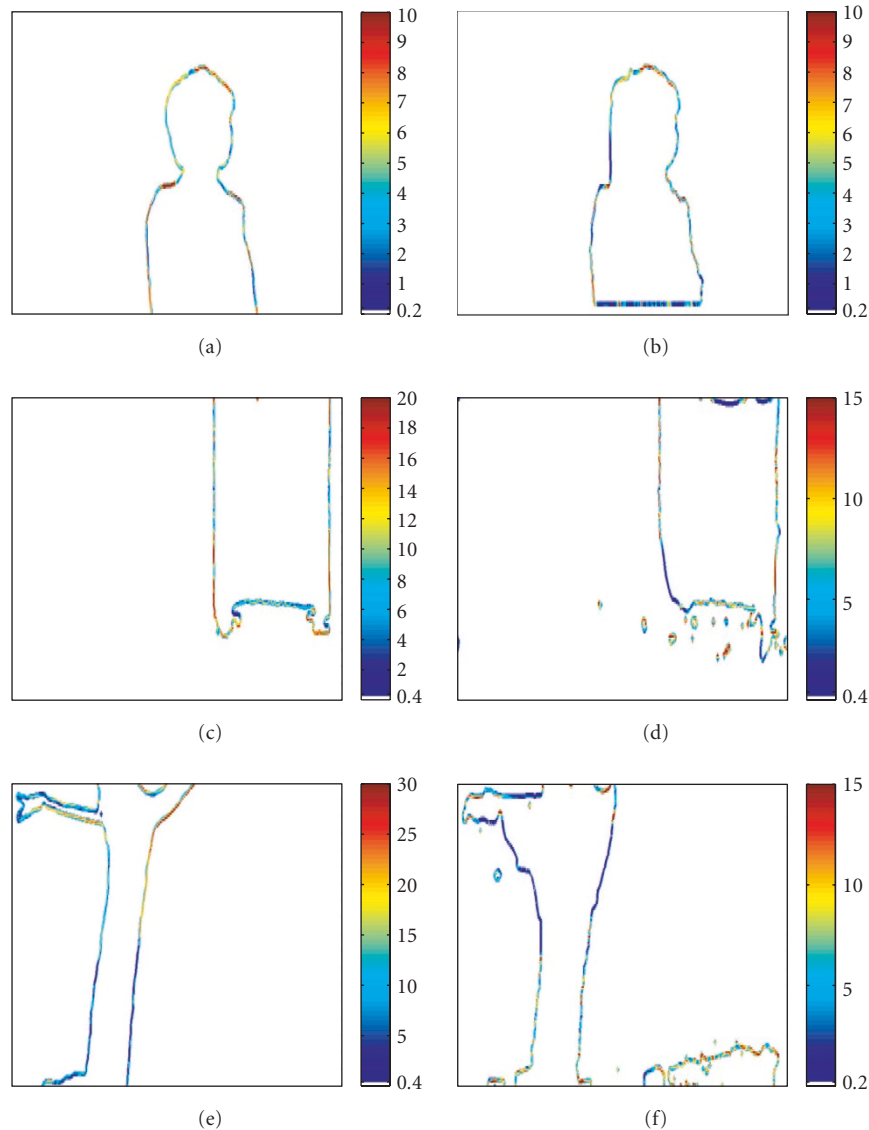
FIGURE 6: Map of intensity of colour contrast along (a), (c), and (e) the boundary of the manually segmented object and (b), (d), and (f) the boundaries of the erroneous object segmentation. The colour bars indicate the magnitude of the contrast in each figure.

Mother and Daughter is a head-and-shoulders sequence. It presents a woman and a young girl talking and moving their heads and hands in front of a simple static background. The colour contrast between background and foreground is low.

Foreman is a head-and-shoulders sequence of a construction worker set against a complex background with low colour contrast.

Stefan is a dynamic sport sequence showing a tennis player against a richly textured background of spectators. As expected, the movement contained in the sequence is very complex.

Manually extracted ground truths and erroneous segmentations have been used in the experiments described below. Examples of ground truths and erroneous segmentations are shown in Figure 5.

### 4.1. Colour disparity metric

The colour disparity metric, CM, is calculated as the value of the NDC computed on the three colour components on the original field, on the position of the boundary taken into account.

We have applied the metric CM to the boundary of both ground truths and erroneous segmentation of video objects moving in the test sequences. The results of such contrast measurement are shown in Figures 6(a), 6(c), and 6(e) for the ground truths and in Figures 6(b), 6(d), and 6(f) for the erroneous segmentations of test sequences Renata, Mobile and Calendar, and Garden. Erroneous parts of the object boundary are consistently signalled for all test sequences by the lowest value of CM. The corresponding values of CM calculated in the corresponding ground truths are much higher.
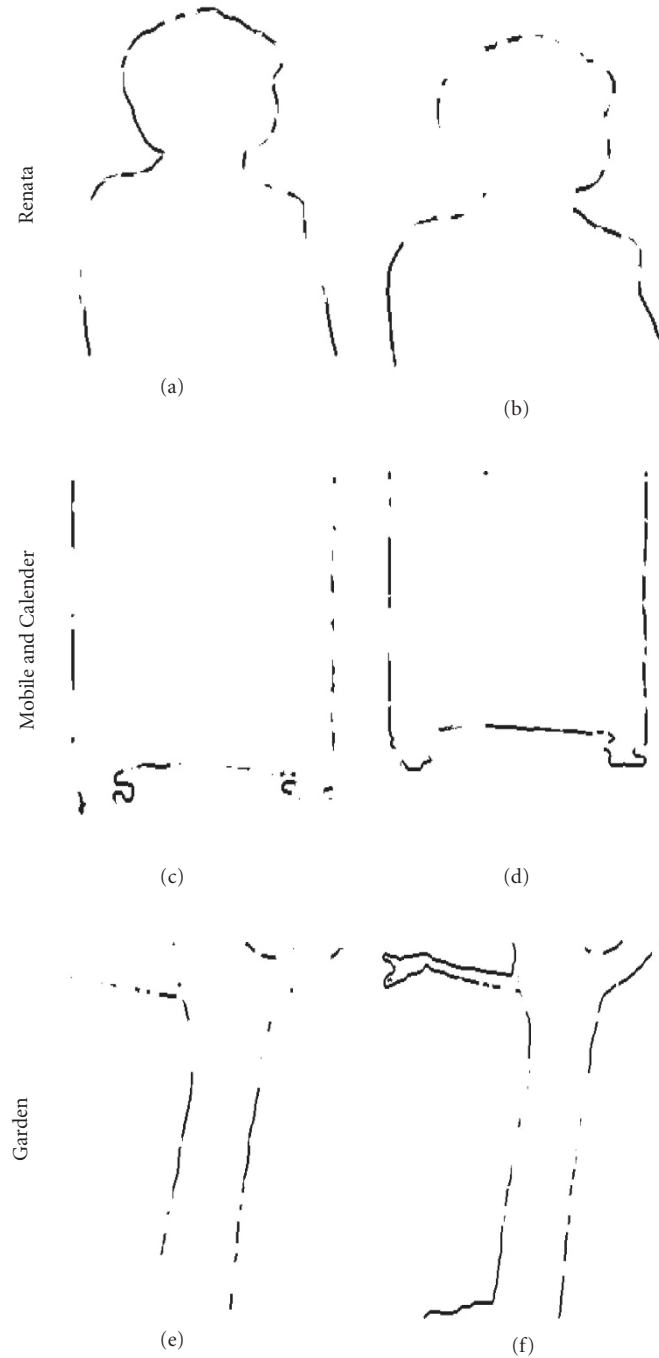
FIGURE 7: (a) Nontextured and (b) textured boundary definition in Renata, Mobile and Calendar, and Garden, respectively.

The most important characteristic of the approach proposed in this paper is the higher sensitivity to a shift in the position of the boundary. Additionally, it is important to verify the influence of noise on the measure, since the proposed method is based on gradient estimation, which tends to be more sensitive to noise, while the approach in [13], which produces the colour disparity metric, $d_C$ (indicated in the diagrams with the legend *Erdem, Tekalp and Sankur*, given the names of the authors of such metric), is based on an average of colour planes.

In order to validate the sensitivity of the method to an incorrect placement of the boundary, the value of CM is calculated for a range of shifts of the motion boundary in the direction of the normal to the boundary at a particular location $p_b$ and compared to $d_C$, calculated on the same boundary points. The boundary is defined by the pixels of the boundary of the manually segmented object. The two contrast measures are normalised with reference to their maximum value, in order to compare them. The sensitivity of the measure is directly proportional to the magnitude of its gradient.

The additional element that needs to be validated is the sensitivity to noise in the image. In order to do so, the boundaries have been divided into two categories: boundaries that lie on a nontextured support, with examples of them shown in Figures 7(a), 7(c), and 7(e) and boundaries that lie on a textured support, with examples of them shown in Figures 7(b), 7(d), and 7(f). The classification into textured and nontextured boundaries is based on [34]. The boundaries that lie on a textured support are expected to suffer from a higher level of noise in the estimation of the gradient.

For the calculation of the contrast measure in [13], a distance $L = 20$ from the boundary and a half range $M = 10$ of the area of calculation of the averages have been used. In order to establish an element of correspondence between the two measures: CM in our work and $d_C$, in [13], we used an applicability function elongated in the direction of the normal to the boundary for the major axis of an ellipse of half length of $R = 30$. Therefore the size of the filter used here is $61 \times 61$ pixels, in order to be comparable to the reference method. In general, the size of the filter depends on the data. The larger the area of missing or uncertain information, the larger the filter. This is because the filter needs to be at least one pixel wider than the largest dimension of the area to be estimated. The speed of the proposed algorithms depends on the size of the filter as well as the resolution of the images. For images of common intermediate format (CIF) resolution $352 \times 288$ pixels and a filter of size $21 \times 21$ pixels, it takes 16 seconds to calculate the disparity metric for each colour channel of the resolution of the frame, with the use of a Matlab-interpreted script on a 433 MHz Intel Celeron CPU. The same considerations apply to the motion disparity metric, in term of filter size and time required for processing a single component of the optic flow.

The contrast metric sensitivity is proportional to the value of the derivative of the disparity metrics, therefore the steeper the descent of the curve representing the metric, the higher the sensitivity. In Figure 8, the comparison of the distortion metrics obtained for all six test sequences are shown, in the case where the object boundary does not lie on a textured support. In Figure 9, the comparison of the distortion metrics obtained for all six test sequences are shown, in the case where the object boundary lies on a textured support. The results obtained using CM are always more sensitive to the presence of the boundary than the ones obtained with the use of $d_C$. The contrast value is oscillating more for CM than $d_C$ in the case of textured boundaries, especially in the cases of Mobile and Calendar and Garden, which contain more texture. However, in the textured regions, the detection of the boundary is clear with CM, while $d_C$ does not differentiate the presence of the object boundary, being almost flat for all values of shift examined.

### 4.2. Motion disparity metric

In Figure 10, the horizontal and vertical components of the optic flow are shown, with the super-imposition of the boundaries of the manually segmented object in case of test sequence Renata. The two components are used in the

calculation of the motion disparity metrics. The motion estimation used here is obtained by a robust motion estimator [28]. This way it is also possible to have a map of motion outliers, shown in Figure 12(a).

In case of the motion measure presented in [13], indicated as $d_M$, the contrast is weighted by the reliability of the motion vectors. In order to implement the reliability measures, the parameters $\sigma_m$ and $\sigma_c$ have been chosen in accordance with the standard deviation of the motion vector and colour planes, respectively. The two components of the reliability measure are shown in Figures 11(a) and 11(b), while their combined effect is shown in Figure 11(c). The weighting scheme proposed here has some disadvantages. In case a motion estimation error occurs in a nontextured area (which is an area where errors in the motion boundary commonly occur), the reliability functions taken into account here do not have any support in order to identify the problem. In Figures 11(a) and 11(b), the errors are shown around the motion boundaries.

In the proposed method, it is possible to distinguish between the signals, that is, the motion estimates, and their *certainty*, which will be used for normalisation of the measure. A robust motion estimator produces a map of the reliability of the estimates, shown in Figure 12(a), where the *outliers* are shown as zeros. This is exactly an example of a certainty map that can be directly used for the purpose of calculating the NDC. The motion outliers will be effectively ignored from the calculation. The information needed at their location is supplied by the local information in a neighbourhood along the normal to the boundary. Moreover, as it is a well-known fact that motion estimators perform poorly in nontextured areas, an additional component of the certainty map is given by the distance of a pixel from textured areas, as shown in Figure 12(b). The rationale for this component of the certainty map is that most motion estimators rely on a neighbourhood search to find a suitable match. With increasing distance from an edge or a textured area, the likelihood of finding a useful reference for motion estimation decreases. We model this dependence directly: the range of the certainty measure goes from 1 to a minimum, $c_{\min} = 1 - d_{\max}$. Here, $d_{\max}$ corresponds to the maximum distance from any textured area. The distance $d$ from the textured areas is scaled in such a way so as to obtain a range of certainty between 1, where an area is textured, and $c_{\min}$, as shown in Figure 12. The third reliability component, shown in Figure 12(c), is a map of the motion boundaries that do not have any correspondence to spatial boundaries, at a distance $d_T = 3$. This is used as an element of spatio-temporal coherence. The three reliability maps are then multiplied together to give the final *certainty* map.

With the proposed method, the motion measure MM is calculated as the average NDC estimated from the horizontal and vertical components of the optic flow, $u$ and $v$, at each point of the boundary $p_b$. In Figure 12(d), the NC of the horizontal flow component, $u$, obtained using the proposed *certainty* map is shown. The boundaries of the manually segmented moving object have been superimposed to give an idea of the shape of the object. The calculation of
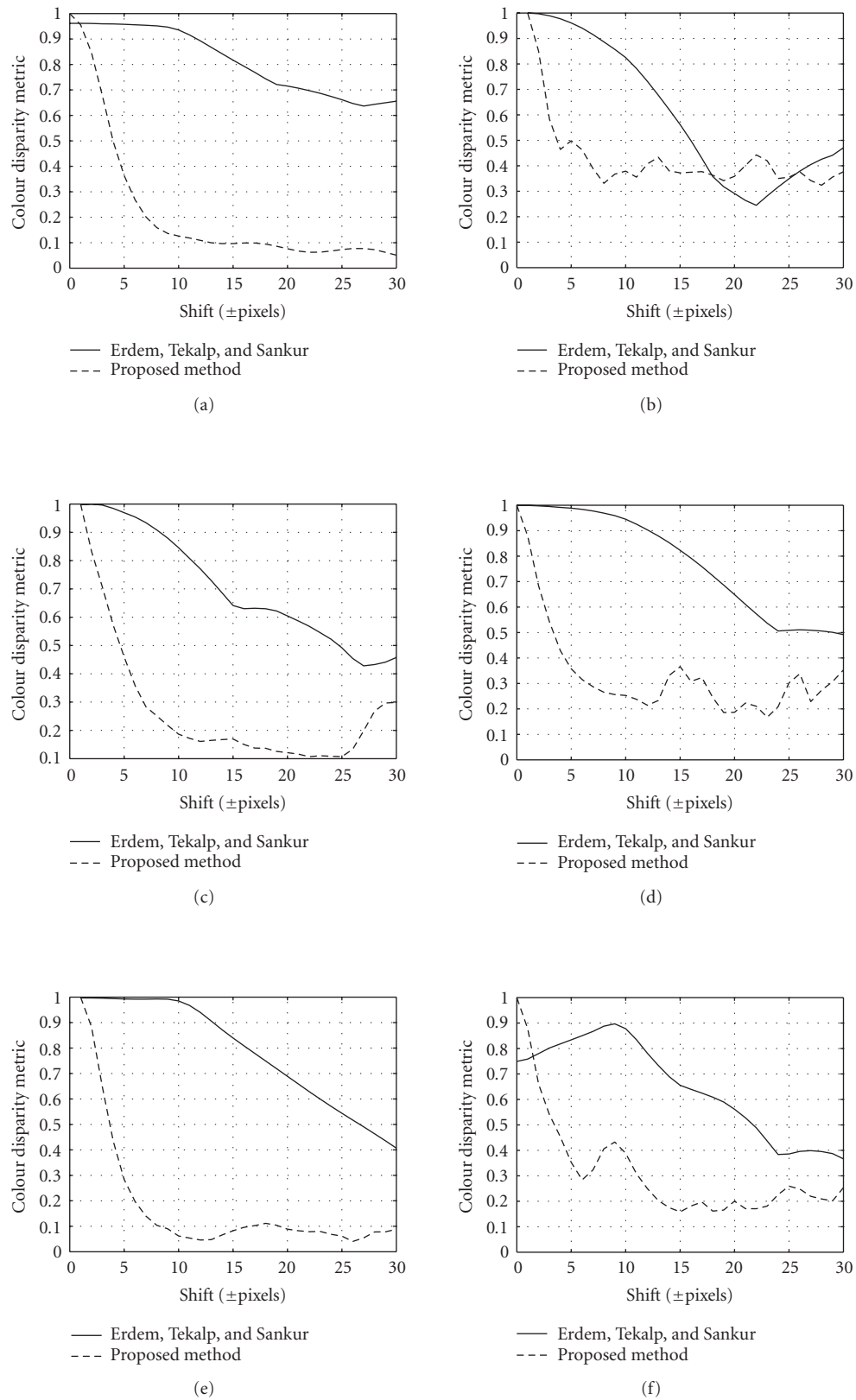
FIGURE 8: Colour disparity metric for nontextured support in 6 MPEG standard test sequences. The sensitivity of the measure is given by the gradient of the metric. The proposed method is shown by the dashed curve. (a) Renata, (b) Mother and Daughter, (c) Mobile and Calendar, (d) Foreman, (e) Garden, and (f) Stefan.
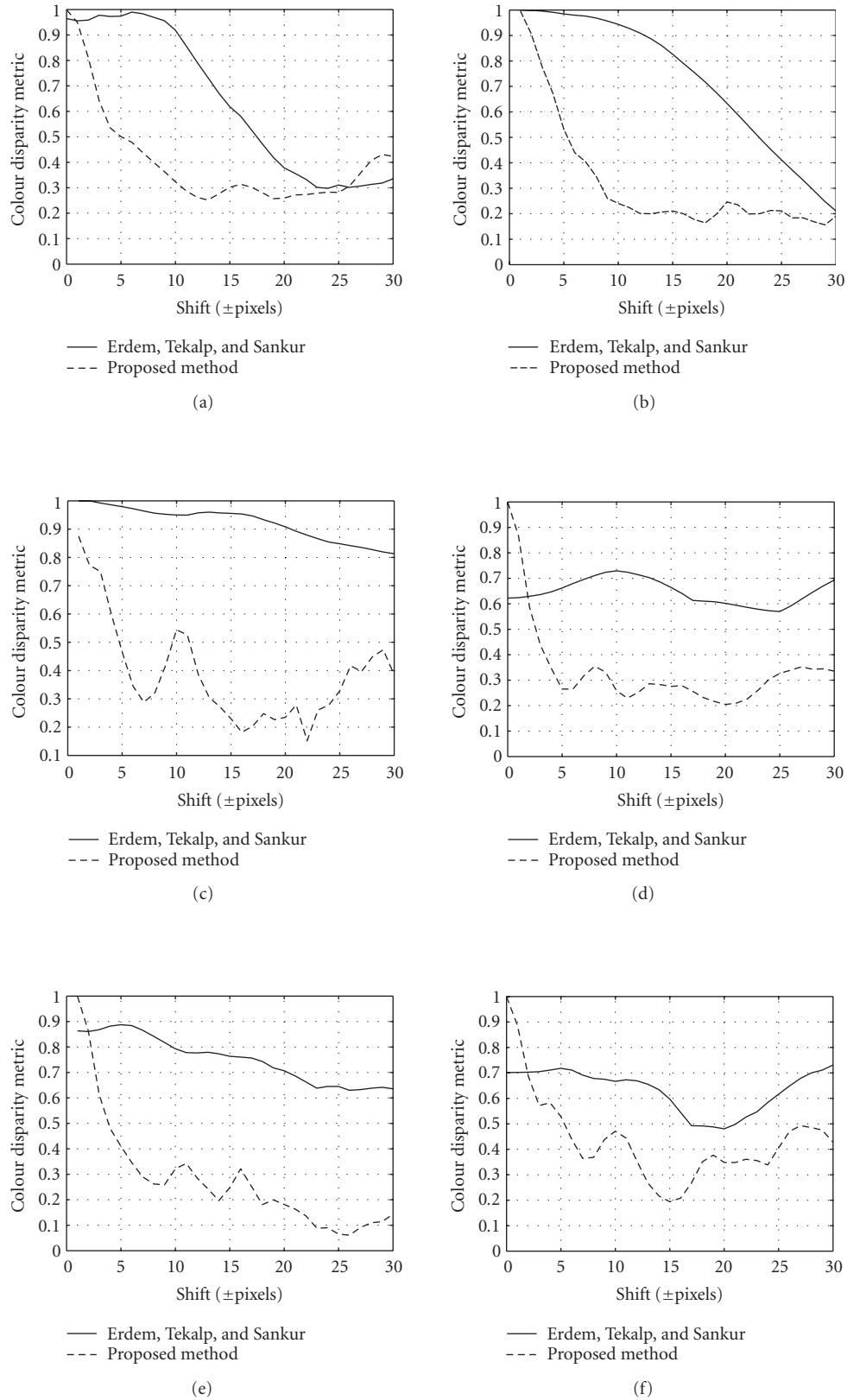
FIGURE 9: Colour disparity metric for textured support in 6 MPEG standard test sequences. The sensitivity of the measure is given by the gradient of the metric. Our proposed method is shown by the dashed curve. (a) Renata, (b) Mother and Daughter, (c) Mobile and Calendar, (d) Foreman, (e) Garden, and (f) Stefan.

(a)                                        (b)

FIGURE 10: Boundaries of the manually segmented object superimposed to (a) the horizontal component of the optic flow and (b) the vertical component of the optic flow.



(a)



(b)
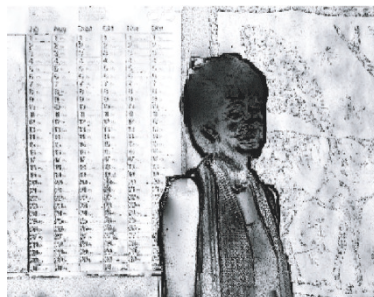


(c)

FIGURE 11: (a), (b) are the two components of the motion reliability measure according to [13], the darker areas are the less reliable areas. In (c), the two elements are combined together, in this case the lighter areas are the more reliable areas.
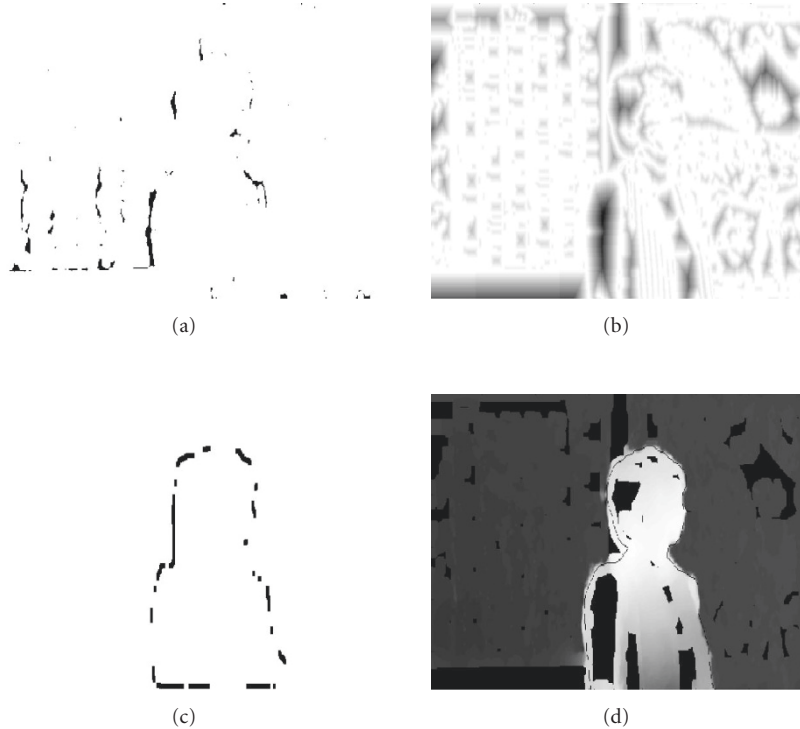
FIGURE 12: (a) Map of motion outliers. (b) Map of distance from textured support, scaled from 1 (on the textured area) to 0 (maximum distance from any textured area for the given field). (c) Map of location of moving object that does not correspond to any colour boundary in the original image. The multiplication of (a), (b), and (c) provides the *certainty map* for the proposed method, while (d) shows the NC obtained with the use of the certainty maps proposed in this method, with superimposed boundaries of the hand-segmented object.

the NC is the first step towards the calculation of the NDC and it gives a clear indication of the transformation that the optic-flow is subjected to as the result of the use of a given certainty map. Comparing this figure with Figure 10(a), the improved correspondence of the optic flow field to the shape of the object along its boundaries is noticeable. Also, the information on inner non-textured regions of the object is lost, because the certainty value associated with these regions in the map of Figure 12(b) is equal or very close to zero. However, this is deliberate because the information at the inner regions of the objects is not relevant for the calculation of MM. In case a specific application would need the information at inner regions, two things could be done: (1) setting the certainty to a value bigger than zero in those areas or (2) using a larger kernel for the applicability function.

In Figure 13, the sensitivity of the motion measure MM is compared to $d_M$. Both measures are calculated along the boundaries of the manually segmented object for the three test sequences. The plotted values are a function of the shift from the correct boundary location, along the normal to the boundary, averaged along all boundary points. The curve obtained using MM is sharper and the maximum has value equal to 1. This means that the measure is much more sensitive and at the same time more accurate towards locating the boundary. The $d_M$ measure never reaches the maximum value of 1, even when the exact boundary, as identified by a human observer, is obtained. Additionally, the plateau shows a lack of sensitivity for the extent of $L$ while there is evidence of sensitivity to noise.

Finally, in Figure 14, the value of the motion measure MM calculated for each point along the boundary of the incorrectly segmented moving objects, for each of the three test sequences, is shown. As in the case of the colour metric CM, the lower values of MM consistently identify for all sequences the presence of an incorrect motion boundary.

### 4.3. Comparative evaluation of spatio-temporal segmentation

In Sections 4.1 and 4.2, we have demonstrated the usefulness and the enhanced sensitivity characteristics of the proposed disparity metric for the evaluation of the quality of object identification on a *local* basis. This means that the evaluation of the erroneous segmentation is based only on the comparison of pixels belonging to one object and one frame. This is useful when a local correction of a generated object is needed.

We further investigate the capability of the proposed metrics to monitor the quality of the segmentation obtained for a given object in each frame of a sequence. This is done by assigning each frame a *global* value of either the colour or motion disparity metric, calculated as the average value along each element of the boundary. We used frames 10–50 of test sequences Mobile and Calendar and Garden.
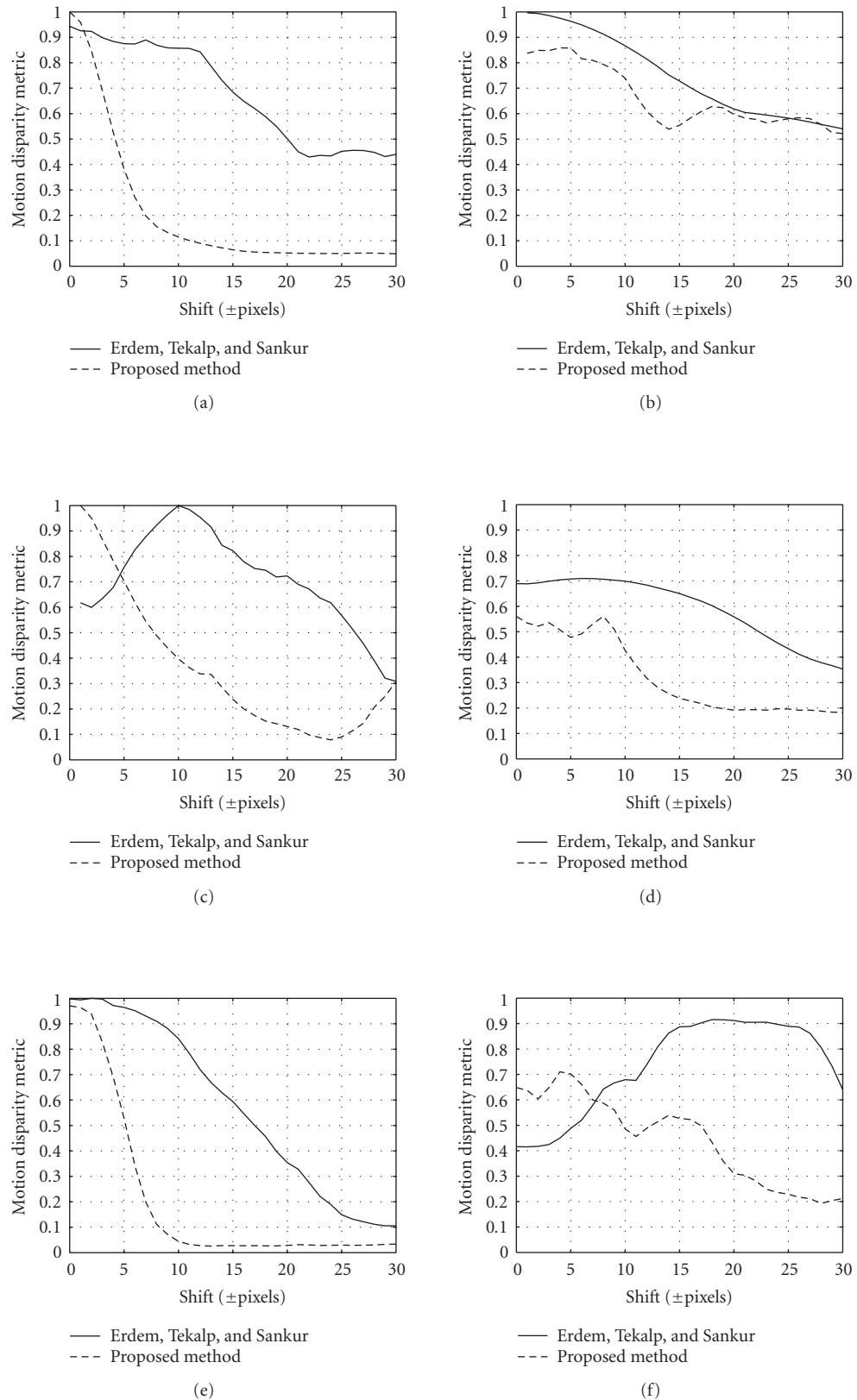
Figure 13: Motion disparity metrics for (a) Renata, (b) Mother and Daughter, (c) Mobile and Calendar, (d) Foreman, (e) Garden, and (f) Stefan. The sensitivity of the measure is given by the gradient of the metric. Our proposed method is shown by the dashed curve.
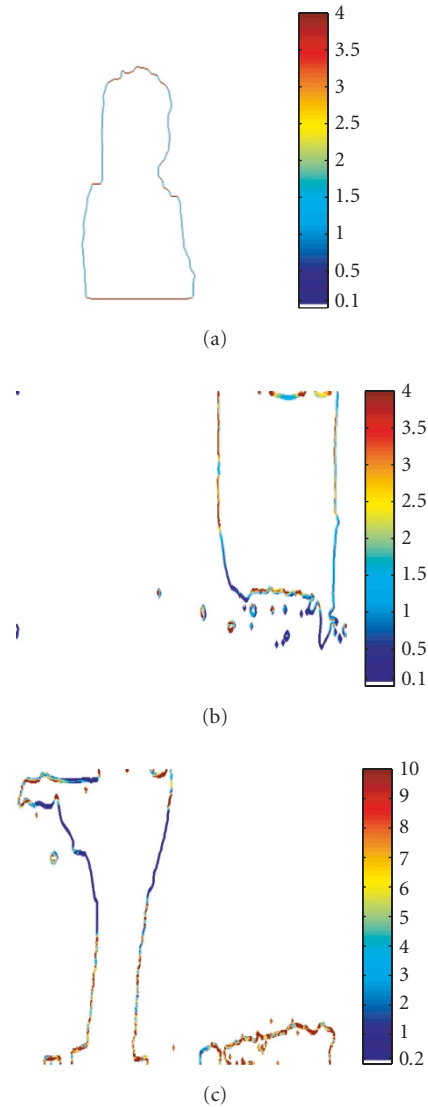
FIGURE 14: Motion disparity metric along the boundaries of the erroneously segmented object for test sequences (a) Renata, (b) Mobile and Calendar, and (c) Garden. The colour bars indicate the magnitude of the contrast in each figure.

Another important use of the evaluation metrics is to compare different spatio-temporal segmentation methods. State-of-the-art spatio-temporal segmentation methods are based mainly on a region-growing paradigm [1]. Since they need to combine spatial and temporal information, they may be classified according to the way this combination is achieved. *Parallel spatio-temporal methods* perform spatial and temporal segmentations separately and then combine the regions formed on the basis of a set of rules. Alternatively, *hierarchical spatio-temporal methods* combine the spatial and temporal information initially using a common similarity measure and they derive regions from it in an iterative fashion. We compare here two methods representative of the two segmentation strategies. The parallel method in [31] is based on a graph-based region-growing method. The hierarchical method in [32] is based on the watershed transform.

Different segmentation methods can be evaluated by monitoring, on a frame-by-frame basis, the value of colour and motion metrics for any object in the sequence. Additionally, it is possible to associate to a segmentation method (for a given sequence), a single figure-of-merit. This can be achieved, as suggested in [14], by summing the colour and motion disparity metrics at each frame and then averaging them over the length of the sequence. We note here that for such a measure to reflect both spatial and temporal quality of the segmentation, the colour and motion features need to be normalised. In this work, we normalised the features with respect to dynamic range, but other methods may be used.

Finally, an important issue of any objective evaluation method is its relevance to subjective quality as perceived by human observers. We present segmentation results obtained for each of the two methods and the two test sequences under

FIGURE 15: Frames 10–50 of test sequence Mobile and Calendar, sampled at regular intervals. The boundaries of the objects generated with (a) parallel [31] and (b) hierarchical spatio-temporal segmentation methods [32] have been blended into the original frames, for the purpose of subjective evaluation [33].

consideration following the recommendation of the COST 211 Quat initiative regarding the presentation of the stimulus for subjective evaluation [33]. Here, instead of blending the original frames to the segmentation masks, we blend the original frames with the boundaries of the segmentation masks, to facilitate the subjective inspection of segmentation quality in Figures 15 and 16.

In Figures 17 and 18, colour and motion disparity metrics are plotted frame-by-frame for two objects of the Mobile and Calendar sequence and for the main moving object in the Garden sequence. The average of the spatial and temporal metrics over all the frames is also shown as a dashed-and-dotted line. This unique value represents the overall evaluation of the methods under consideration. According to this
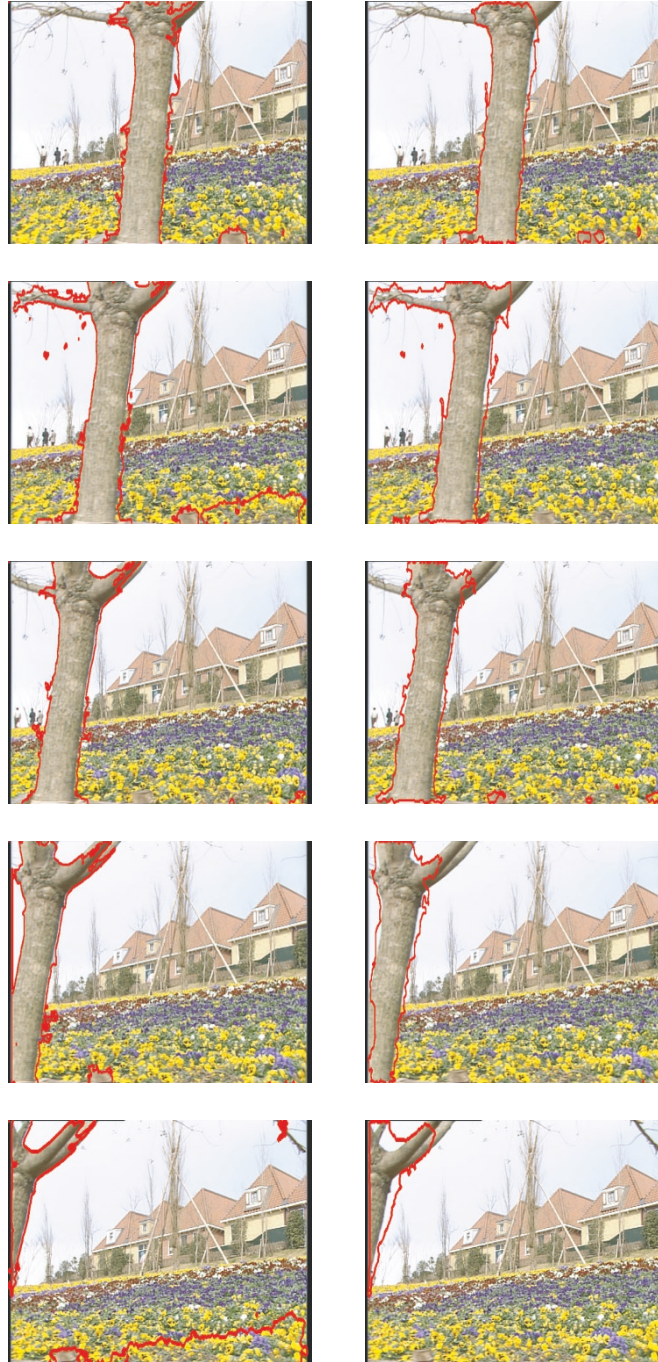
FIGURE 16: Frames 10–50 of test sequence Garden, sampled at regular intervals. The boundaries of the objects generated with (a) parallel [31] and (b) hierarchical spatio-temporal segmentation methods [32] have been blended into the original frames, for the purpose of subjective evaluation [33].

measure in Figure 18, the difference between the two methods is significant in the case of the Garden sequence, where the parallel method performs better. In fact, from a subjective viewpoint, we notice that the closeness of the segmentation boundaries to perceived object boundaries is better for the parallel method. Moreover, a frame-by-frame inspection exposes a fluctuation of the metric according to object segmentation errors resulting in a number of pixels misclassified outside and/or inside of the object. An example of that is provided by the outline of the toy train in Mobile and Calendar for the parallel method in Figure 17. Here a number of pixels are erroneously attached to the object under consideration rendering the plotted metric fairly variable from one frame to the next.
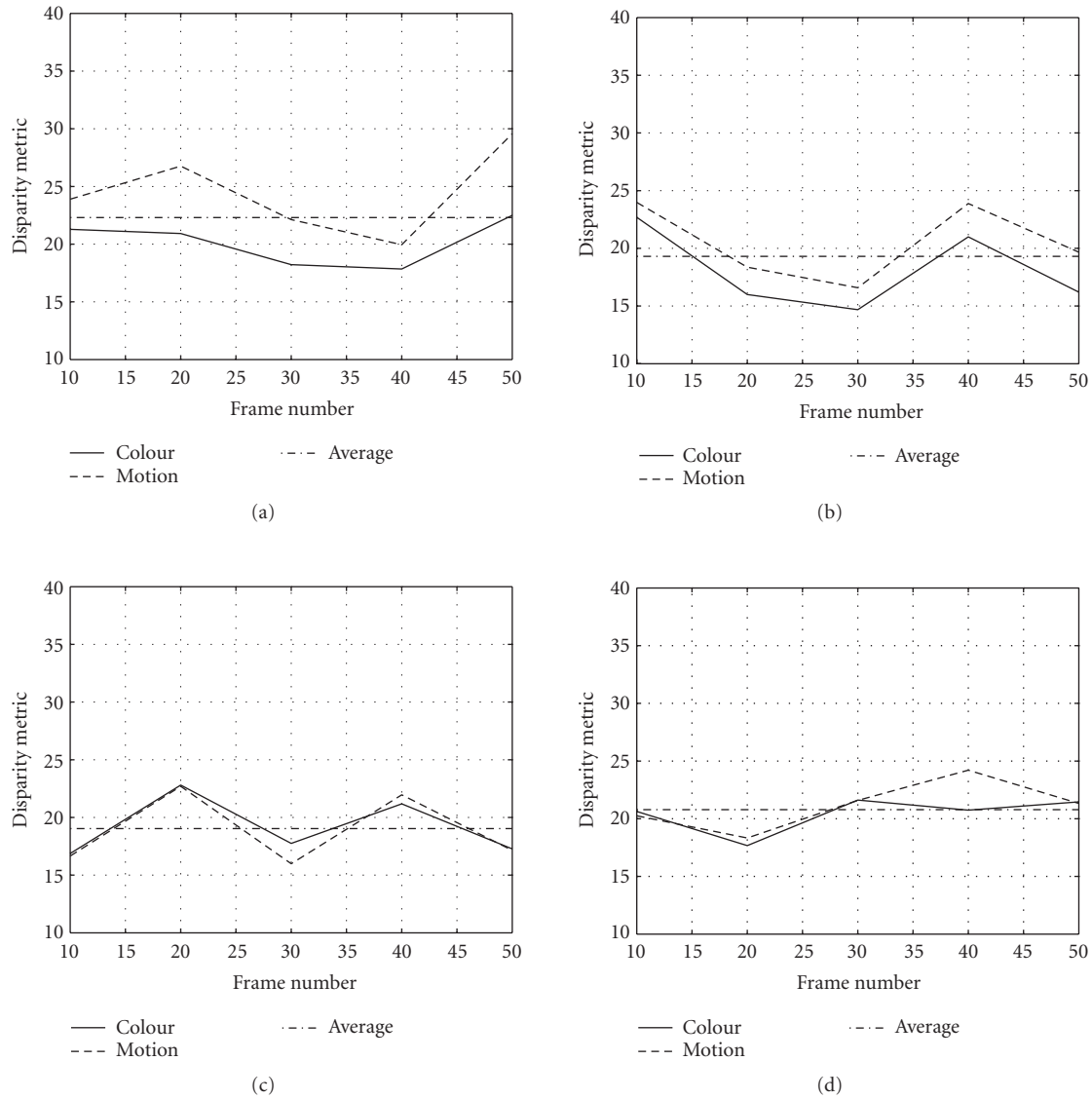
FIGURE 17: Comparative evaluation of two spatio-temporal segmentation methods using the disparity metrics proposed here, calculated on two objects of the test sequence Mobile and Calendar (a) Parallel method (object: Calendar), (b) hierarchical method (object: Calendar), (c) parallel method (object: train), and (d) hierarchical method (object: train).

Given the effectiveness of these measures, one might ask whether it would be possible to use them to drive segmentation algorithms in the first place, rather than just employ them retroactively for evaluation purposes. For straightforward spatio-temporal segmentation, we have already noted that the vast majority of methods are region-based methods. This is because the main aim is the creation of meaningful video objects, often in the shape of regions. In the proposed method, the metrics are calculated on a boundary basis, therefore edges are targeted rather than regions. For this reason, one might only consider the proposed evaluation technique as complementary to segmentation techniques used to produce the video objects in the first place. For example, we can envisage the use of our metrics in a two-stage process,

where a video object previously identified using conventional region-based segmentation, is refined locally with boundary and neighbourhood information, as discussed in Sections 4.1 and 4.2.

## 5. CONCLUSIONS

In this paper, we have presented a unified method for single-stimulus quality assessment of segmented video. According to this method colour and motion features of a moving sequence are taken into consideration and their changes across segment boundaries are monitored. Features are estimated using a local neighbourhood which preserves the topological integrity of segment boundaries. Furthermore the proposed
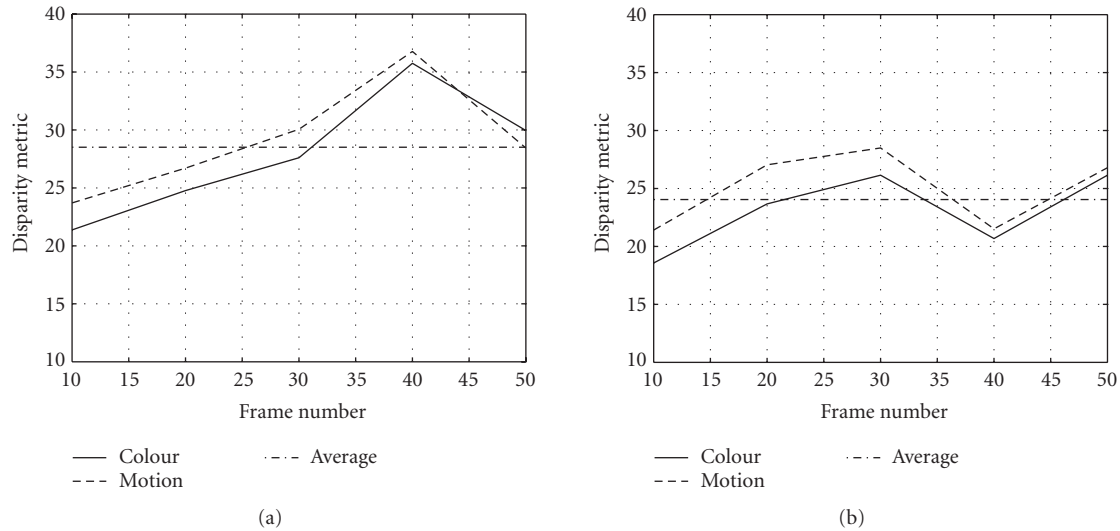
FIGURE 18: Comparative evaluation of two spatio-temporal segmentation methods using the disparity metrics proposed here, calculated on one object of the test sequence Garden. (a) Parallel method. (b) Hierarchical method.

method addresses the problem of unreliable and/or unavailable feature estimates by applying the normalized differential convolution (NDC). Our experimental results have suggested that the proposed method outperforms competing methods in terms of sensitivity as well as noise immunity for a variety of standard test sequences.

## REFERENCES

[1] P. Salembier and F. Marques, "Region-based representations of image and video: segmentation tools for multimedia services," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1147–1169, 1999.

[2] D. Zhang and G. Lu, "Segmentation of moving objects in image sequence: a review," *Circuits Systems and Signal Processing*, vol. 20, no. 2, pp. 143–183, 2001.

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

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

[5] X. Zhang and B. A. Wandell, "Color image fidelity metrics evaluated using image distortion maps," *Signal Processing*, vol. 70, no. 3, pp. 201–214, 1998.

[6] A. M. van Dijk and J.-B. Martens, "Subjective quality assessment of compressed images," *Signal Processing*, vol. 58, no. 3, pp. 235–252, 1997.

[7] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seunties, "A survey of perceptual quality issues in three-dimensional television systems," in *Stereoscopic Displays and Virtual Reality Systems X*, vol. 5006 of *Proceedings of SPIE*, pp. 313–326, Santa Clara, Calif, USA, January 2003.

[8] International Telecommunication Union ITU.R BT.500-11: *Methodology for the subjective assessment of the quality of television pictures*, 2002.

[9] International Telecommunication Union. ITU.R BT.710-4: *Subjective assessment methods for image quality in high-definition television*, 1998.

[10] M. H. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," in *Visual Communications and Image Processing (VCIP '03)*, vol. 5150 of *Proceedings of SPIE*, pp. 573–582, Lugano, Switzerland, July 2003.

[11] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntiens, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 381–391, 2004.

[12] A. Cavallaro and T. Ebrahimi, "Object-based video: extraction tools, evaluation metrics, and applications," in *Visual Communications and Image Processing (VCIP '03)*, vol. 5150 of *Proceedings of SPIE*, pp. 1–8, Lugano, Switzerland, July 2003.

[13] C. E. Erdem, B. Sankur, and A. M. Tekalp, "Performance measures for video object segmentation and tracking," *IEEE Trans. Image Processing*, vol. 13, no. 7, pp. 937–951, 2004.

[14] C. E. Erdem and B. Sankur, "Performance evaluation metrics for object-based video segmentation," in *Proc. 11th European Signal Processing Conference (EUSIPCO '02)*, vol. 2, pp. 917–920, Toulouse, France, September 2002.

[15] P. Correia and F. Pereira, "Objective evaluation of relative segmentation quality," in *Proc. IEEE International Conference on Image Processing (ICIP '00)*, vol. 1, pp. 308–311, Vancouver, BC, Canada, September 2000.

[16] A. Cavallaro, E. D. Gelasca, and T. Ebrahimi, "Objective evaluation of segmentation quality using spatio-temporal context," in *Proc. IEEE International Conference on Image Processing (ICIP '02)*, vol. 3, pp. 301–304, Rochester, NY, USA, September 2002.

[17] P. Villegas and X. Marichal, "Perceptually-weighted evaluation criteria for segmentation masks in video sequences," *IEEE Trans. Image Processing*, vol. 13, no. 8, pp. 1092–1103, 2004.

[18] P. L. Correia and F. Pereira, "Objective evaluation of video segmentation quality," *IEEE Trans. Image Processing*, vol. 12, no. 2, pp. 186–200, 2003.

[19] P. L. Correia and F. Pereira, "Classification of video segmentation application scenarios," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 5, pp. 735–741, 2004.

[20] C. E. Erdem, A. M. Tekalp, and B. Sankur, "Metrics for performance evaluation of video object segmentation and tracking without ground-truth," in *Proc. IEEE International Conference on Image Processing (ICIP '01)*, vol. 2, pp. 69–72, Thessaloniki, Greece, October 2001.

[21] H. Knutsson and C.-F. Westin, "Normalized and differential convolution: methods for interpolation and filtering of incomplete and uncertain data," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR '93)*, pp. 515–523, New York, NY, USA, June 1993.

[22] C.-F. Westin, K. Nordberg, and H. Knutsson, "On the equivalence of normalized convolution and normalized differential convolution," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '94)*, vol. 5, pp. 457–460, Adelaide, SA, Australia, April 1994.

[23] C.-F. Westin and H. Knutsson, "Processing incomplete and uncertain data using subspace methods," in *Proc. 12th IAPR International Conference on Pattern Recognition*, vol. 3, pp. 171–173, Jerusalem, Israel, October 1994.

[24] R. Piroddi and M. Petrou, "Dealing with irregular samples," in *Advances in Imaging and Electron Physics*, P. W. Hawkes, Ed., vol. 132, pp. 109–165, Elsevier, Amsterdam, The Netherlands, 2004.

[25] M. Petrou, R. Piroddi, and S. Chandra, "Irregularly Sampled Scenes," in *Image and Signal Processing for Remote Sensing X*, vol. 5573 of *Proceedings of SPIE*, pp. 319–333, Maspalomas, Gran Canaria, Spain, September 2004.

[26] T. Q. Pham and L J. van Vliet, "Normalized averaging using adaptive applicability functions with applications in image reconstruction from sparsely and randomly sampled data," in *Proc. 13th Scandinavian Conference on Image Analysis (SCIA '03)*, vol. 2749 of *Lecture Notes in Computer Science*, pp. 485–492, Halmstad, Sweden, June–July 2003.

[27] B. Rieger and L. J. van Vliet, "Curvature of *n*-dimensional space curves in grey-value images," *IEEE Trans. Image Processing*, vol. 11, no. 7, pp. 738–745, 2002.

[28] M. J. Black, D. J. Fleet, and Y. Yacoob, "Robustly estimating changes in image appearance," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 8–31, 2000.

[29] R. Piroddi, *Multiple-feature object-based segmentation of video sequences*, Centre for Vision, Speech and Signal Processing, University of Surrey, 2004.

[30] R. Koenen, *From MPEG-1 to MPEG-21: creating an interoperable multimedia infrastructure*, International Organisation for Standardisation—Organisation Internationale de Normalisation ISO/IEC JTC1/SC29/WG11 (Coding of Moving Pictures and audio), 2001.

[31] A. A. Alatan, E Tuncel, and L. Onural, "A rule-based method for object segmentation in video sequences," in *Proc. IEEE International Conference on Image Processing (ICIP '97)*, vol. 2, pp. 522–525, Santa Barbara, Calif, USA, October 1997.

[32] J. G. Choi, S.-W. Lee, and S.-D. Kim, "Spatio-temporal video segmentation using a joint similarity measure," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 279–286, 1997.

[33] A. 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 Syst. Video Technol.*, vol. 8, no. 7, pp. 802–813, 1998.

[34] R. Piroddi and T. Vlachos, "Multiple-feature segmentation of moving sequences using a rule-based approach," in *Proc. 13th British Machine Vision Conference (BMVC '02)*, vol. 1, pp. 353–362, Cardiff , UK, September 2002.

---

**R. Piroddi** received a Laurea degree in electronic engineering from the University of Cagliari, Italy, in 1999. She was awarded a Ph.D. degree in electronic engineering from the University of Surrey, UK, in 2004 for her work on object-based segmentation of video sequences. From October 2002 to August 2005, she worked on irregularly sampled signal and image processing as a Research Fellow at the Centre for Vision, Speech and Signal Processing, University of Surrey, UK. Since September 2005, she has been a Research Fellow in the Department of Electrical and Electronic Engineering, Imperial College London, UK, working on biologically inspired computer vision algorithms. Her research interests include image/signal processing, with emphasis on applications to medical imaging, geoscience and remote sensing, object-based video processing and compression, cognitive vision, biologically-motivated computer vision paradigms and information representation. She is the author of 10 conference and journal articles.

**T. Vlachos** received a Dipl.-Ing degree from the University of Patras, Greece, in 1985 and the M.S. degree from the University of Maryland, USA, in 1986 both in electrical engineering. For his work on image and video coding, he was awarded the Ph.D. degree from Imperial College in 1993. From 1985 to 1987, he held research positions at the University of Maryland and the Institute for Systems Research working on digital communication systems and networks. Between 1988 and 1992, he was a European Commission Fellow at Imperial College and was associated with Philips Research Laboratories, UK, working on image analysis, image processing, and video coding for very low bit rate and broadcasting applications. From 1993 to 1997, he was a Research Engineer at the BBC R&D Department where he led various projects on bit-rate reduction for digital HDTV and archive restoration. He joined CVSSP at the University of Surrey in 1997, where he is now a Senior Lecturer in multimedia signal processing. He is a Chartered Engineer, and a Member of the Technical Chamber of Greece and the IEE. Current research interests are in the areas of video compression, motion estimation, and archive restoration.