Published in: Proc. Visual Communications and Image Processing (VCIP), Beijing, China, Vol. 5960, pp. 111-120, 12-15 July 2005.

Model based Segmentation of Motion Fields in Compressed Video Sequences using Partition Projection and Relaxation

Siripong Treetasanatavorn*[a], Uwe Rauschenbach[b], Jörg Heuer[b], and André Kaup[a]

[a]University of Erlangen-Nuremberg, Chair of Multimedia Communications and Signal Processing, Cauerstraße 7, D-91058 Erlangen, Germany;
[b]Siemens AG, CT IC 2, Otto-Hahn-Ring 6, D-81739 Munich, Germany

ABSTRACT

In the context of visual signal analysis for media adaptation, this paper presents a stochastic method for segmentation of motion vector fields in compressed videos. The video sequence is analysed and partitioned into temporally associated, motion-semantic regions by three analysis steps. The global motion based reliability analysis assesses the reliability extent of each encoded displacement vector. Upon this measure array, an initial partition is approximated by the region displacement prediction. The subsequent relaxation procedure optimises the partition shape, coverage, and location using the *stochastic motion coherency* model and the second-order random field based contour smoothness. The visually convincing results are demonstrated from the standard sequences.

Keywords: Compressed video analysis for media adaptation, Stochastic motion segmentation, Stochastic object tracking

1. INTRODUCTION

The past years have witnessed the growing demands for multimedia data at different forms, qualities, and characteristics. This is due to the expanding range of terminal device capabilities as well as usage situations and preferences. Particularly in the video messaging scenario,[1] the future services require video adaptation to effectively scale with an increasing number of terminal classes, configurations, and usage contexts under a limited system resource. However, a success of the video adaptation strongly depends on the comprehensibility of the adapted video presentations perceived by end users. This prerequisite is naturally fulfilled by incorporating the video content analysis, such that the syntactic and semantic video content is retrieved and employed in the video adaptation as far as possible.

In the addressed context, this contribution presents the displacement field analysis and segmentation method for the compressed video sequence. This work extends the initial analysis model[2] for the segmentation of displacement field sequence into motion-coherent regions, and the corresponding temporal association (i.e., tracking) of the inter-frame regions. The method assesses the reliability extent of the displacement vectors by applying the global motion based motion coherency analysis. This process ensures that only reliable displacement vectors shall be utilised in the segmentation. Based on this result, the first displacement field is segmented using a single-field optimisation method.[2] The algorithm estimates a set of region motion parameters, and projects the current partition to the subsequent frame. Through this process, a rough region constellation of the next frame is initialised using the current segmentation result. The stochastic relaxation adapts the predicted partition based on new displacement field statistics. As such, the final partition shall exhibit the statistical optimality in terms of motion coherency and contour smoothness. This projection and relaxation procedure is iterated for the subsequent frames until the end of sequence. It is noted that the outlined video analysis method is suitable for a video analysis scenario constrained by low computational complexity and no-user interaction.[3]

*Send correspondence to Siripong Treetasanatavorn (siripongtr@ieee.org) or André Kaup (kaup@LNT.de). The work appeared on this paper was developed during his industrial appointment at Siemens AG (Corporate Technology, CT IC 2 Network and Multimedia Communications, Munich, Germany) under Siemens *Youth and Knowledge* Doctoral Research Fellowship Grant.

Published in: Proc. Visual Communications and Image Processing (VCIP), Beijing, China, Vol. 5960, pp. 111-120, 12-15 July 2005.

A number of methods for video segmentation and tracking were present in the literature. As displacement fields from the compressed videos do not necessarily obey true semantics on the captured scenes, arrays of confidence measures are usually estimated in the preprocessing step[4] that is followed by field classification or segmentation, and partition refinement.[5,6] A number of statistical approaches gain more and more popularity due to robustness to uncertainty of the acquired video data.[7,8] Further literature in video segmentation can be found in Ref. 9, 10.

On video tracking, most techniques require an initial process selecting appropriate visual features[11,12] that is followed by analysis of object contours,[13,14] tracked edges,[15] or, most recently, region motion trajectory.[16] A fusion of multiple sources such as motion and color was reported as a robust segmentation-tracking solution.[17] Since high-level visual semantics perceived by humans are difficult to analyse and extract solely by machine, another important research track applies cues from human-computer interaction as additional analysis inputs to the segmentation and tracking algorithm.[18]

The paper is organised as follows. Sect. 2 introduces important terminologies for the model and algorithm derivation. Sect. 3 presents the basis analysis model, the *stochastic motion coherecy*. Sect. 4 discusses the optimisation algorithm that applies the global motion-based reliability assessment, as well as the partition projection and relaxation method. Sect. 5 presents the experimental results. Sect. 6 concludes the paper.

2. NOMENCLATURE

$\mathbf{x}$	a two-dimensional (2-D) Cartesian displacement vector coordinate on the field lattice structure
$\mathbf{v}(\mathbf{x})$	a 2-D displacement vector at coordinate $\mathbf{x}$
$\mathbf{v'}(\mathbf{x},k)$	a 2-D modelled displacement vector at coordinate $\mathbf{x}$, given k-th displacement model
$\hat{\mathbf{v}}(\mathbf{x}\|\mathbf{n},k)$	a 2-D displacement vector predictor given neighbour $\mathbf{v}(\mathbf{n})$ and k-th displacement model
Ψ_k	the k-th vector membership set of region ψ_k, i.e., the partitioning result
$\mathbf{T}_k$	the k-th 2-D displacement model of region ψ_k
$\mathbf{V}_k$	the k-th 2-D velocity (first-order derivative of displacement) model of region ψ_k
$\mathbf{A}_k$	the k-th 2-D acceleration (second-order derivative of displacement) model of region ψ_k
$\mathbf{S}_{k,j}$	the projection estimation of the k-th region at time τ_j

3. STOCHASTIC MOTION COHERENCY MODEL

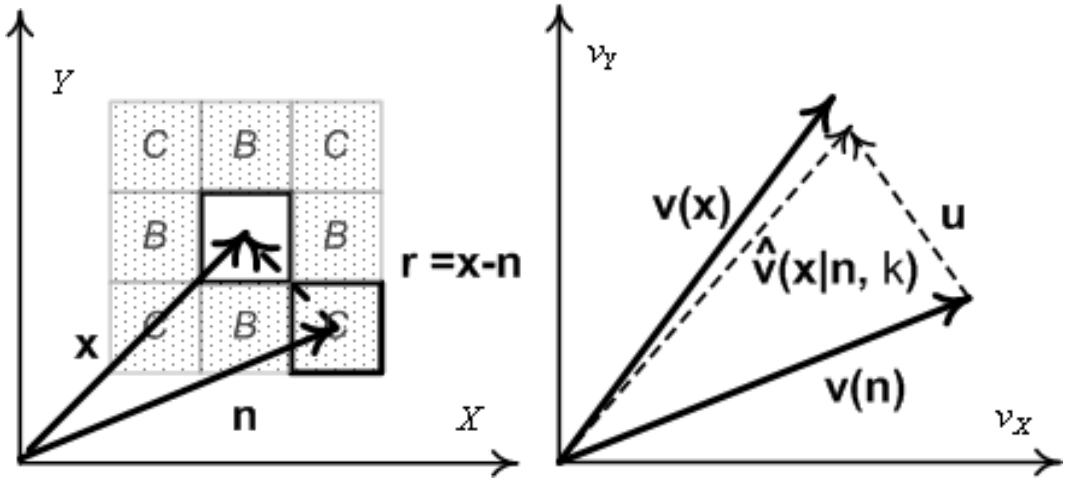
The basis model[2] evaluates the probability density $\Pr(\mathcal{Q}|\mathcal{V})$ of partition $\mathcal{Q}$, consisting of λ regions, given an observed displacement field $\mathcal{V}$. Through the use of Bayes rule and *maximum a posteriori* (MAP) estimation,[19] the probability $\Pr(\mathcal{Q}|\mathcal{V})$ can be expressed by the multiplication of local/region motion coherency likelihood $\Pr(\mathcal{V}|\mathcal{Q})$ and the *a priori* region boundary density $\Pr(\mathcal{Q})$:

$$\Pr(\mathcal{Q}|\mathcal{V}) \propto \Pr(\mathcal{V}|\mathcal{Q}) \cdot \Pr(\mathcal{Q}) = \underbrace{\Pi_\alpha(\mathcal{V}|\mathcal{Q}) \cdot \Pi_\beta(\mathcal{V}|\mathcal{Q})}_{Likelihood} \cdot \underbrace{\Pi_\chi(\mathcal{Q})}_{Priori}, \tag{1}$$

which can be written in a comprehensive form as:

$$\Pr(\mathcal{Q}|\mathcal{V}) \propto \frac{1}{Z} \cdot \underbrace{\exp\left[-\sum_{k=1}^{\lambda}\left\{G_k \cdot \sum_{\mathbf{x}\in\Psi_k}\Delta_\alpha(\mathbf{x},k)\right\}\right]}_{Local\ Motion\ Coherency} \cdot \underbrace{\exp\left[-\sum_{k=1}^{\lambda}\left\{H_k \cdot \sum_{\mathbf{x}\in\Psi_k}\Delta_\beta(\mathbf{x},k)\right\}\right]}_{Region\ Motion\ Coherency} \cdot \underbrace{\exp\left[-\mathcal{N}_B B - \mathcal{N}_C C\right]}_{A\ Priori\ Density}. \tag{2}$$

The likelihood assesses the motion coherency probability at two levels. At the neighborhood level the local smoothness is ensured through the $\Delta_\alpha(\mathbf{x},k)$ observation (cf. Eq. (4)) at each vector coordinate $\mathbf{x}$ and the k-th local coefficient G_k. At the region level, the model guarantees that each vector fits to the assigned region model through the region incoherence $\Delta_\beta(\mathbf{x},k)$ (cf. Eq. (11)) and the k-th region coefficient G_k. The affine displacement model[20] $\mathbf{T}_k$ of the region Ψ_k, $k = 1,\ldots,\lambda$ describes a displacement vector at the coordinate $\mathbf{x}$ by $\mathbf{v'}(\mathbf{x},k) = \mathbf{M}_k\mathbf{x} + \mathbf{t}_k$, with $\mathbf{M}_k \in \Re^{2x2}$ and $\mathbf{t}_k \in \Re^2$ being the region model parameters. The third multiplicand represents the *a priori* density of region borders. This function



(a) The second-order neighbourhood system $\mathcal{G}(\mathbf{x})$ of principal displacement coordinate $\mathbf{x}$ plotted in the context of *relative shift vector* $\mathbf{r}$ based on neighbour coordinate $\mathbf{n}$

(b) The analysis of predictor $\hat{\mathbf{v}}(\mathbf{x}|\mathbf{n},k)$ via *displacement vector justifier* u that rectifies the spatial deviance between $\mathbf{v}(\mathbf{x})$ and neighbour $\mathbf{v}(\mathbf{n})$ using the k-th region model $\mathbf{T}_k$ (cf. (6)-(9))

Figure 1. Illustration of the second-order neighbourhood system and the analysis of displacement vector predictor

measures the partition state by counting a number of vector pairs at the region borders, $\mathcal{N}_B$ in the horizontal or vertical directions and $\mathcal{N}_C$ for the diagonal with B and C being the corresponding coefficients. The parameter Z is a common normalisation constant of these three multiplicands.

3.1. The Likelihood

The first multiplicand $\Pi_\alpha(\mathcal{V}|\mathcal{Q})$ defines the *local motion coherency* of the partition $\mathcal{Q}$ with respect to the displacement field $\mathcal{V}$. The coherency estimation is proceeded within the second-order neighbourhood system $\mathcal{G}(\mathbf{x})$ or the eight nearest neighbours of the Gibbs-Markov random field[19, 21, 22] (cf. Fig. 1(a)),

$$\Pi_\alpha(\mathcal{V}|\mathcal{Q}) = \frac{1}{Z_\alpha}\exp\left[-\sum_{k=1}^{\lambda}\left\{G_k \cdot \sum_{\mathbf{x}\in\Psi_k}\Delta_\alpha(\mathbf{x},k)\right\}\right], \tag{3}$$

with Z_α being a normalisation constant and $\Delta_\alpha(\mathbf{x},k)$ a *local incoherence function* indicating the median of prediction errors $\delta(\mathbf{x},\mathbf{n},k)$ between a principal displacement vector $\mathbf{v}(\mathbf{x})$ and its predictors $\hat{\mathbf{v}}(\mathbf{x}|\mathbf{n},k)$. It is observable that the local motion coherency favours small prediction errors being estimated by the local incoherence function in the neighbourhood. That means, the lower the local incoherence, the higher the local motion coherency likelihood shall be. The local incoherence function $\Delta_\alpha(\mathbf{x},k)$ is expressed as:

$$\Delta_\alpha(\mathbf{x},k) = \textbf{median}_{\,\mathbf{n}\in\,\mathcal{G}(\mathbf{x})\cap\Psi_k}\left\{\delta(\mathbf{x},\mathbf{n},k)\right\}, \tag{4}$$

with $\mathcal{G}(\mathbf{x})\cap\Psi_k$ being a coordinate set in the region Ψ_k that overlaps neighbourhood system $\mathcal{G}(\mathbf{x})$ of the principal displacement vector $\mathbf{v}(\mathbf{x})$. The prediction error $\delta(\mathbf{x},\mathbf{n},k)$ is calculated at each vector component in relation to the predictor $\hat{\mathbf{v}}(\mathbf{x}|\mathbf{n},k)$ by

$$\delta(\mathbf{x},\mathbf{n},k) = \left|v_X(\mathbf{x}) - \hat{v}_X(\mathbf{x}|\mathbf{n},k)\right| + \left|v_Y(\mathbf{x}) - \hat{v}_Y(\mathbf{x}|\mathbf{n},k)\right|. \tag{5}$$

The predictor $\hat{\mathbf{v}}(\mathbf{x}|\mathbf{n},k)$ of the principal displacement vector $\mathbf{v}(\mathbf{x})$ is calculated based on the neighbour displacement vector $\mathbf{v}(\mathbf{n})$, whose coordinate is related to the principal coordinate $\mathbf{x}$ by a relative shift vector $\mathbf{r}$. Fig. 1(a) depicts that the neighbour coordinate $\mathbf{n}$ can be written in terms of its related shift vector $\mathbf{r} = \mathbf{x} - \mathbf{n}$. In the second-order neighbourhood system, eight shift vectors $\mathbf{r}$ can be calculated in advance based on the position constellation of the eight neighbours $\mathbf{n}$ in relation to the principal coordinate $\mathbf{x}$.

A calculation of the prediction error $\delta(\mathbf{x},\mathbf{n},k)$ in (5) requires the definition of predictor $\hat{\mathbf{v}}(\mathbf{x}|\mathbf{n},k)$. For simplicity, we introduce an intermediate notion of *displacement vector justifier* $\mathbf{u}$ that justifies the interpretation of neighbour displacement vector $\mathbf{v}(\mathbf{n})$ at the principal coordinate $\mathbf{x}$ using the k-th region model (cf. Fig. 1(b)), i.e.,

$$\hat{\mathbf{v}}(\mathbf{x}|\mathbf{n},k) = \mathbf{v}(\mathbf{n}) + \mathbf{u}. \tag{6}$$

Upon the notion of $\mathbf{r}$ in Fig. 1(a), the displacement vector justifier $\mathbf{u}$ can be formulated in conjunction with the k-th region model as:

$$\mathbf{u} = \mathbf{v'}(\mathbf{x},k) - \mathbf{v'}(\mathbf{n},k) = \mathbf{v'}(\mathbf{n}+\mathbf{r},k) - \mathbf{v'}(\mathbf{n},k). \tag{7}$$

Since the affine model is linear, we obtain

$$\mathbf{u} = \mathbf{M}_k(\mathbf{n}+\mathbf{r}) + \mathbf{t}_k - (\mathbf{M}_k\mathbf{n} + \mathbf{t}_k) = \mathbf{M}_k\mathbf{r}, \tag{8}$$

that can be substituted to Eq. (6):

$$\hat{\mathbf{v}}(\mathbf{x}|\mathbf{n},k) = \mathbf{v}(\mathbf{n}) + \mathbf{M}_k\mathbf{r}. \tag{9}$$

This derivation demonstrates that the predictor $\hat{\mathbf{v}}(\mathbf{x}|\mathbf{n},k)$ can be calculated from the model parameter matrix $\mathbf{M}_k$, as well as a neighbour displacement vector $\mathbf{v}(\mathbf{n})$ and a vector $\mathbf{r}$ that depends on the neighbourhood constellation, provided $\mathbf{n} \in \mathcal{G}(\mathbf{x}) \cap \Psi_k$ (cf. Eq. (4)). This conclusion enables the realisation of local incoherence $\Delta_\alpha(\mathbf{x},k)$ based on the neighbourhood observation and the chosen random-field neighbourhood system.

In Eq. (1), the local motion coherency likelihood is regularised by the second multiplicand $\Pi_\beta(\mathcal{V}|\mathcal{Q})$. The latter probability defines the likelihood assessing how well each displacement vector fits to the assigned region model, thus termed as the *region motion coherency*:

$$\Pi_\beta(\mathcal{V}|\mathcal{Q}) = \frac{1}{Z_\beta}\exp\left[-\sum_{k=1}^{\lambda}\left\{H_k \cdot \sum_{\mathbf{x}\in\Psi_k}\Delta_\beta(\mathbf{x},k)\right\}\right], \tag{10}$$

where Z_β denotes a normalisation constant and $\Delta_\beta(\mathbf{x},k)$ the *region incoherence function*:

$$\Delta_\beta(\mathbf{x},k) = |v_X(\mathbf{x}) - v'_X(\mathbf{x},k)| + |v_Y(\mathbf{x}) - v'_Y(\mathbf{x},k)|. \tag{11}$$

This function accumulates the absolute differences between the observed $\mathbf{v}(\mathbf{x})$ and the modelled displacement vector $\mathbf{v'}(\mathbf{x},k)$ for each vector component. Following the formulations in Eq. (3) and (10), the likelihood can be written as:

$$\Pr(\mathcal{V}|\mathcal{Q}) = \underbrace{\frac{1}{Z_\alpha}\exp\left[-\sum_{k=1}^{\lambda}\left\{G_k \cdot \sum_{\mathbf{x}\in\Psi_k}\Delta_\alpha(\mathbf{x},k)\right\}\right]}_{Local\ Motion\ Coherency} \cdot \underbrace{\frac{1}{Z_\beta}\exp\left[-\sum_{k=1}^{\lambda}\left\{H_k \cdot \sum_{\mathbf{x}\in\Psi_k}\Delta_\beta(\mathbf{x},k)\right\}\right]}_{Region\ Motion\ Coherency}, \tag{12}$$

It is demonstrated that the complete notion of the affine motion coherency likelihood $\Pr(\mathcal{V}|\mathcal{Q})$ consists of two parts: the local and the region motion coherency. Both analysis terms obey the exponential decay adopted from the Gibbs distributions, which incorporate the cost or penalty formulations in the neighbourhood system of Markov random field and the region

membership fit. At the parameters of the exponential functions, the model specifies the incoherence functions $\Delta_\alpha(\mathbf{x}, k)$ and $\Delta_\beta(\mathbf{x}, k)$ enumerating the degree of vector inconsistency in the local neighbourhood and the observed regions, respectively. These incoherence functions are evaluated based on λ affine models, corresponding to λ observed regions of partition $\mathcal{Q}$.

3.2. The A-Priori Region Boundary Density

The last term in Eq. (1) is the *a priori* density of the region shapes or boundaries given an observed partition $\mathcal{Q}$. The model chooses the density that favours regions of compact shapes and smooth boundaries akin to most physical objects.[19,22] It assigns a large probability to a partition that comprises regions of this desired property. As such, the priori is modelled by a subclass of the 2-D Gibbs-Markov random field in the form of

$$\Pi_\chi(\mathcal{Q}) = \frac{1}{Z_\epsilon} \exp\left[-\mathcal{H}(\mathcal{Q})\right] = \frac{1}{Z_\epsilon} \exp\left[-\mathcal{N}_B B - \mathcal{N}_C C\right], \tag{13}$$

with Z_ϵ being a normalisation constant and the energy $\mathcal{H}(\mathcal{Q})$ assessing the state of partition $\mathcal{Q}$ by counting a number of inhomogeneous cliques or displacement vector pairs at the region borders (being assigned with different region labels). In the second-order random field, the energy is estimated based on the number of inhomogeneous cliques at eight neighbours of all displacement vectors on the field. As inhomogeneous cliques with diagonal or horizontal/vertical configurations influence different degrees of boundary roughness, the model specifies weights to the counts in both cases independently, i.e., B to the number of horizontal or vertical border pairs ($\mathcal{N}_B$) and C to the diagonal ones ($\mathcal{N}_C$).

4. DISPLACEMENT FIELD SEQUENCE SEGMENTATION PRINCIPLE

This section presents the video sequence segmentation principle based on the aforementioned stochastic motion coherency model. The vector reliability assessment measures the confidence extent of each coded displacement vector on the field. It forms a partition analysis foundation. The partition of the first frame is estimated based on the single-frame optimisation technique. The subsequent frames are estimated and refined by the projection and relaxation scheme until the end of video sequence. Note that the latter process insures that the obtained partitions are optimal according to the different field statistics. Fig. 2 summarises the overall segmentation or partitioning principle.

4.1. Vector Reliability Assessment

The method applies the local motion coherency, i.e., the first multiplicand of Eq. (12), to assess the reliability extent of the displacement vectors. The local smoothness with respect to the global motion (i.e., considering the region $k = \lambda = 1$) is investigated. As such, the displacement vectors which are smooth according to the affine-based spatial variance (cf. Eq. (9)) are considered reliable. The reliability measure of the vector at coordinate $\mathbf{x}$ is expressed as:

$$w(\mathbf{x}) = \frac{1}{Z_\alpha} \exp\left[-\sqrt{\frac{\mathcal{N}_{\Psi_\lambda}}{\sigma^2_{\Delta_\alpha(\mathbf{x},\lambda)}}} \cdot \Delta_\alpha(\mathbf{x}, \lambda)\right]. \tag{14}$$

In order that this measure is statistically justified, coefficient G_λ of the local incoherence $\Delta_\alpha(\mathbf{x}, \lambda)$ is chosen at the reciprocal of the normalised standard deviation, being calculated from $\mathcal{N}_{\Psi_\lambda}$ local incoherence observations. The parameter Z_α is a constant ensuring that measure $w(\mathbf{x})$ lies between 0 and 1. The displacement vectors at the frame borders are not considered in this process.

4.2. Partition Projection: Region Displacement Prediction

The partition projection algorithm enables the segmentation of the displacement field sequence, provided the partition of the first displacement field (using the single-frame optimisation[2]). The method initialises the subsequent partition by approximating the label projection upon each displacement vector. First, let us consider the case that a label is located on a 1-D spatial field that is being displaced in the continuous-time domain. Assuming acceleration $a(\tau)$ is a linear function of velocity, $a(\tau) = m - n \cdot v(\tau)$, the label projection $s(\tau)$ can conveniently be calculated from:

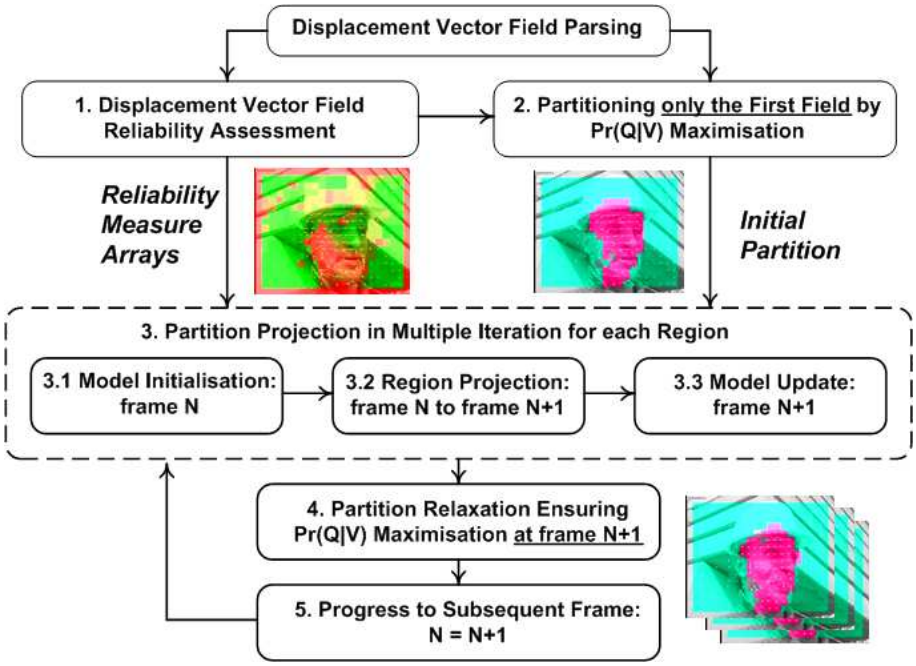


Figure 2. Overall Partitioning Principle

$$s(\tau) = \int\int a(\tau)\, d^{\,2}\tau = \int \frac{m}{n}(1 - e^{-n\tau})d\tau = s_0 + \frac{m}{n}\left[\tau + \frac{e^{-n\tau}}{n}\right], \tag{15}$$

where s_0 denotes the initial projection of the integral. This case requires parameter estimates $\hat{m}$ and $\hat{n}$ to properly identify projection function $s(\tau)$. To overcome this problem, the derivation can be simplified by using the second-order approximation[23] which assumes the constant acceleration, m. As such, Eq. (15) is rewritten by:

$$s(\tau) = \int\int a(\tau)\, d^{\,2}\tau \approx \int v_0 + m\tau\, d\tau = s_0 + v_0\tau + \frac{1}{2}m\tau^2, \tag{16}$$

where v_0 denotes the initial velocity of the integral, and s_0 denotes the initial projection similarly to Eq. (15). Using this result, the projection on a 2-D spatial field in the discrete-time domain (according to the frame sampling configuration of the video sequence being analysed) can be derived in the following.

The projection $\mathbf{S}_{k,j}$ of the k-th region at time τ_j is based on the affine displacement model $\mathbf{T}_{j,k}$ at time τ_j, where $\mathbf{v'}(\mathbf{x}, k, j) = \mathbf{M}_{k,j}\mathbf{x} + \mathbf{t}_{k,j}, \mathbf{x} \in \Psi_k, \mathbf{M}_{k,j} \in \Re^{2x2}$, $\mathbf{t}_{k,j} \in \Re^2$ and that of the two previous frames at times τ_{j-1}, τ_{j-2}, $\mathbf{T}_{k,j-1}$ and $\mathbf{T}_{k,j-2}$, respectively. The displacement model $\mathbf{T}_{j,k}$ is estimated using the weighted linear regression upon the membership function Ψ_k. Assuming accelerations are constant, or $\mathbf{A}_{k,j} = \mathbf{A}_{k,j-1} = \mathbf{A}_{k,j-2}$, the k-th region projection estimate $\mathbf{S}_{k,j}$ for the region membership Ψ_k at time τ_j is estimated by (cf. Eq. (16)):

$$\mathbf{S}_{k,j} \approx \mathbf{S}_{k,j-1} + \mathbf{V}_{k,j-1} \cdot (\tau_j - \tau_{j-1}) + \frac{1}{2}(\mathbf{T}_{k,j-1} - \mathbf{T}_{k,j-2}) \cdot (\frac{\tau_j - \tau_{j-1}}{\tau_{j-1} - \tau_{j-2}})^2. \tag{17}$$

With the constant acceleration or $\mathbf{A}_{k,j-2} = \mathbf{A}_{k,j-1}$, the velocity $\mathbf{V}_{k,j-1}$ in this equation can be identified by

$$\mathbf{V}_{k,j-1} \approx \mathbf{V}_{k,j-2} + (\mathbf{T}_{k,j-2} - \mathbf{T}_{k,j-3}) \cdot \frac{\tau_{j-1} - \tau_{j-2}}{(\tau_{j-2} - \tau_{j-3})^2} \approx \mathbf{V}_{k,j-2} + (\mathbf{T}_{k,j-1} - \mathbf{T}_{k,j-2}) \cdot \frac{\tau_j - \tau_{j-1}}{(\tau_{j-1} - \tau_{j-2})^2}. \tag{18}$$

A substitution of Eq. (18) to Eq. (17) yields:

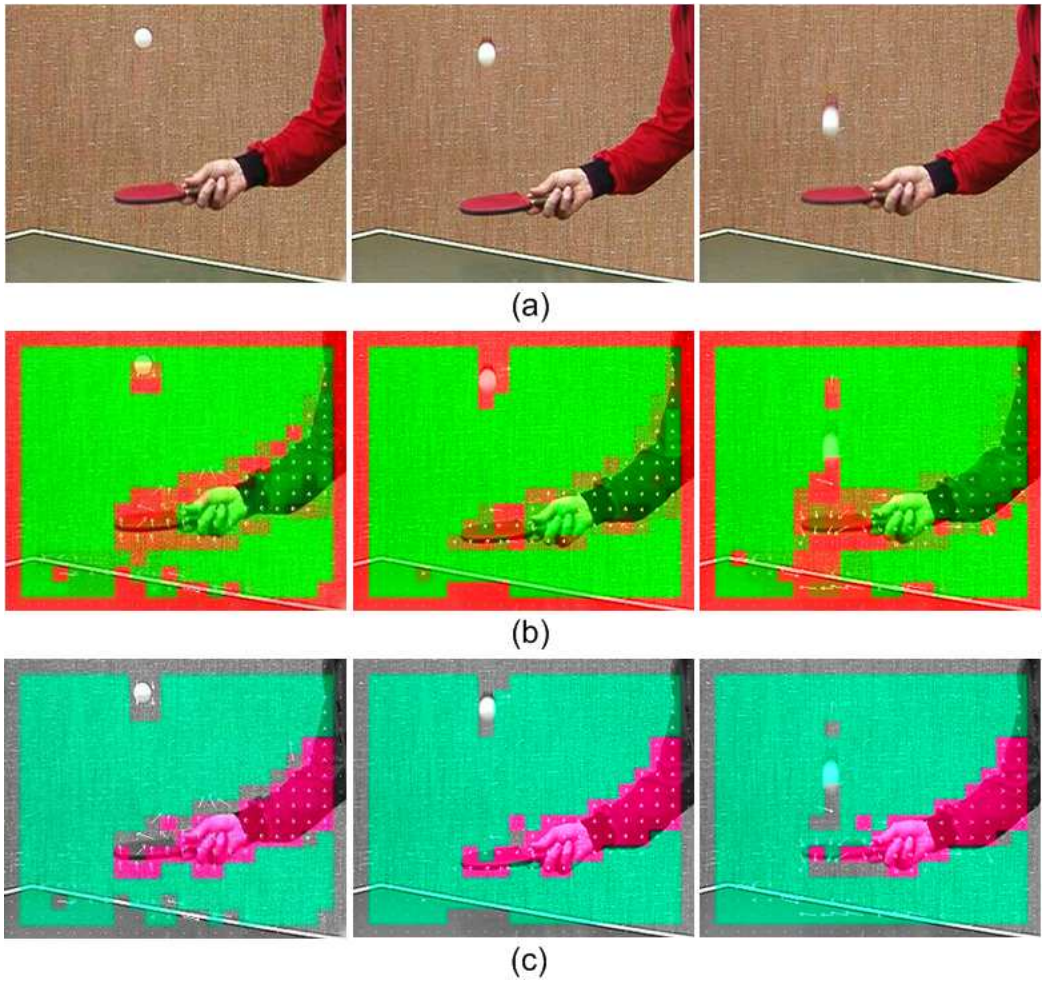


Figure 3. Reliability assessment and segmentation results from sequence *Table Tennis* at frames 4, 7 and 10 (This figure contains color information best viewed by color printout.)

$$\mathbf{S}_{k,j} \approx \mathbf{S}_{k,j-1} + \underbrace{\mathbf{V}_{k,j-2} \cdot (\tau_{j-1} - \tau_{j-2}) + \frac{3}{2} (\mathbf{T}_{k,j-1} - \mathbf{T}_{k,j-2}) \cdot (\frac{\tau_j - \tau_{j-1}}{\tau_{j-1} - \tau_{j-2}})^2}_{projection\ vector\ estimate}. \tag{19}$$

This formulation allows the projection of labels in the k-th region from the previous frame at time τ_{j-1} to the current frame at time τ_j by using the *projection vector estimate* and the projection result at the previous frame (i.e., $\mathbf{S}_{k,j-1}$). The displacement model $\mathbf{T}_{j,k}$ of the current frame is updated after the projection using the weighted linear regression upon the membership function Ψ_k. Through the iteration $k = 1, \ldots, \lambda$, the entire partition at time τ_j is initialised.

4.3. Partition Relaxation

The projected partition is refined to ensure the maximal probability $Pr(\mathcal{Q}|\mathcal{V})$ based on the new given motion field $\mathcal{V}$. The algorithm evaluates the probability $Pr(\mathcal{Q}|\mathcal{V})$ by taking the negative logarithm of (2). This leads to the MAP cost function:

$$-\log\{Pr(\mathcal{Q}|\mathcal{V})\} \propto \sum_{k=1}^{\lambda} \Bigg[\underbrace{G_k \cdot \sum_{\mathbf{x} \in \Psi_k} \Delta_\alpha(\mathbf{x}, k)}_{Local\ Heterogeneity} + \underbrace{H_k \cdot \sum_{\mathbf{x} \in \Psi_k} \Delta_\beta(\mathbf{x}, k)}_{Region\ Heterogeneity} \Bigg] + \underbrace{\mathcal{N}_B B + \mathcal{N}_C C}_{Contour\ Roughness} + \underbrace{\log(Z)}_{Constant}. \tag{20}$$

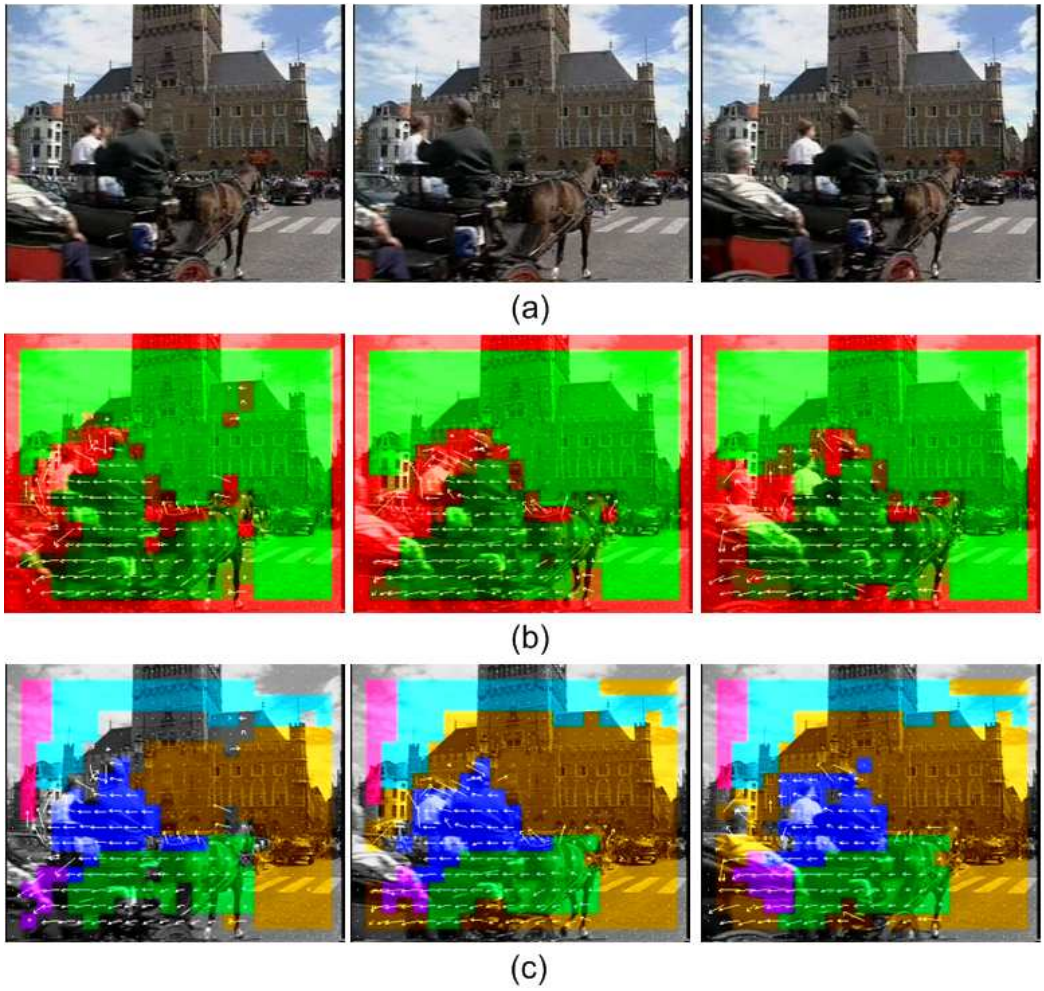


Figure 4. Reliability assessment and segmentation results from sequence *Lancaster*[24] at frames 4, 7 and 10 (This figure contains color information best viewed by color printout.)

The optimal partition $\mathcal{Q}$, that maximises $Pr(\mathcal{Q}|\mathcal{V})$ in terms of the local and region motion coherency and region boundary smoothness, shall minimise this cost function. To arrive at this optimality, the algorithm investigates at every region border if a label substitution may decrease this MAP cost function. At each vector, the substitution label set is gathered from the second-order neighbourhood system (i.e., 8 nearest neighbours) plus a new label. The latter case enables the new object detection on the new displacement field. If two or more substitution tests reduce the MAP cost function, only the configuration which leads to the highest function reduction shall take place. This region reassignment scheme proceeds in multiple raster-scan iterations until the substitution does not decrease the MAP cost, i.e., the optimal partition is found.

The projection and relaxation scheme in Sect. 4.2 and Sect. 4.3 is repeated for the next frame pairs and until the rest of the sequence as shown in Fig. 2.

5. EXPERIMENTAL RESULTS

Sequences *Table Tennis*, *Documentary about buildings*, *Lancaster Television (Lancaster)*,[24] and *Foreman* in CIF format were experimented. This test utilised an MPEG-4 encoder[25] configured at a 25-fps frame rate and an IBBP group of frames. The displacement fields were estimated using a 16-pixel search range and a 512-kbps rate control (TM5 algorithm). Fig. 3, 4, and 5 depict the experimental results from P-type frames 4, 7 and 10 (cf. Subfig. 3, 4, and 5(a)) of the three test sequences. Based on the parsed displacement vector fields of this video test set, the implemented algorithm generated the

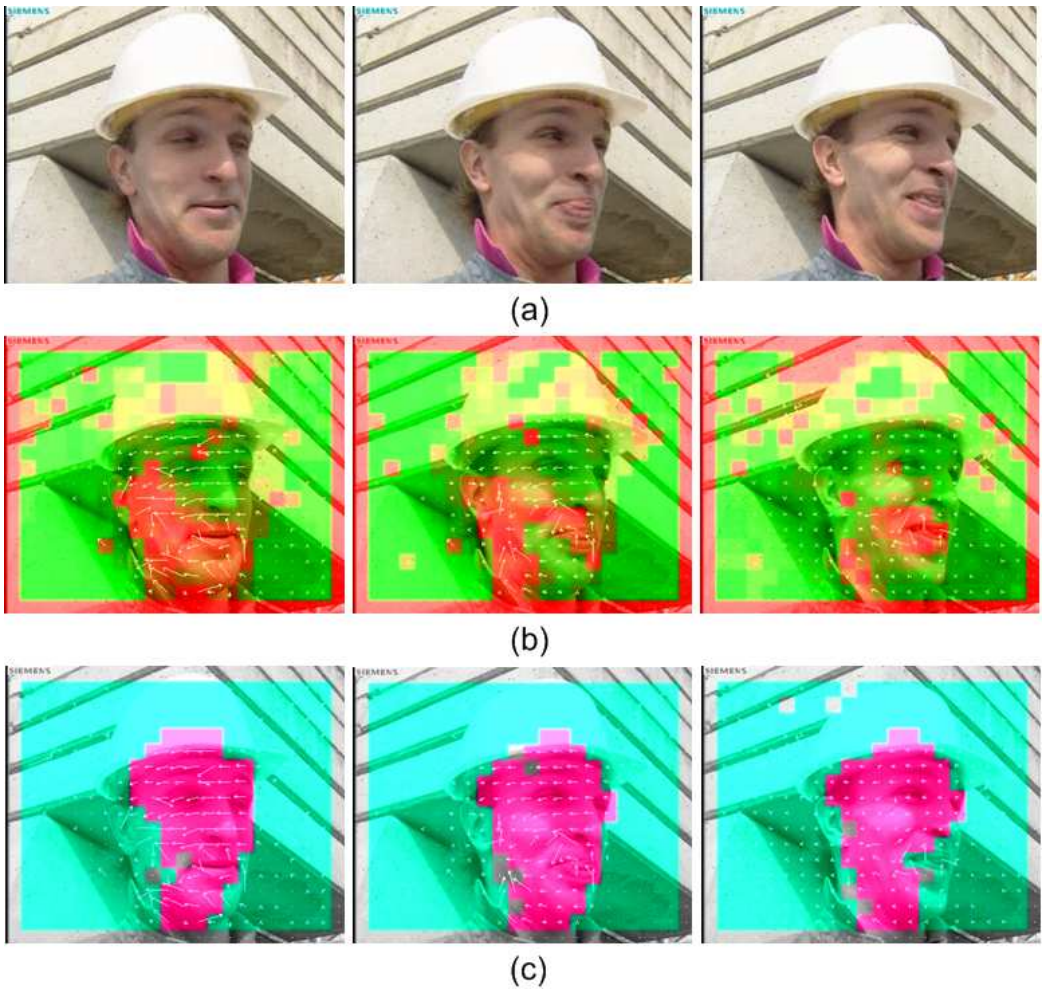


Figure 5. Reliability assessment and segmentation results from sequence *Foreman* at frames 4, 7 and 10 (This figure contains color information best viewed by color printout.)

reliability measure arrays corresponding to the vector field sources (cf. Subfig. 3, 4, and 5(b)). Blocks of the absolute green color represent the highest confident value, while the absolute red for the lowest one. In the segmentation (cf. Subfig. 3, 4, and 5(c)), a unique color illustrates a tracked region. No color is used at the blocks of unreliable displacement vectors.

It is observable that unreliable vectors influenced by new coded signals and/or estimation imprecision of the chosen coding algorithm were efficiently identified (in a red-color range). These vectors were marked mostly around the arm in Fig. 3(b), at the left border of the coach in Fig. 4(b), as well as on the left part of foreman's face in Fig. 5(b). The partitioning and tracking algorithm functioned well to extract the arm in Fig. 3(c), the coach in Fig. 4(c), and the foreman's face in Fig. 5(c) out of the background regions. However, too many regions were detected at the coach and the sky regions in Fig. 4, as a single affine model may not efficiently model complex region motion. The success in temporal association or tracking is evident from the region color preservation at the sequence level.

6. CONCLUSION

This paper presents a stochastic method for segmentation and tracking of the displacement field sequences from compressed videos. Based on the stochastic motion coherency model, the displacement field sequence is analysed and partitioned to multiple temporally associated, motion-semantic regions using the global motion based reliability analysis, partition prediction, and stochastic relaxation. The visually convincing results were demonstrated from the standard sequences.

ACKNOWLEDGMENTS

The authors wish to thank Klaus Illgner for many fruitful discussions, as well as Jens Bialkowski and Marcus Barkowsky for assistance in result visualisation.

REFERENCES

1. A. Kaup, S. Treetasanatavorn, U. Rauschenbach, and J. Heuer, "Video analysis for universal multimedia messaging," in *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, pp. 211–215, (Santa Fe, NM), April 2002.
2. S. Treetasanatavorn, U. Rauschenbach, J. Heuer, and A. Kaup, "Stochastic motion coherency analysis for motion vector field segmentation on compressed video sequences," in *Proceedings of the International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, (Montreux, Switzerland), April 2005.
3. P. L. Correia and F. Pereira, "Classification of video segmentation application scenarios," *IEEE Transactions on Circuits and Systems for Video Technology* **14**(5), pp. 735–741, May 2004.
4. R. Wang, H.-J. Zhang, and Y.-Q. Zhang, "A confidence measure based moving object extraction system built for compressed domain," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, **V**, pp. 21–24, (Geneva, Switzerland), May 2000.
5. O. Sukmarg and K. Rao, "Fast object detection and segmentation in MPEG compressed domain," in *Proceedings of the IEEE TENCON*, (Kuala Lumpur, Malaysia), September 2000.
6. R. Babu, K. Ramakrishnan, and S. Srinivasan, "Video object segmentation: a compressed domain approach," *IEEE Transactions on Circuits and Systems for Video Technology* **14**(4), pp. 462–474, April 2004.
7. A.-R. Mansouri and J. Konrad, "Multiple motion segmentation with level sets," *IEEE Transactions on Image Processing* **12**(2), pp. 201–220, February 2003.
8. M. M. Chang, A. M. Tekalp, and M. I. Sezan, "Simultaneous motion estimation and segmentation," *IEEE Transactions on Image Processing* **6**, pp. 1326–1333, Sep. 1997.
9. A. Tekalp, *Handbook of Image and Video Processing*, ch. 4.9 Video Segmentation, pp. 383–400. Academic Press, 2000.
10. A. Tekalp, *Digital Video Processing*, Prentice Hall, Upper Saddle River, NJ, 1995.
11. J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Seattle, WA), June 1994.
12. J. Wiklund and G. H. Granlund, "Image sequence analysis for object tracking," in *Proceedings of the Scandinavian Conference on Image Analysis*, pp. 641–648, (Stockholm, Sweden), June 1987.
13. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision* **1**(4), pp. 321–331, Jan. 1988.
14. S. Sun, D. R. Haynor, and Y. Kim, "Semiautomatic video object segmentation using vsnakes," *IEEE Transactions on Circuits and Systems for Video Technology* **13**(1), pp. 75–82, January 2003.
15. T. Drummond and R. Cipolla, "Real-time visual tracking of complex structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(7), pp. 932–946, July 2002.
16. V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "Video object segmentation using Bayes-based temporal tracking and trajectory-based region merging," *IEEE Transactions on Circuits and Systems for Video Technology* **14**(6), pp. 782–795, June 2004.
17. F. Moscheni, F. Dufaux, and M. Kunt, "Object tracking based on temporal and spatial information," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, **4**, pp. 1914–1917, (Atlanta, GA), May 1996.
18. C. Gu and M.-C. Lee, "Semiautomatic segmentation and tracking of semantic video objects," *IEEE Transactions on Circuits and Systems for Video Technology* **8**(5), pp. 572–584, Sept. 1998.
19. T. Aach and A. Kaup, "Bayesian algorithms for adaptive change detection in image sequences using Markov random fields," *Signal Processing: Image Communication* **7**, pp. 148–160, August 1995.
20. C. Stiller and J. Konrad, "Estimating motion in image sequences," *IEEE Signal Processing Magazine* **16**, pp. 70–91, July 1999.
21. S. Li, *Markov Random Field Modeling in Image Analysis*, Springer-Verlag, Tokyo, 2001.
22. R. Mester and U. Franke, "Statistical model based image segmentation using region growing, contour relaxation and classification," in *Proceedings of the SPIE Visual Communications and Image Processing*, pp. 616–624, (Cambridge, MA), November 1988.
23. S. Jeannin, A. Divakaran, and B. Mory, *Introduction to MPEG-7 Multimedia Content Description Interface*, ch. Motion Descriptors, pp. 261–280. John Wiley & Sons, West Sussex, England, 2002.
24. MPEG, "Licensing agreement for the MPEG-7 content set." ISO/IEC JTC1/SC29/WG11/N2466, Atlantic City, October 1998.
25. Microsoft, "ISO/IEC 14496 Video Reference Software." Microsoft-FDAM1-2.3-001213.