VIDEO SEGMENTATION USING SPATIO-TEMPORAL INFORMATION

Yong-Wan Kim  Yo-Sung Ho
Kwangju Institute of Science and Technology
572 Ssangam-dong Kwangsan-gu, Kwangju, 506-712, Korea
herley@gob.kjist.ac.kr

ABSTRACT

This paper presents a new segmentation algorithm that processes segmentation by combining spatial segmentation and motion information. This algorithm uses morphological filters and watershed algorithm. The proposed segmentation algorithm consists of six parts: preprocessing, image simplification, feature extraction, decision, region merging and postprocessing. By combining spatial and temporal information intelligently, we can get visually meaningful segmentation results. Simulation results show the efficiency of the proposed technique.

1. INTRODUCTION

As video processing techniques have been advanced, a new functionality of content-based interactivity is needed to allow users flexible manipulation of video objects. This functionality is useful for many applications, such as the authorization system at video studio and the digital audio-visual multimedia database, and so on.

Growing needs for multimedia services, such as information retrieval from multimedia databases, interactive TV, videoconferencing applications, invoke an increasing efforts for low bit rate coding to allow easy manipulation of the content of a moving scene. Since the popular MPEG-1 and MPEG-2 can not provide such functionalities, a special attention is paid now to development of coding methods based on the analysis-synthesis approach.

Recently, the MPEG group has been working on establishing a new standard for low-rate audio-video coding. In object-based coding, we extract objects in the video sequence, find relationship between them, and process them based on their property for supplying content-based interactivity for the users.

Object-based algorithms segment image contents into a set of objects or regions according to a given model, and estimate their parameters (e.g., color, shape, and motion) that can then be encoded. The segmentation masks form the base for content-based functionalities as envisaged by MPEG-4. Segmentation is an indispensable tool for such an approach. Thus several segmentation algorithms were proposed to detect moving objects.

M. Hörster proposed an interactive approach in which an image is first divided into unchanged and changed areas. Each connected changed area defines an object: and motion parameters describing the motion of each object are computed. Within each object, regions of connected pixels whose motion is not well described by the estimated motion parameters are detected. The algorithm then proceeds recursively by estimating the motion of these objects [2].

C. Advíl proposed an segmentation algorithm based on optical flow field [8]. However, this algorithm has huge amount of computational load and this model requires explicit constraints such as smoothness.

P. Salembier proposed a segmentation algorithm using morphological tools, such as morphological filters and watershed [5]. However, the segmentation produced by only spatial information or motion information may have false contours. In some cases, background and foreground may be merged [5]. Recently, an attempt has been made to use the information both from spatial and temporal domains which results in a more meaningful segmentation for perception [4,9]. For these reasons, we propose a new efficient spatio-temporal segmentation to get more meaningful and accurate segmentation results. Proposed segmentation algorithm consists of six parts: preprocessing, image simplification, feature extraction, decision, region merging and postprocessing. We propose change detection as preprocessing to prevent the foreground and the background merging maximally and postprocessing to eliminate large shape information parts with small texture information. By combining the spatial and the temporal information intelligently, we can get visually meaningful segmentation results.

2. SPATIO-TEMPORAL SEGMENTATION

Our proposed segmentation algorithm consists of six parts, as shown in Fig.1.

1) Change detection
2) Image simplification
3) Feature extraction
4) Boundary decision
5) Region merging
6) Postprocessing

At first, change detector differentiates changed and unchanged regions. Segmentation algorithm applies only the changed regions. Changed regions are considered as one object, and not carried out any operations. The unchanged regions are simplified by morphological filtering. A marker extraction is a seed which is coherent in both motion and
luminance. Region boundaries are decided by a watershed algorithm that incorporates motion and luminance information simultaneously. With this result, regions having similar motions are merged. Finally, the result after region merging is postprocessed to eliminate redundant parts.

2.1. Change Detection

The conventional segmentation algorithm produced only by the spatial information or temporal information may have false contours. In some cases, background and foreground may be merged. The recent attempt has been made to use information both from the spatial and the temporal domains, which results in a more meaningful segmentation. However, such conventional spatio-temporal segmentation algorithms have the probabilities to be merged foreground into background. Thus we propose the change detection as a preprocessing step for segmentation to prevent this probability maximally. The unchanged region is considered as one object of the background and the segmentation is applied only on the changed region.

In the benefit of change detection, we can reduce the amount of computational load in a large video format such as CCIR601.

2.2. Image Simplification

In this step, image is simplified to make it easier to segment. The simplification controls the nature and amount of information that is kept for segmentation at this level of hierarchy.

A large number of morphological tools rely on two basic sets of transformations, known as erosion and dilations. If \( f(x) \) denotes an input signal and \( M_n \) a window (or a flat structuring element) of size \( n \), the erosion and dilation by \( M_n \) are given by:

Erosion: \( \delta_n (f)(x) = \text{Min}\{f(x - y), y \in M_n\} \)

Dilation: \( \epsilon_n (f)(x) = \text{Max}\{f(x + y), y \in M_n\} \)

The second set of erosions and dilations involves geodesic transforms. They are always defined with respect to a reference function \( r \). The geodesic dilation of size one is defined as the minimum between the dilation of size one of the original function \( f \) and the reference function \( r \). The geodesic erosion is defined by duality:

Geodesic dilation of size 1: \( \delta^{(1)}(f,r) = \text{Min}\{\delta_1 (f), r\} \)

Geodesic erosion of size 1: \( \epsilon^{(1)}(f,r) = -\delta^{(1)}(-f, -r) \)

Geodesic dilations and erosions of arbitrary size are defined by iterations. For example, the geodesic dilation of infinite size, also called reconstruction by dilation (by erosion), is given by:

Reconstruction by dilation:
\[
\gamma^{(\text{rec})}(f,r) = \delta^{(\text{rec})}(f,r) = \delta^{(1)}(\cdots \delta^{(1)}(f,r) \cdots, r)
\]

Reconstruction by erosion:
\[
\varphi^{(\text{rec})}(f,r) = \epsilon^{(\text{rec})}(f,r) = \epsilon^{(1)}(\cdots \epsilon^{(1)}(f,r) \cdots, r)
\]

Opening by reconstruction:
\[
\gamma^{(\text{rec})}(\epsilon_n(f), f)
\]

Closing by reconstruction:
\[
\varphi^{(\text{rec})}(\delta_n(f), f)
\]

The most popular filter by reconstruction is the opening by reconstruction of erosion. Of course, duality, a closing can be defined:

Open-Close by reconstruction:
\[
\varphi^{(\text{rec})}(\delta_n(\gamma^{(\text{rec})}(\epsilon_n(f), f)), \gamma^{(\text{rec})}(\delta_n(f), f))
\]

These filters have a simplification effect on the signal, but preserve the contour information [5].

2.3. Feature Extraction

The goal of this step is to detect the presence of homogeneous regions. It produces markers identifying the interior of the regions that will be segmented. The marker defines the set of pixels that surely belong to the region and, in general, it defines the major part of the region interior. For size-oriented segmentation, the marker extraction consists of labeling the interior of large flat zones after the simplification we detect the spatial and temporal markers indivis-
2.4. Boundary Decision

After feature extraction, however, a large number of pixels are not yet assigned to any region. These pixels correspond to uncertainty areas mainly concentrated around the contours of the regions. Assigning these pixels to a given region can be viewed as a decision process that precisely defines the partition. The classical morphological decision tool is the watershed. It is generally used on the morphological gradient of the image to segment. However, we use a different version of the watershed algorithm working on the original signal [7]. This watershed algorithm is a region growing process. The set of markers is extended until they occupy all the available space. During the extension, pixels of the uncertainty areas are assigned to a given marker. Watershed algorithm works in two distinct steps: queue initialization and region growing. The initialization consists of putting the location of all pixels corresponding to the interior of the regions in the queue of highest priority. The region growing consists of extracting a pixel from the queue. If the pixel does not yet belong to a region, because of the filling procedure, it has at least one neighboring region. Therefore, a distance(similarity) between the current pixel and each neighboring region is computed. The pixel is assigned to the region corresponding to the smallest distance. Then, if the current pixel has some neighbors that do not belong to any region, these neighbors are put in the queue with a priority. As can be seen, any pixel that is put in the queue has at least one neighboring region.

2.5. Region Merging

The segmentation result in boundary decision may have different luminance properties, but may have the same motion parameter set. Therefore, elimination of redundant regions is needed for coding since fewer regions permit a smaller bit rate allocated to the transmission of region shapes and motion parameters. This step corresponds to the motion-based region merging [5].

2.6. Postprocessing

The result in region merging may have large shape information regions with small texture information. These parts should be eliminated in MPEG-4 coding framework. Therefore, we propose the postprocessing to eliminate these parts using morphological opening and closing.

3. SIMULATION RESULTS

Experiments have been carried out on successive frames of sequences of QCIF format, “Miss America” and “Claire” Fig.(a) and Fig.(b) show the test frames and Fig.(c) is the segmentation result by M. Hötter. And Fig.(d) is the result before the region merging. Fig.(e) is the result after the region merging. Fig.(f) is the final segmentation result by the proposed scheme.

Figures from the simulation results, the proposed scheme gives more meaningful and accurate shape information and segmentation regions rather than that of M. Hötter. In addition: we see the increase of coding efficiency by postprocessing. This spatio-temporal approach is well suited to coding application: such as object- and region-based coding.

4. CONCLUSIONS

Segmentation is an indispensable tool for content-based coding algorithm. Therefore, we have presented a new segmentation that processes segmentation by combining spatial and motion information. By combining spatial and temporal information intelligently, we can get visually meaningful segmentation results. We have presented an unsupervised spatio-temporal segmentation techniques. This algorithm is mainly devoted to very low bit rate video coding applications and no assumption is made about the sequence content. Experimental results demonstrate the effectiveness of the proposed method. This method can be well suited for low bit rate video coding.
REFERENCES


