Image Sequence Analysis for Emerging Interactive Multimedia Services—The European COST 211 Framework

A. Aydın Alatan, *Member, IEEE*, Levent Onural, *Senior Member, IEEE*, Michael Wollborn,
Roland Mech, Ertem Tuncel, and Thomas Sikora, *Senior Member, IEEE*

(Invited Paper)

Abstract— **Flexibility and efficiency of coding, content extraction, and content-based search are key research topics in the field of interactive multimedia. Ongoing ISO MPEG-4 and MPEG-7 activities are targeting standardization to facilitate such services. European COST Telecommunications activities provide a framework for research collaboration. COST 211$^{\mathrm{bis}}$ and COST 211$^{\mathrm{ter}}$ activities have been instrumental in the definition and development of the ITU-T H.261 and H.263 standards for video-conferencing over ISDN and videophony over regular phone lines, respectively. The group has also contributed significantly to the ISO MPEG-4 activities. At present a significant effort of the COST 211$^{\mathrm{ter}}$ group activities is dedicated toward image and video sequence analysis and segmentation—an important technological aspect for the success of emerging object-based MPEG-4 and MPEG-7 multimedia applications. The current work of COST 211 is centered around the test model, called the Analysis Model (AM). The essential feature of the AM is its ability to fuse information from different sources to achieve a high-quality object segmentation. The current information sources are the intermediate results from frame-based (still) color segmentation, motion vector based segmentation, and change-detection-based segmentation. Motion vectors, which form the basis for the motion vector based intermediate segmentation, are estimated from consecutive frames. A recursive shortest spanning tree (RSST) algorithm is used to obtain intermediate color and motion vector based segmentation results. A rule-based region processor fuses the intermediate results; a postprocessor further refines the final segmentation output. The results of the current AM are satisfactory; it is expected that there will be further improvements of the AM within the COST 211 project.**

Index Terms— **Camera motion estimation, change detection, content-based search, COST 211, data fusion, image segmentation, interactive multimedia, motion analysis, motion-based segmentation, MPEG-4, MPEG-7, object tracking, video processing, video segmentation.**

I. INTRODUCTION

INTERACTIVE multimedia services will strongly influence and even dominate the future of telecommunications. The flexibility and efficiency of the used coding systems, as well as the ability to efficiently access and search particular content of interest in distributed databases, are essential for the success of these emerging services. In this regard the ISO MPEG-4 standard has attracted much attention recently for providing a technical standardized solution for content-based access and manipulation for these applications. The standard is targeted for flexible interactive multimedia applications with provisions for content access and manipulation [1], [17], [18].[1] The new ISO MPEG-7 initiative will further standardize a *Multimedia Content Description Interface* with the aim to ease content search and user-controlled content streaming for a variety of database and broadcasting environments [19].

Anticipating the rapid convergence of telecommunications, computer, and TV/film industries, the MPEG group officially initiated the MPEG-4 standardization phase in 1994—with the mandate to standardize algorithms for audio–visual coding in multimedia applications, allowing for interactivity, high compression and/or universal accessibility, and portability of audio and video content. In addition to provisions for efficient coding of conventional image sequences, MPEG-4, which will be standardized in 1998, will provide a representation of the audio and video data that can be accessed and manipulated on an audio–visual object basis, even in the compressed domain at the coded data level with the aim to use and present the objects in a highly flexible way. In particular, future multimedia applications as well as computer games and related applications are seen to benefit from the increased interactivity with the audio–visual content.

The MPEG-4 standard will assist the coding of objects in image sequences separately in different object layers. Thus, in MPEG-4 images as well as image sequences can be considered to be arbitrarily shaped—in contrast to the standard MPEG-1 and MPEG-2 definitions. Here it is envisioned that video sequences are decomposed into individual objects in a scene and the objects are encoded entirely separately in individual

---

[1] See also http://wwwam.hhi.de/mpeg-video.

object layers. In general, this provides to the user an extended content-based functionality (the ability to separately access and manipulate video content) and it is also possible to achieve increased image quality for a number of applications. This will require the segmentation of video sequences into the objects of interest prior to coding. The segmentation of images and video into these separate objects (arbitrarily shaped regions in images) is not specified by the MPEG-4 standard; however, this can be an extremely difficult task for many applications. If video was originally shot in a studio environment using the chroma-key technology, image segmentation can be easily performed—e.g., for news sequences and scenes generated in a more elaborate virtual studio. If no chroma-key segmentation is available—as for most scenes under investigation—the segmentation of the objects of interest needs to be performed using automatic or semiautomatic segmentation algorithms. As of this writing, the authors were not aware of the existence of a universal algorithm that could potentially solve the segmentation problem. The video segmentation task still remains to a large extent an unsolved problem, resulting in a variety of tools and algorithms described in literature—each of them specialized and optimized for a specific segmentation task to be performed [2]–[16]. Even then, in many applications, a considerable amount of user interference with the segmentation process is required to indicate to the algorithm where the objects of interest are, and to ensure stable and precise results.

The purpose of this paper is to provide an overview of the segmentation algorithms developed in the framework of the European COST 211$^{\text{ter}}$ activity. The COST 211$^{\text{ter}}$ group is a research collaboration of partners from European countries. The main focus in the group is to develop tools for the segmentation of image sequences—both automatic and semiautomatic allowing user interaction—to assist MPEG-4 applications. The group has a so-called *Test Model* approach adopted—similar to the Test Model approach in MPEG—to develop and optimize analysis tools and algorithms in a collaborative manner in a common environment under controlled conditions.

The service and application profiles envisioned by the COST 211 group include:

- real-time communications (e.g., MPEG-4 conversational services such as surveillance or video conferencing);
- retrieval services (e.g., access to MPEG-4 video or audio objects stored on a remote data base);
- interactive distribution services (e.g., distribution of news clips with user-defined content).

The paper is organized as follows. The general framework of the European COST activity is described in Section II with particular emphasis on the scope and objectives of the COST 211$^{\text{ter}}$ initiative. Section III outlines the *Test (Analysis) Model* approach used in COST 211$^{\text{ter}}$; the algorithms defined by COST 211$^{\text{ter}}$ for automatic and semiautomatic segmentation of image sequences are described in that section, too. Results obtained from standard image test sequences are also presented in Section III. Finally, Section IV concludes the paper.

II. CONTENT ANALYSIS FOR EMERGING INTERACTIVE MULTIMEDIA SERVICES—THE COST 211$^{\text{ter}}$ FRAMEWORK

A. The COST Framework in General

The European COST (Cooperation Europeenne dans le recherche scientifique et technique) Telecommunications activities—initiated by the European Community (EU)—provide an open and flexible framework for R&D cooperation in Europe. COST actions involve precompetitive basic research or activities of public utility—in particular research topics which are of strategic importance for the development of the information society.

In contrast to other EU research programs, collaboration within COST allows utmost freedom to the participant to choose and conduct their research. All COST 211 actions have been and are following a flexible framework which is open to a large number of members and allows long periods of cooperation. Each COST action focuses on specific topics for which there is interest in particular COST countries, with the primary aim to raise intellectual property rights. Any COST country can join any action by signing the Memorandum of Understanding (MoU) which is the legal basis of the action even though it in fact resembles an expression of good faith rather than a legally binding document. Each MoU governs the joint aims, the type of activity to be pursued, the terms of participation, and compliance with intellectual property rights.

B. History of COST 211

The present COST 211$^{\text{ter}}$ action is a follow-up project to COST 211 and 211$^{\text{bis}}$, all dealing with redundancy reduction techniques applied to video signals. The two preceding COST 211 projects paved the way to the creation and maintenance of a high level of expertise in the field of video coding in Europe and resulted in a strong European contribution to the standardization process in this field. The digital video standards adopted and widely deployed today are strongly influenced by the results of the COST 211 actions and the label COST 211 is well known and respected worldwide in the field of video coding.

Action COST 211 started in 1976 when the development of digital video was in its infancy. The result of this phase was a video coding algorithm allowing the first digital videoconference system at 2 Mbit/s, and this in turn led to the first ITU standard in this field (H.120). The follow-up action, COST 211$^{\text{bis}}$, improved to a large extent the efficiency of the coding algorithm allowing videophone and videoconferencing at ISDN rates and further ITU standards emerged (in particular H.261). COST 211$^{\text{ter}}$ has been a major contributor to the standardization activity in MPEG (MPEG 1, 2 and 4) and ITU. The COST projects cover more than 20 years of cooperation and strengthened considerably the European position and influence in the field of video coding. COST 211$^{\text{quat}}$ action will be the follow-up platform for cooperation in this field for another five years, starting in 1998. The COST 211 actions are considered to be complementary to other EU projects in this field.

TABLE I
MAIN RESEARCH ITEMS COVERED BY THE COST 211[ter] GROUP

| Functionality	Possible Applications
Audio-Visual Content Identification	e.g. the automatic detection of content of interest in video scenes, such as a person, etc.
Content and/or Feature Extraction	e.g. the automatic extraction of content of interest in complex video or audio scenes
Tracking of Content and/or Features	e.g. the tracking of content or features of interest over time in an audio or video scene allowing user interaction with the tracking process
Selective Coding of Objects	for the separate compression of content data to allow separate access and query to the content in a data-base
Improved Coding Efficiency	for the reconstruction of content with sufficient quality by means of sophisticated compression technology
Selective Error Protection	for a robust storage or transmission of the most important content in error prone environments
Content-Based Visual Database Query &Indexing	e.g. the visual query for content of interest
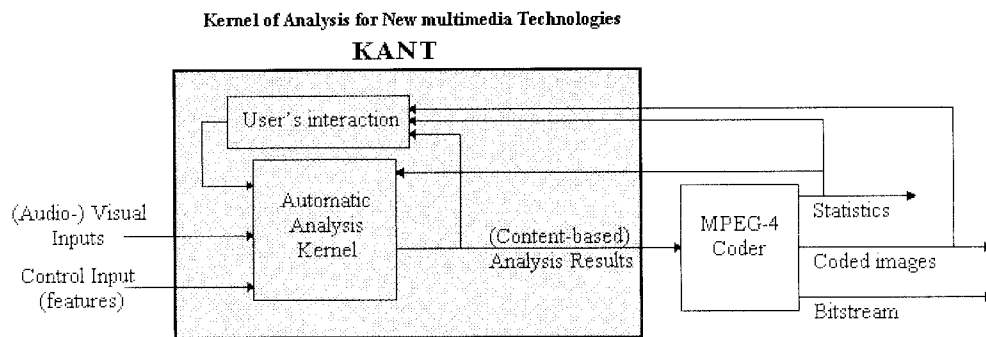
Fig. 1. The European COST 211[ter] AM model: KANT—broad overview.

C. Current Objectives: Image Analysis for Emerging Multimedia Applications

The main objective of the current COST 211[ter] action group is to improve the efficiency of redundancy reduction and to develop content analysis techniques for video signals to assist future multimedia applications. In particular the group focuses on content-oriented processing for emerging interactive multimedia services based on the ongoing ISO MPEG-4 standardization phase [17][2] as well as the new ISO MPEG-7 initiative [19]. The current research items covered by COST 211[ter] group are outlined in Table I. The aim is to define and develop a set of tools assisting these new services in the analysis, characterization and processing of their video and audio signals.

The basics of the current COST scenario, which is called KANT (Kernel of Analysis for New Multimedia Technologies), is outlined in Fig. 1. KANT provides the desperately needed segmentation tools which would then give life to the MPEG-4 operation. The KANT is also the basis of the current Analysis Model (AM). The AM consists of a set of functional blocks and their interconnections, together with algorithms for each of those blocks and their full software implementation. Thus, AM provides a particular solution to the outlined KANT.

The participants in the project simultaneously undertake and coordinate research and development work with the aim of applying analysis and coding techniques to video and audio signals to assist emerging interactive multimedia applications.

The focus items are researched by means of:

- investigation through nonreal-time computer simulation of algorithms and tools based on a Test Model approach (see below);
- implementation and optimization by software;
- trials together with selected applications and users in order to evaluate the efficiency of the developed tools.

The action primarily focuses on studies applied to video signals, but in the future will associate as much as possible with equivalent studies applied to audio signals.

D. The COST 211[ter] Test Model Approach

The cooperation within the COST 211 group is centered around a test model, called the *Analysis Model* for the purposes of COST 211. This approach is adopted to investigate,
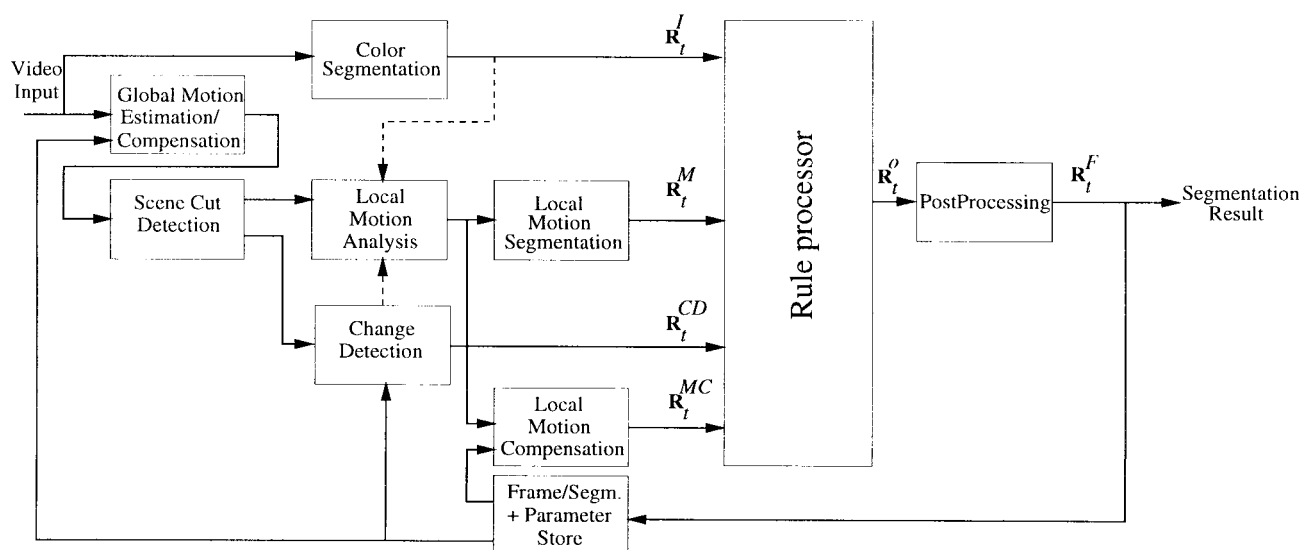
---

[2] See also http://wwwam.hhi.de/mpeg-video.

Fig. 2. The European COST 211$^{\text{ter}}$ AM model.

compare, and optimize algorithms for image and video analysis in an experimental approach. In contrast to research conducted on an individual basis, this approach enables the comparison of competing technology based on a *test-bed* and under agreed experimental conditions and performance measures.

The purpose of a test model in COST—very similar to that in MPEG—is to describe completely defined "Common Core" algorithms, such that collaborative experiments performed by multiple independent parties can produce identical results and will allow the conduction of "Core Experiments" under controlled conditions in a common environment. A test model specifies the formats for the input and the output. It fully specifies the algorithm for image analysis and the criteria to judge the results.

The COST 211$^{\text{ter}}$ meeting in Ankara, Turkey, in October 1996 witnessed the definition of the 1st AM—which consists of a full description of tools and algorithms for automatic and semiautomatic image sequence segmentation (object detection, extraction, and tracking). The AM was then refined in further meetings and progressed to its third version. In addition to the full description of the AM algorithm, a software implementation was developed by the group to allow partners a convenient way to perform experimentation and to integrate provisions for improvement.

The defined AM will be further refined and improved in a collaborative way by developing, exchanging, and testing algorithms and software.

III. THE COST ANALYSIS MODEL

Current approaches for segmentation of video sequences in order to detect moving objects can be subdivided into three classes: *intensity parameter based*, *motion parameter based*, and *simultaneous motion estimation and segmentation* methods. (See, for example [20].) The intensity-based methods use spatio-temporal intensity information without explicitly segmenting the motion parameters. These methods are usually based on a *change detection* approach followed by motion

estimation. They separate moving objects from background regions, but cannot distinguish between different moving objects [21]–[24]. The algorithms in the second class use the motion information which may be explicitly provided as an input, or may be estimated by further processing from the available intensity data. Given the motion data, each image is segmented into a number of regions with similar motion, using an algorithm like the K-means algorithm [25], modified Hough transform [26], Bayesian segmentation [27] or merging [28]. Although multiple objects can be located successfully by such methods, due to occlusions and the "chicken-egg" problem between motion estimation and segmentation, the object boundaries may not be determined accurately. The methods in the third class simultaneously update the motion estimates and the segmentation; therefore, they are accepted to be the most powerful approaches among all [29], [30]. However, such methods are unattractive due to their high computational complexity.

A. Basic Structure of the Analysis Model

The main idea behind the approach taken by the current COST 211 Analysis Model is the fusion of various intermediate results by a set of rules for a better segmentation result [31], [32]. Currently, motion information, color information, and intensity changes are the main sources of clues used to get the intermediate results which are then fused. Furthermore, the temporally accumulated segmentation information is also used during the fusion to provide temporal continuity for a better final result and tracking. The block diagram of the current version (version 3.0) is shown in Fig. 2. The AM version 3.0 provides two modes of operations. In the first mode, the result is a binary mask which distinguishes the moving objects from the static/moving background. In the second mode, the moving objects are further segmented into multiple objects.

In order to cope with possible camera motion and changing image content, a camera motion compensation and a scene cut detection are applied within a preprocessing unit. Intermediate segmentation results are fused by the rule processor, leading

to the final segmentation results which are then refined by an appropriate postprocessing step, if necessary.

The output of the AM are the segmentation masks; some internal parameters which are found during the computations can also be taken as byproduct outputs. The AM structure is completely independent of applied input sequence, or variations in an input sequence. Therefore, it provides fully automated segmentation. The functions of the AM blocks are described in the following subsections.

B. Global Motion Compensation and Scene-Cut Detection

Given two successive frames I_t and I_{t-1} of a video sequence captured by a static or moving camera, an apparent camera motion is estimated and compensated [23]. A single rigid planar scene is assumed to move in the three-dimensional (3-D) space; the eight parameters a_1–a_8, which can reflect any kind of such rigid motion including zooming and panning, are estimated [21]. For each pixel (X_{t-1}, Y_{t-1}) in frame I_{t-1} the corresponding pixel (X_t, Y_t) in frame I_t is given by

$$X_t = \frac{a_1 X_{t-1} + a_2 Y_{t-1} + a_3}{a_7 X_{t-1} + a_8 Y_{t-1} + 1}$$
$$Y_t = \frac{a_4 X_{t-1} + a_5 Y_{t-1} + a_6}{a_7 X_{t-1} + a_8 Y_{t-1} + 1}.$$

The camera motion is estimated by regression considering only the pixels within the background regions of the previous image.

Then, a following step finds those areas which do not fit into the "single rigid plane" assumption. In these areas the estimated displacement is pelwise refined by performing a full search within a squared area of limited size. The estimated displacement vector for the background is further improved by excluding these "failure" areas. The camera motion is compensated only if a moving camera has been detected.

In case of a scene cut between two consecutive frames, the motion compensation is meaningless. Therefore, detection of scene cuts improves the performance. The proposed scene cut detector [23] evaluates whether the difference between the current original image I_t and the camera motion compensated previous image exceeds a given threshold; the evaluation is performed only within the background region of the previous frame.

C. Analysis of Color, Motion, and Intensity Changes

1) Color Segmentation: In this step, the current frame is divided into a predefined number of regions using only the color information. Most of the still image segmentation tools in the literature can be used to find the output mask, $\mathbf{R}_t^I$, which describes the regions with coherent intensity and labels them (see for example, [33] and [34]). The goal, however, is to have regions whose boundaries coincide with the boundaries of the real (semantic) objects in the scene: each region must belong to only one semantic object. In our proposed algorithm, a recursive shortest spanning tree (RSST)-based segmentation method is used to segment the current frame into some regions each having uniform intensity [35]. RSST has the advantage of not imposing any external constraints on the image. Some other methods, such as split-merge algorithm

(see, for example, [36]), which requires segments consisting of nodes of a quadtree, can produce artificial region boundaries. Furthermore, RSST segmentation permits simple control over the number of regions, and therefore amount of detail, in the segmentation image. The simulation results on still images support the superior image segmentation performance of this method [37].

2) Motion Analysis, Segmentation, and Compensation: The motion between two consecutive frames is estimated. Among many available motion estimation algorithms (see, for example, [20]), a three-level Hierarchical Block Matching (HBM) algorithm [38] is used due to its acceptable results with quite low computational demand. The estimated block motion vectors are interpolated in order to obtain a dense motion field. Better motion estimation methods may be employed at the expense of increased computational complexity.

Using the RSST algorithm, the estimated motion vector field is segmented in a similar manner to color segmentation. During this process, the two components of the motion vectors at each pixel are used instead of the three color components of the color segmentation stage. The resultant output is denoted as $\mathbf{R}_t^M$ in Fig. 2. Since the motion estimation step might contain some matching errors and occlusions, the resulting segmentation field is expected to be coarse. The locations of the objects, however, are usually found correctly.

For the continuity of the extracted objects, the previous segmentation results should be included in the rule-based data fusion process. Tracked object information is inserted into the current segmentation process. Using the available previous segmentation result at $t-1$ and the estimated motion information, a temporal prediction of the current segmentation mask is obtained as $\mathbf{R}_t^{MC}$. By the help of $\mathbf{R}_t^{MC}$, not only a better segmentation, but also the tracking of the individual objects in the scene, can be achieved. Moreover, the status of the objects (i.e., halted, newly exposed) can be determined by including the information in this mask to the current segmentation process.

3) Change Detection: The change detection mask between two successive frames is estimated. In this mask, pixels for which the image luminance has changed due to a moving object are labeled as "changed." The algorithm for the estimation of the change detection mask [22] can be subdivided into several steps, which are described in the following.

First, an initial change detection mask between the two successive frames is generated by global thresholding the frame difference. In a second step, boundaries of changed image areas are smoothed by a relaxation technique using local adaptive thresholds [39], [40]. Thereby, the algorithm adapts frame-wise automatically to camera noise [22]. In order to finally get temporal stable object regions, a memory is used in the following way: The mask after thresholding is connected with the previous $\mathbf{R}_t^O$. Specifically, the mask after thresholding is extended by pixels which are set to foreground in the $\mathbf{R}_t^O$ of the previous frame. This is based on the assumption that all pixels which belonged to the previous $\mathbf{R}_t^O$ should belong to the current change detection mask. In order to avoid infinite error propagation, however, a pixel from the previous $\mathbf{R}_t^O$ is only labeled as changed in the change detection mask, if it was also

labeled as changed in one of the last L frames. The value L denotes the depth of the memory, which adapts automatically to the sequence by evaluating the size and motion amplitudes of the moving objects in the previous frame. By the last step, the mask is simplified and small regions are eliminated.

D. Rule-Based Region Processing

The four input segmentation masks of the region processor have different properties. Although the color segmentation mask $\mathbf{R}_t^I$ is usually oversegmented, it contains the most reliable boundaries whereas the boundaries of both $\mathbf{R}_t^M$ and $\mathbf{R}_t^{MC}$ are blurred or not very accurate. However, $\mathbf{R}_t^M$ locates the disjoint objects with some semantic meaning in the current frame and $\mathbf{R}_t^{MC}$ contains the previous segmentation information. Finally, the change detection mask $\mathbf{R}_t^{CD}$ reliably labels stationary and moving areas. Thus, the rule processor can use the boundaries supplied by $\mathbf{R}_t^I$ in order to form $\mathbf{R}_t^O$ by merging the regions of $\mathbf{R}_t^I$ appropriately, taking into account the information provided by the other segmentation masks. Moreover, it can also use $\mathbf{R}_t^{MC}$ to track the objects throughout a sequence.

Currently, there are two modes of the AM, providing different functionalities. The first mode allows a reliable detection of all moving objects without being able to distinguish between objects with different motion. In the second mode, it is possible to further distinguish between these moving objects. Both modes are described in the following subsections. The module for postprocessing is only used if the second mode is active.

1) Detection of Moving Objects and Background Regions (Mode 1): In this mode, only the results from the change detection, color segmentation, and local motion analysis are used in order to distinguish between moving objects (denoted as foreground in the following rule) and background. Thus, the resultant segmentation mask is binary.

In the first step, the uncovered background areas are eliminated from the estimated change detection mask as in [21] and [22], resulting in an initial object mask. Only the estimated motion information for pixels within the changed regions is used. A pixel is set to "foreground" if both the starting and ending points of the corresponding displacement vector are in the "changed" area of the change detection mask. If not, it belongs to uncovered background and is therefore set to "background."

Since the color segmentation has accurate boundary information, the object boundaries on the final mask, i.e., the segmentation result, are copied from the color segmentation result whenever appropriate. The decision rule is as follows.

Rule 1—Foreground detection: If at least t% of the pixels within a region of the color segmentation were detected as foreground in the initial object mask, all pixels of the region already detected as foreground and all pixels within a correction range of N pixel with respect to the boundary of the initial object mask are set to foreground, and the other pixels are set to background.

Rule 2—Background detection: If less than t% of the pixels within a region of color segmentation were detected as foreground in the initial object mask, all pixels of the

region already detected as background and all pixels within a correction range of M pixel with respect to the boundary of the initial object mask are set to background, and the other pixels are set to foreground.

2) Extraction of Moving Objects (Mode 2): Let $\mathbf{I}_{t,i} \subseteq \mathbf{R}_t^I$ define the regions in $\mathbf{R}_t^I$ where $\mathbf{R}_t^I = \cup_i^{i_{\max}} \mathbf{I}_{t,i}$ and $\mathbf{I}_{t,i} \cap \mathbf{I}_{t,j} = \emptyset$ for $i \neq j$. The relationships between $\mathbf{R}_t^M$ and $\mathbf{M}_{t,m}$, $\mathbf{R}_t^{MC}$ and $\mathbf{K}_{t,k}$, $\mathbf{R}_t^{CD}$ and $\mathbf{C}_{t,c}$ can be defined in a similar way. The objects at time t are also defined as $\mathbf{O}_{t,q} \subseteq \mathbf{R}_t^F$ with $q \in \{1 \cdots q_{\max,t}\}$, where $q_{\max,t}$ shows the number of objects at time t. Let $a(\cdot)$ represent the *area* operator which gives the area (i.e., the number of pixels) of a region. Let $\mathbf{R}^X$ and $\mathbf{R}^Y$ be two different segmentations over the same lattice; $\mathbf{R}^X = \cup_g^{g_{\max}} \mathbf{X}_g$ and $\mathbf{R}^Y = \cup_h^{h_{\max}} \mathbf{Y}_h$. A *projection operator* $\mathcal{P}$ is defined between each $\mathbf{Y}_h$ and $\mathbf{R}^X$ as

$$\mathcal{P}(\mathbf{Y}_h, \mathbf{R}^X) \doteq \mathbf{X}_{g'} \quad \ni \quad g' = \arg\max_g \left\{ a(\mathbf{X}_g \cap \mathbf{Y}_h) \right\}. \tag{1}$$

The first step of rule-based processing for Mode 2 is finding a corresponding $\mathbf{M}_{t,m'}$ (similarly $\mathbf{K}_{t,k'}$) region for each $\mathbf{I}_{t,i}$ region. This is accomplished by finding $\mathcal{P}(\mathbf{I}_{t,i}, \mathbf{R}_t^M)$ and $\mathcal{P}(\mathbf{I}_{t,i}, \mathbf{R}_t^{MC})$. This mapping increases the accuracy of the boundaries in $\mathbf{R}_t^M$ and $\mathbf{R}_t^{MC}$ as shown in Fig. 3. At the end of this step, each $\mathbf{I}_{t,i}$ region will have a corresponding $\mathbf{M}_{t,m'}$ and a $\mathbf{K}_{t,k'}$ regions.

Using the projection operator of (1), we also define a *group operator* $\mathcal{G}$ as

$$\mathcal{G}(\mathbf{Y}_h, \mathbf{R}^X, g) \doteq \left\{ \mathbf{Y}_h : \mathcal{P}(\mathbf{Y}_h, \mathbf{R}^X) = \mathbf{X}_g \right\}. \tag{2}$$

This operator gives the set of $\mathbf{Y}_h$ regions whose projections onto $\mathbf{R}^X$ give the same region $\mathbf{X_g} \subset \mathbf{R}^X$.

Before stating the rules for segmentation of the objects, an auxiliary set $\mathbf{B}$ is defined as

$$\mathbf{B}_{t,x,y} \doteq \left\{ \mathbf{I}_{t,i} : \mathcal{P}(\mathbf{I}_{t,i}, \mathbf{R}_t^{MC}) = \mathbf{K}_{t,y}; \mathcal{P}(\mathbf{I}_{t,i}, \mathbf{R}_t^M) \neq \mathbf{M}_{t,x} \right\} \tag{3}$$

which relates the previous and current motion with an intensity region $\mathbf{I}_{t,i}$. $\mathbf{B}_{t,x,y}$ represents the set of $\mathbf{I}_{t,i}$ whose projection onto $\mathbf{R}_t^{MC}$ gives regions $\mathbf{K}_{t,y}$ *and* projection onto $\mathbf{R}_t^M$ gives regions other than $\mathbf{M}_{t,x}$. The auxiliary set $\mathbf{B}$ is very useful for compact description of the decision rules given below.

The following rules are applied to obtain the mask, $\mathbf{R}_t^O$, at the output of the rule processor [31]. The rules are applied not for all (x,y) pairs, but for (x,y) pairs for which $\mathcal{G}(\mathbf{I}_{t,i}, \mathbf{R}_t^M, x) \cap \mathcal{G}(\mathbf{I}_{t,i}, \mathbf{R}_t^{MC}, y) \neq \emptyset$.

Rule 1—Tracking of objects: At time t, for a given y (that is, for a previously segmented object),

If $(\mathbf{B}_{t,x,y} = \emptyset)$ then the output object is given by
$$\mathbf{O}_{t,y} = \mathcal{G}(\mathbf{I}_{t,i}, \mathbf{R}_t^{MC}, y)$$

An empty $\mathbf{B}_{t,x,y}$ corresponds to the case where a group of $\mathbf{I}_{t,i}$ regions which belonged to past object y are now labeled as x; it is then concluded that x and y actually denote the same semantic object. Therefore, it is decided that this object is being *tracked*. In other words, the objects from current and previous time instants are *matched*; there are no new objects. These tracked regions which have the same y label (i.e., the same motion behavior) are merged in order to construct a region in $\mathbf{R}_t^O$; this constructed region gets the common
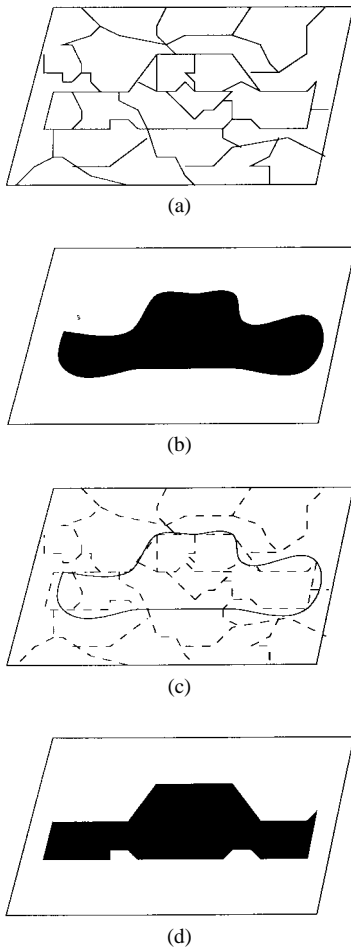
Fig. 3. The projection of $\mathbf{I}$ regions onto $\mathbf{R}_t^{MC}$ and correction of boundaries. (a) Color segmentation. (b) Motion compensated segmentation. (c) Color segmentation projected onto motion compensated segmentation. (d) Corrected boundaries of motion compensated segmentation.

motion-based label y. This rule covers four distinct cases: an object (e.g. background) continues its stationarity, a moving object continues its motion, a previously segmented stationary object begins to move (but it is still identified as the same object since its label has not changed), and a moving object halts (again, it is identified as the same object even if it is stopped, now). The needed extra information to identify which one of these four cases has occurred comes from the past and current change detection masks, $\mathbf{R}_t^{CD}$ and $\mathbf{R}_{t-1}^{CD}$, respectively.

Rule 2—Newly exposed objects: At time t, for a given y (i.e., previously segmented object),
   If $(\mathbf{B}_{t,x,y} \neq \emptyset)$ then
     If $(\mathcal{P}(\mathbf{O}_{t-1,y}, \mathbf{R}_{t-1}^{CD}) = stationary\ region)$ then
       If $(\mathcal{P}(\mathbf{R}_t^{CD}, \mathcal{G}(\mathbf{I}_{t,i}, \mathbf{R}_t^M, x)) = stationary\ region)$ then
         $\mathbf{O}_{t,y} = \mathcal{G}(\mathbf{I}_{t,i}, \mathbf{R}_t^M, x)$
       else
         $\mathbf{O}_{t,\text{new}} = \mathcal{G}(\mathbf{I}_{t,i}, \mathbf{R}_t^M, x)$

This rule is in effect if a region (object) in the stationary region is detected to change its label. This could happen if a still object (which was indistinguishable from the background) has now started to move. Thus, the old background is now split into two: one of them is the new background $\mathbf{O}_{t,y}$, and the other one is $\mathbf{O}_{t,\text{new}}$ which has a newly generated label.

Rule 3—Articulated motion of objects: At time t, for a given y (i.e., previously segmented object),
   If $(\mathbf{B}_{t,x,y} \neq \emptyset)$ then
     If $(\mathcal{P}(\mathbf{O}_{t-1,y}, \mathbf{R}_{t-1}^{CD}) = moving\ region)$ then
       If $(\mathcal{P}(\mathbf{R}_t^{CD}, \mathcal{G}(\mathbf{I}_{t,i}, \mathbf{R}_t^M, y)) = stationary\ region)$ then
         $\mathbf{O}_{t,\text{new}} = \mathcal{G}(\mathbf{I}_{t,i}, \mathbf{R}_t^M, x)$
       else
         $\mathbf{O}_{t,y} = \mathcal{G}(\mathbf{I}_{t,i}, \mathbf{R}_t^M, x)$

This rule is in effect if a region (part of an object), which has been moving, is detected to change its label. This could happen if a moving region which was thought to be a single object due to the uniformity of its motion is found out to be actually consisting of multiple objects; some of these objects now comes to a halt. Although this is an extraordinary situation, in some cases this rule might be necessary. If this is the case, some parts of the previously moving object now keeps the same label, whereas the newly stopped part is now recognized as a new object with a new label.

An example for the application of the proposed rules: The proposed set of rules can be clarified by the help of a simple example which is shown in Fig. 4. We can assume that $\mathbf{R}_t^I$, $\mathbf{R}_t^M$ and $\mathbf{R}_t^{MC}$ masks are obtained by the methods explained in the previous section. When all the regions in $\mathbf{R}_t^I$ are projected onto $\mathbf{R}_t^M$, $\{H, I, J\}$ and $\{E, F, G\}$ color regions are assigned to two differently moving regions, respectively. All the remaining color regions, $\{A, B, C, D\}$, are labeled as stationary. The color regions in $\mathbf{R}_t^I$ are also projected onto $\mathbf{R}_t^{MC}$ in a similar manner.

It can be easily realized that projected color regions of the previously moving object-1 in $\mathbf{R}_t^{MC}$ and one of the moving regions in $\mathbf{R}_t^M$ are same; i.e. $\mathbf{B}_{t,x,y} = \emptyset$. These color regions correspond to $\{E, F, G\}$. Hence, according to the proposed Rule 1, a moving object is being tracked and the color regions, $\{E, F, G\}$, should be merged to carry object-1 to the output segmentation mask. However, the situation is different for the color regions of the previously stationary object-0 in Fig. 4. These previously stationary regions, $\{A, B, C, D, H, I, J\}$, have different corresponding motion regions in $\mathbf{R}_t^M$. More specifically, while the $\{A, B, C, D\}$ regions belong to a stationary region in $\mathbf{R}_t^M$, $\{H, I, J\}$ regions are found out to be moving at the current time instant. Hence, a new object is said to exposed according to Rule 2 and it is assigned with a new object number. Following this rule, $\{H, I, J\}$ color regions are merged to obtain this new object in the final mask, where as rest of the color regions, $\{A, B, C, D\}$ are left to be the background object-0.

E. Postprocessor

The postprocessor is only used if mode 2 of the rule processor is active, because, after the rule-based processing step, it is still possible to have one semantic object segmented into multiple neighboring regions in $\mathbf{R}_t^O$. Moreover, some small unnecessary regions can appear due to motion estimation errors. A postprocessing step improves the segmentation result. On the contrary to splitting step in rule-based processing, regions are merged according to two criteria during the post-processing step. The first criteria is related to small regions: If
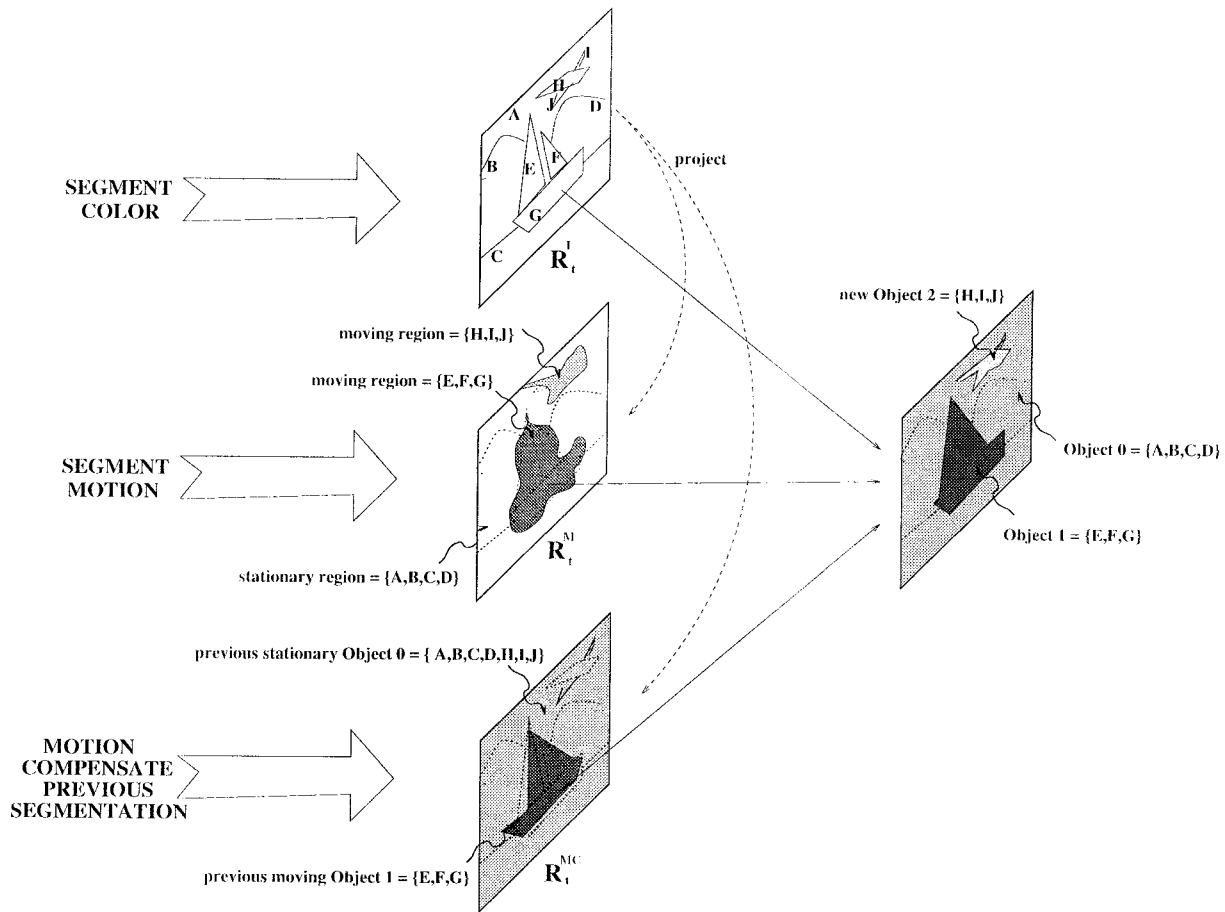
Fig. 4. A simple example for the proposed scheme.

the area of a region in $\mathbf{R}_t^O$ is smaller than a predefined threshold, then this region is merged with one of its neighbors to form the final segmentation, $\mathbf{R}_t^F$. If more than one neighboring region exists, the one with the same label and largest area is chosen. This leads to elimination of some small unnecessary regions. The second criteria tries to merge erroneously divided objects: If a region in $\mathbf{R}_t^O$ is moving and if it is a neighbor to any other moving region with a similar motion, these two regions are merged in $\mathbf{R}_t^F$. The step is necessary if an object is erroneously split in the previous (motion segmentation) steps.

F. Simulation Results

1) Mode 1—Moving Objects and Background Detection: Some exemplary results for the automatic segmentation of video sequences for video object generation, obtained by the described algorithm, are shown in Fig. 5. The quality of the segmented objects is quite good. The temporal coherency, i.e., the accuracy of the estimated objects in time, that cannot be shown in printed form, is also observed to be satisfactory for an automatic algorithm.

2) Mode 2—Detection of Differently Moving Objects: The simulation results for Mother & Daughter sequence are presented. Some typical results are shown in Fig. 6. In Fig. 6(a) and (b), the original fortieth and forty-fifth frames are shown, respectively. The color segmentation mask $\mathbf{R}_t^I$ of the original 45th frame for 256 regions, obtained using the RSST

method, can be found in Fig. 6(c). Fig. 6(d) shows the motion segmentation mask $\mathbf{R}_t^M$ based on the motion vectors, which are obtained using the HBM method between 40th and 45th frames. The previous segmentation mask is motion compensated by the estimated motion vectors to obtain $\mathbf{R}_t^{MC}$, which is shown in Fig. 6(e). The final segmentation result $\mathbf{R}_t^F$ in Fig. 6(f) is obtained after processing the four segmentation masks using the stated rules and the postprocessing step. In these frames, the daughter begins to move and it is denoted as a newly exposed object, whereas the mother is being tracked as an object which continues its motion. It should be noted that the boundaries of both objects are accurate.

G. Discussion on COST Analysis Model

The COST Analysis Model can be summarized as merging of small regions with color coherence according to their motion content at previous and current time instants. Previous merge-based methods use either color or motion or two of them together during all stages of their merging process [26], [29], [21], [27]. In the proposed scheme, while an RSST algorithm first merges regions using only the input color information, after some point only motion information is utilized to continue to the merging process. Such a novel approach usually results with accurate object boundaries at the correct locations.
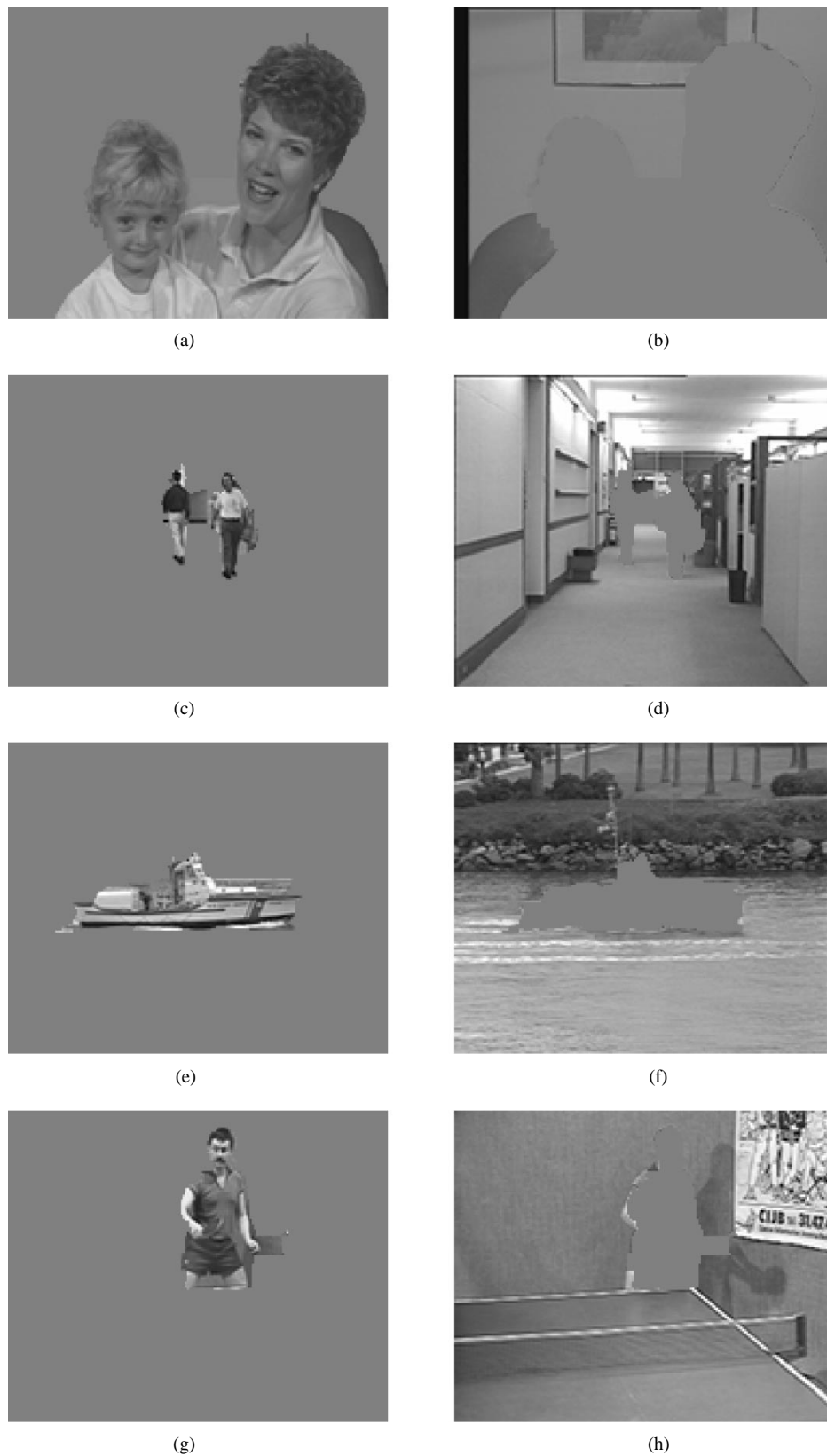
Fig. 5. Exemplary segmentation results for mode 1: (a), (b) mother & daughter, (c), (d) hall-monitor, (e), (f) coastguard, and (g), (h) table-tennis.

As long as a color region does not overlap with two different semantic objects, the boundaries of the segmentation result are expected to be accurate. However, occlusions might cause problems in some cases. When an occluding region, which is obviously neighboring to the moving object, corresponds to a small section of a smoothly varying part of the scene, this occluding region will not be included in the moving object. Since the majority of the points in this smoothly
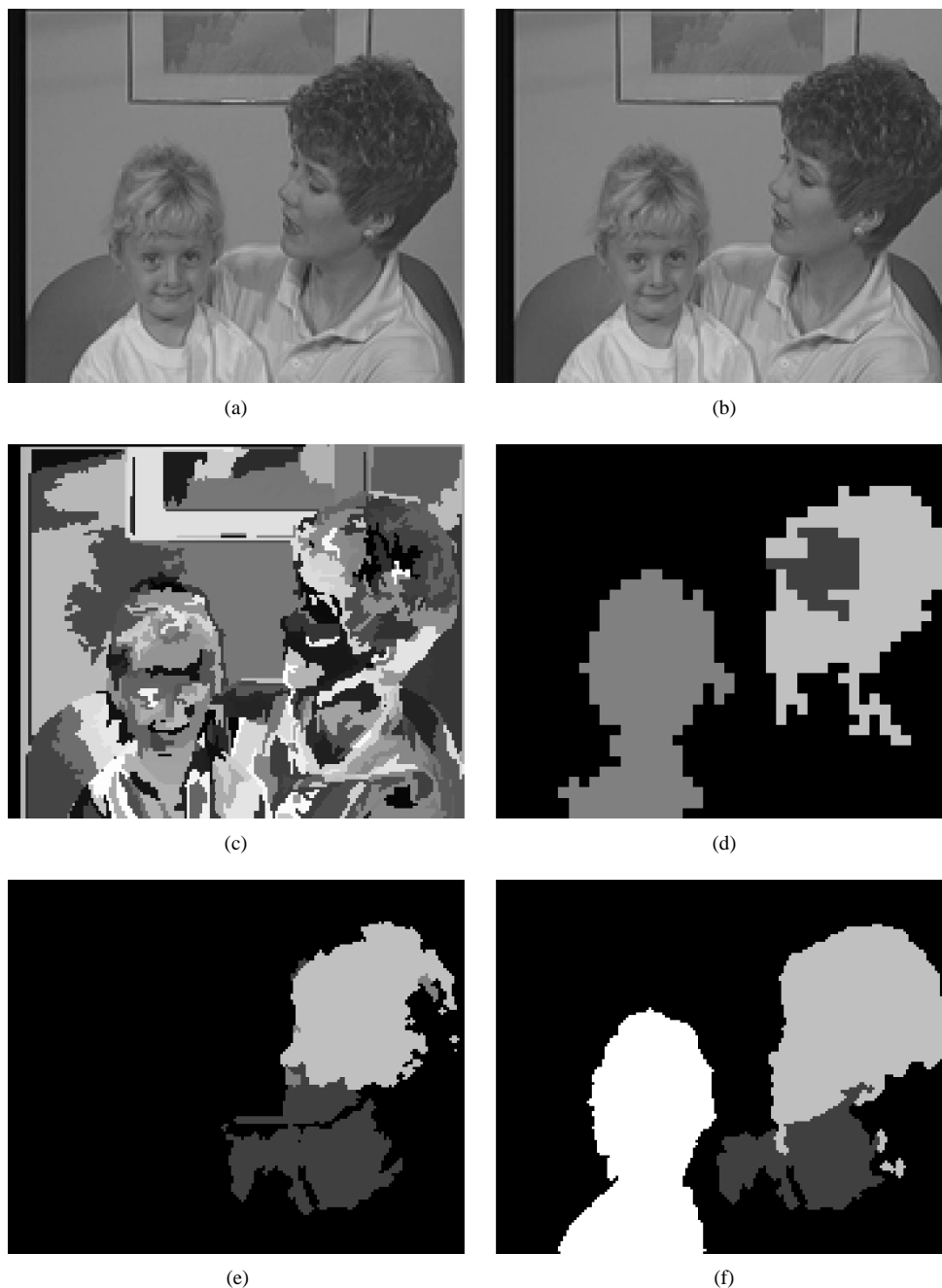
Fig. 6. Exemplary results for mode 2: original (a) fortieth and (b) forty-fifth frames of the mother & daughter sequence. (c) $\mathbf{R}_t^I$, (d) $\mathbf{R}_t^M$, (e) $\mathbf{R}_t^{MC}$, and (f) $\mathbf{R}_t^F$.

varying region will usually belong to a stationary background, occluding points are included in a stationary object and hence, they are not merged with that object. However, if there is a textured area in the scene corresponding to the occlusion area, it is possible to handle that part as a newly exposed object. Eventually, the algorithm removes such erroneous objects from the scene, after they stay stationary more than some predetermined duration of time.

IV. CONCLUSION

The described algorithms concentrate on segmentation of video sequences into objects, and tracking of these objects

in time. The chosen approach for the segmentation is to preanalyze the video sequence, and then in a second step, to evaluate the results from the preanalysis in a so-called rule processor. Thus, the fusion of different kinds of information is the essential feature. Currently, color, motion, and intensity changes are used; however, it is easy to extend the algorithm by adding new sources of information. Having a structure which is independent of the features of the analyzed input sequence, the AM model gives fully automated segmentation results. Even though the Analysis Model is pretty new, the presented results already show promising performance for various test sequences with different complexities.

Although the presented AM requires no user interaction and thus provides automatic segmentation, the KANT scenario allows user interaction. It is expected that user interaction will improve the segmentation results, for example, by pointing a region for analysis instead of the entire frame. User interaction might not be needed for rather simpler sequences. However, it may be unavoidable in case of complex scenes. The effect of user interaction and the subsequent improvements in segmentation quality are not examined in this paper. Rather, the goal is to find out the performance of a fully automated approach.

The presented results are the outcome of a voluntary collaboration of many companies and institutions throughout Europe within COST 211. Concentration of these efforts within a collaborative framework allows efficient transfer of knowledge during the precompetitive period of research, and thereby increases the expertise in this area. Keeping in mind the great success of the COST work items related to H.261 and H.263 in the past, one can conclude that the COST Analysis Model could develop into an important milestone in the long history of video-related research. The current algorithms will be further improved in a collaborative way, within COST 211^{quat}.

REFERENCES

[1] L. Chiariglione, "MPEG and multimedia communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 5–18, Feb. 1997.
[2] M. Hotter, R. Mester, and F. Muller, "Detection and description of moving objects by stochastic modeling and analysis of complex scenes," *Signal Processing: Image Communication*, vol. 8, pp. 281–293, 1996.
[3] D. Zhong and S. Chang, "Spatio-temporal video search using the object based video representation," in *Proc. 1997 IEEE Int. Conf. Image Processing ICIP-97*, Oct. 1997, vol. I, pp. 21–24.
[4] R. Oten, R. J. P. de Figueireda, and Y. Altunbasak, "Simultaneous object segmentation, multiple object tracking and alpha map generation," in *Proc. 1997 IEEE Int. Conf. Image Processing ICIP-97*, Oct. 1997, vol. I, pp. 69–73.
[5] P. Csillas and L. Böröczky, "Iterative motion-based segmentation for object-based video coding," in *Proc. 1997 IEEE Int. Conf. Image Processing ICIP-97*, Oct. 1997, vol. I, pp. 73–76.
[6] F. Morier, J. Benois-Pineau, D. Barba, and H. Sanson, "Robust segmentation of moving image sequences," in *Proc. 1997 IEEE Int. Conf. Image Processing ICIP-97*, Oct. 1997, vol. I, pp. 719–722.
[7] J. R. Morros and F. Marques, "Stable segmentation-based coding of video sequences addressing content-based functionalities," in *Proc. 1997 IEEE Int. Conf. Image Processing ICIP-97*, Oct. 1997, vol. II, pp. 1–4.
[8] P. de Smet and D. DeVleeschauwer, "Motion-based segmentation using a threshold merging strategy on watershed segments," in *Proc. 1997 IEEE Int. Conf. Image Processing ICIP-97*, Oct. 1997, vol. II, pp. 490–494.
[9] C. Gu and M. Lee, "Semantic video object segmentation and tracking using mathematical morphology and perspective motion mode," in *Proc. 1997 IEEE Int. Conf. Image Processing ICIP-97*, Oct. 1997, vol. II, pp. 514–517.
[10] F. Moscheni and S. Bhattacharjee, "Region merging for spatio-temporal segmentation," in *Proc. 1996 IEEE Int. Conf. Image Processing ICIP-96*, Sept. 1996, vol. I, pp. 501–504.
[11] C. K. Cheong and K. Aizawa, "Structural motion segmentation based on probabilistic clustering," in *Proc. 1996 IEEE Int. Conf. Image Processing ICIP-96*, Sept. 1996, vol. I, pp. 505–508.
[12] Y. Lin, Y. Chen, and S. Y. Kung, "Object-based scene segmentation combining motion and image cues," in *Proc. 1996 IEEE Int. Conf. Image Processing ICIP-96*, Sept. 1996, vol. I, pp. 957–960.

[13] S. Siggelkow, R. Grigat, and A. Ibenthal, "Segmentation of image sequences for object oriented coding," in *Proc. 1996 IEEE Int. Conf. Image Processing ICIP-96*, Sept. 1996, vol. II, pp. 477–480.
[14] J. W. Park and S. U. Lee, "Joint image segmentation and motion estimation for low bit rate video coding," in *Proc. 1996 IEEE Int. Conf. Image Processing ICIP-96*, Sept. 1996, vol. II, pp. 501–504.
[15] M. Schultz and T. Ebrahimi, "Matching-error based criterion of region merging for joint motion estimation and segmentation techniques," in *Proc. 1996 IEEE Int. Conf. Image Processing ICIP-96*, Sept. 1996, vol. II, pp. 509–512.
[16] E. Chalom and V. M. Bove, "Segmentation of an image sequence using multi-dimensional image attributes," in *Proc. 1996 IEEE Int. Conf. Image Processing ICIP-96*, Sept. 1996, vol. II, pp. 525–528.
[17] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 19–31, Feb. 1997.
[18] MPEG Group, "Overview of the MPEG-4 version 1 standard," ISO-/IEC JTC1/SC29/WG11, Doc. no. 1909, Oct. 1997 [Online]. Available WWW: http://wwwam.HHI.DE/mpeg-video/standards/mpeg-4.htm
[19] MPEG-7 Requirements Group, "MPEG-7 context and objectives," ISO/IEC JTC1/SC29/WG11, Doc. no. 1920, Oct. 1997 [Online]. Available WWW: http://drogo.cselt.stet.it/mpeg/standards/mpeg-7.htm
[20] A. M. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[21] M. Hotter and R. Thoma, "Image segmentation based on object oriented mapping parameter estimation," *Signal Processing*, vol. 15, pp. 315–334, 1988.
[22] R. Mech and M. Wollborn, "A noise robust method for segmentation of moving objects in video sequences," in *Proc. 1997 IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP 97*, Apr. 1997.
[23] R. Mech and M. Wollborn, "A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera," in *Proc. Workshop on Image Analysis for Multimedia Interactive Services WIAMIS'97*, Louvain-la-Neuve, Belgium, June 1997.
[24] R. Mech and P. Gerken, "Automatic segmentation of moving objects (partial results of core experiment N2)," Doc. ISO/IEC JTC1/SC29/WG11 MPEG97/1949, Bristol, U.K., Apr. 1997.
[25] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. Image Processing*, vol. 3, pp. 625–638, Sept. 1994.
[26] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, pp. 384–402, July 1985.
[27] D. W. Murray and B. F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 220–228, Mar. 1987.
[28] G. D. Borshukov, G. Bozdagi, Y. Altunbasak, and A. M. Tekalp, "Motion segmentation bu multistage affine classification," *IEEE Trans. Image Processing*, vol. 6, pp. 1591–1594, Nov. 1997.
[29] M. Chang, M. I. Sezan, and A. M. Tekalp, "A bayesian framework for combined motion estimation and scene segmentation in image sequences," in *Proc. 1994 IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP 94*, 1994, pp. 221–224.
[30] M. M. Chang, A. M. Tekalp, and M. I. Sezan, "Simultaneous motion estimation and segmentation," *IEEE Trans. Image Processing*, vol. 6, pp. 1326–1333, Sept. 1997.
[31] A. A. Alatan, E. Tuncel, and L. Onural, "A rule-based method for object segmentation in video sequences," in *Proc. 1997 IEEE Int. Conf. Image Processing ICIP 97*, Oct. 1997, vol. II, pp. 522–525.
[32] A. Alatan, E. Tuncel, and L. Onural, "Object segmentation via rule-based data fusion," in *Proc. Workshop on Image Analysis for Multimedia Interactive Services WIAMIS'97*, Louvain-la-Neuve, Belgium, June 1997, pp. 51–56.
[33] R. M. Haralick and L. G. Shapiro, "Survey: Image segmentation techniques," *Computer Vision, Graphics and Image Processing*, vol. 29, pp. 100–132, 1985.
[34] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 641–647, June 1994.
[35] O. J. Morris, M. J. Lee, and A. G. Constantinides, "Graph theory for image analysis: An approach based on the shortest spanning tree," *Proc. Inst. Elect. Eng.*, vol. 133, pp. 146–152, Apr. 1986.
[36] R. C. Gonzales and R. E. Woods, *Digital Image Processing*. Norwood, MA: Addison-Wesley, 1992.
[37] M. J. Biggar, O. J. Morris, and A. G. Constantinides, "Segmented-image coding: Performance comparison with the discrete cosine transform," *Proc. Inst. Elect. Eng.*, vol. 135, pp. 121–132, Apr. 1988.
[38] M. Bierling, "Displacement estimation by hierarchical block matching," in *Proc. SPIE Visual Communications and Image Processing 88*, 1988, pp. 942–951.

[39] T. Aach, A. Kaup, and R. Mester, "Statistical model-based change detection in moving video," *Signal Processing*, vol. 31, no. 2, pp. 165–180, Mar. 1993.

[40] T. Aach, A. Kaup, and R. Mester, "Change detection in image sequences using Gibbs random fields: A Bayesian approach," in *Proc. Int. Workshop on Intelligent Signal Processing and Communication Systems*, Sendai, Japan, Oct. 1993, pp. 56–61.

A. Aydın Alatan (S'91–M'97) was born in Ankara, Turkey, in 1968. He received the B.S. degree from Middle East Technical University, Ankara, Turkey, in 1990, the M.S. and DIC degrees from Imperial College of Science, Medicine and Technology, London, U.K., in 1992, and the Ph.D. degree from Bilkent University, Ankara, Turkey, in 1997, all in electrical engineering. He was a British Council scholar in 1991 and 1992.

He joined the Center for Image Processing at the ECSE Department of Rennselaer Polytechnic Institute in 1997, where he currently is a Postdoctoral Research Associate. He current research interests are image/video compression, object-based coding, motion analysis, 3-D motion models, nonrigid motion analysis, Gibbs models, rate-distortion theory, active meshes, and robust video transmission over ATM networks.

Levent Onural (S'82–M'85–SM'91) received the B.S. and M.S. degrees in electrical engineering from Middle East Technical University, Ankara, Turkey, in 1979 and 1981, respectively, and the Ph.D. degree in electrical and computer engineering from the State University of New York at Buffalo in 1985. He was a Fulbright scholar between 1981 and 1985.

After holding a Research Assistant Professor position at the Electrical and Computer Engineering Department of the State University of New York at Buffalo, he joined the Electrical and Electronics Department of Bilkent University, Ankara, Turkey, where he is currently a full Professor. He visited the Electrical and Computer Engineering Department of the University of Toronto on a sabbatical leave from September 1994 to February 1995. His current research interests are in the area of image and video processing, with emphasis on very low bit rate video coding, texture modeling, nonlinear filtering, holographic TV, and signal processing aspects of optical wave propagation.

Dr. Onural was the organizer and first chair of the IEEE Turkey Section; he served as the chair of the IEEE Circuits and Systems Turkey Chapter from 1994 to 1996, and as the IEEE Region 8 (Middle East and Africa) Student Activities Committee Chair. He is currently the IEEE Regional Activities Board Vice Chair in charge of student activities. In 1995, he received the Young Investigator Award from TUBITAK, Turkey.
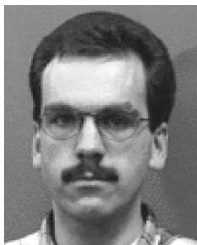
Michael Wollborn received the Dipl.-Ing. degree in electrical engineering from the University of Hannover, Hannover, Germany, in 1991.

Since 1991, he has been with the Institute für Theoretische Nachrichtentechnik und Informationsverarbeitung at the University of Hannover, Germany, where he works in the area of video analysis and coding, in particular at low bit rates. During his work he was involved in several projects on these topics, supported either by industrial partners, by research centers such as the Deutsche Forschungsgemeinschaft (DFG), or by the European Community (EU). In his current EU project, he is leading a work group on the development of algorithms for the new ISO/MPETG-4 multimedia coding standard. His present research interests and activities cover improved motion estimation and compensation for video coding and image sequence analysis, in particular detection and shape estimation of moving objects. He is a member of the European COST 211 group, and from the very beginning he was actively contributing to and involved in the development of the ISO/MPEG-4 standard.

Roland Mech received the Dipl.-Inform. degree in computer science from the University of Dortmund, Dortmund, Germany, in 1995.

Since 1995 he has been with the Institute für Theoretische Nachrichtentechnik and Informationsverarbeitung at the University of Hannover, Germany, where he works in the areas of image sequence analysis and image sequence coding. He is a member of the simulation subgroup of the European project COST 211 and contributes actively to the ISO/MPEG-4 standardization activities. His present research interests and activities cover image sequence analysis, especially 2-D shape estimation of moving objects, and the application of object-based image coding.

Ertem Tuncel was born in Antalya, Turkey, in 1974. He received the B.Sc. degree from Middle East Technical University, Ankara, Turkey, in 1995 and the M.Sc. degree from Bilkent University, Ankara, Turkey, in 1997. He is currently pursuing the Ph.D. degree at the Department of Electrical and Computer Engineering at the University of California, Santa Barbara.

Thomas Sikora (M'93–SM'96), for a photograph and biography, see this issue, p. 800.