ABSTRACT

We present a novel motion-based video indexing scheme for fast content-based browsing and retrieval in a video database. The proposed technique constructs a dictionary of prototype objects to support query by motion. The first step in our approach extracts moving objects by analyzing layered images constructed from the coarse data in a 3-D wavelet decomposition of the video sequence. These images capture motion information only. Moving objects are modeled as collections of interconnected rigid polygonal shapes in the motion sequences that we derive from the wavelet representation. The motion signatures of the object are computed from the rotational and translational motions associated to the elemental polygons that form the objects. These signatures are finally stored as potential query terms.

1. INTRODUCTION

The fundamental step in content-based multimedia database system is to index individual multimedia objects with searchable compactly represented descriptors [1]. Well organized indexing structures facilitate the fast and practical browsing and direct access to the desired data objects in response to object-oriented searching. Such indexing procedure can be constructed by creating semantic descriptors based on information that identify and represent each multimedia object model. Therefore, to define and extract a single or multiple key features from each object data is an essential task in indexing construction. The motion feature is an important characteristic contained in video data. This unique property differentiates the dynamic video data from the static imagery data.

In this paper, we propose a method to index the motion features of video objects for content-based searching and browsing. To attack this task, our proposed scheme performs the motion analysis of video objects in multiresolution/multiscale manner. We extract moving objects by analyzing layered images constructed from the coarse data in a 3-D wavelet decomposition of the video sequence. These images capture motion information only. Next, we fit a polygon to each connected region in the motion sequence (the wavelet domain images which capture the motion information). The rotational and translational motions associated to the center of gravity of these polygons are extracted. Finally, these motions are stored and indexed using a progressive resolution approach similar to [10]. Specifically, we combine a VQ approach with a coarse wavelet representation of the motion data to perform indexing and retrieval.

Most previous motion indexing approaches are based on global motions at frame level (e.g., camera zooming, panning and tilting). Although useful for some purpose, it is hard to investigate the local motion properties of video objects. Recently, some object-based motion indexing techniques have been presented [2, 3, 4, 5]. A variety of motion features are derived by several different methods. The motion vectors have been most widely adopted as an indexing clue for representing the motion properties. Traditionally, to examine camera operations, the motion vectors are computed by the optical flow, block-matching method and the Hough transform [6, 7, 8, 9]. Unfortunately, these procedures have a drawback of computational complexity, especially on processing of a long video data.

Note that our approach, like that of [5], relies ultimately on a low resolution wavelet representation of the motion for indexing and retrieval. Unlike the approach of [5], we extract moving objects by analyzing layered images constructed from the coarse data in a 3-D wavelet decomposition of the video sequence. This leads to a simple clue for representing the motion properties. Traditionally, to examine camera operations, the motion vectors are computed by the optical flow, block-matching method and the Hough transform [6, 7, 8, 9]. Unfortunately, these procedures have a drawback of computational complexity, especially on processing of a long video data.

2. PROGRESSIVE SPATIO-TEMPORAL SEGMENTATION OF MOVING OBJECTS

2.1. Progressive Resolution Analysis of a Video

The first step in the object-oriented video indexing is to locate a particular VO (video object) of interest. The VOs are carried by video shot that is regarded as a basic temporal segment unit of a long video sequence. We decompose a continuous video sequence into isolated shots using the temporal segmentation method proposed in [11]. In this method, shot boundaries are determined by examining the coarse data in a 3-D wavelet transform of the video track. To reduce the computational complexity (especially for long video sequences), we begin by computing coarse reduced frames from a 2-D wavelet transform of a temporally sub-sampled video sequence. Through this spatio-temporal progressive analysis scheme, a set of isolated video shots are efficiently decomposed from video sequences. We extend this 3-D wavelet transform of a video data to locate and segment the motion contour of video objects within
The wavelet transform can provide a coarse representation of multiple moving objects with layers. Intuitively, we expect the motion sequence corresponding to the most rapid moving object to appear in the higher successive subbands.

We begin by separating prominent moving objects from the effect of global camera motions (e.g., zoom and pan) by combining the approach of [3] with our multiresolution motion decomposition. Next, we construct multi-layered image sequences [12] $Layer_k(m, n, t)$ of multiple moving objects using the direct temporal correlation of wavelet transform at several adjacent subbands image sequences $W_l(m, n, t)$ in time axis as follows:

$$Layer_k(m, n, t) = \prod_{l=1}^{j} W_l(m, n, t), \quad i, j = 1, 2, 3, \ldots$$

Here $k$ is the layer number, $l$ is the number of temporal subband and $i \leq j$. By examining the layered images, we can segment the video sequence into the individual moving objects. We tested this technique on the TableTennis sequence. Figure 3 shows that two moving objects (ball and a player’s arm) are effectively classified as different objects based on associated motion amounts.

3. MOTION ANALYSIS OF OBJECTS

3.1. Rigid Moving Object

In this section, we describe a model for temporal variations of the spatial intensity pattern on the motion sequence. Generally, a 3-D moving object is analyzed by projecting its time-varying activities onto 2-D image plane. The motion characteristics, such as translation and rotation, of a moving object can be represented as follows,

$$\vec{v}_{l+1} = A \vec{v}_l + \vec{d},$$

where $A$ contains a scale and rotation factor, and $\vec{d}$ is a displacement vector ($\vec{d} = [\Delta x \Delta y]^T$). In order to obtain these motion parameters, we first compute the motion gravity from the intensity values within the area of motion contour. Such a motion gravity is located by a method similar to [13]. Then, a polygon (e.g., ellipse or rectangle) is fitted to the motion contour of each object of interest. An object encompasses the distribution of the motion intensity around the centroid of motion (motion gravity). The major axis of such a polygon is simply derived based on the relative position between the furthest intensity pixel and the centroid (Figure 2). The size of a polygon may be changed in each successive frames according to the amount and distribution of the motion intensities. However, because we focus on the positions of a motion centroid and the major axis, the time-varying size of a polygon has no effect on our motion feature analysis. Recall that all our methods of identifying the motion contour and deriving the fitting polygon are performed on the reduced coarse image sequence computed by a 3-D wavelet transform of a video data. Thus, we can achieve the highly reduced computational complexity in even a large video database.

3.2. Multiresolution Analysis of Motion Features

From the location of a centroid and axes of a fitting polygon, we compute the motion indexing features representing
the motion behavior of objects. First, an indexing feature is extracted from translational motion \( \mathbf{d} \). The position tracking of a centroid effectively reflects the trajectory of objects in the translational movement. In addition, from the position of the polygon's axes, it is possible to derive the rotational motion feature \( \mathbf{A} \) from the angle value of a moving rigid object around the hinge. Due to the inherent spatially coarse representation in our proposed approach, these motion signatures may be slightly inaccurate in describing the temporal motion behavior of objects. Furthermore, very slowly moving objects in a long video sequence need to be efficiently represented in a compact form (i.e., in shorter time-scale). For this purpose, we take a wavelet transform of \( \mathbf{A} \) and \( \mathbf{d} \) since it is capable of compactly represent their temporal variation at different scales and resolution. Note that we keep track of the lowest 3 scales. This provides a filtering of inaccurate motion vectors (Figure 4).

As explained earlier, the dictionaries are constructed in terms of these multiresolution subbands of motion features from video objects. Based on the dictionaries, progressive searching and matching to the query by motion is performed in an efficient manner.

4. EXPERIMENTAL RESULTS

We tested our proposed technique on two different categorized motion (translation and rotation) of video sequences. All image sequences are spatially filtered and subsampled by 2-D wavelet transform at 30 × 40 resolution from the original size of 320 × 240. Figure 3 shows the effectiveness of the first step of our procedure for detecting and classifying the changing parts of moving objects using the temporal 1-D wavelet transform. On the correlation between adjacent subbands, a rough classification between a ball and player’s arm is shown. In Figure 4, a vertically moving (up-and-down) ball is successfully captured by an ellipse centered on the gravity of motion intensity. A simple trajectory of gravity's position is also illustrated. Such a global view of moving object provides a user with effective and efficient cue for motion-based indexing of a large video database. The rotational movement of a particular object around a hinge is shown in Figure 5. A salesman’s ankle and a part of left upper-arm are captured by a rectangle centered on the motion intensity. From the rotational movement of the axes of this rectangle around the hinge (salesman’s elbow), the rotational motion parameter (i.e., the angle value) is computed without any additional work.

5. CONCLUSIONS

We presented a new object-based motion indexing scheme of video objects. Through its progressive multiresolution and multiscale properties, the proposed approach speeds up the process of constructing the compact description of moving objects and is capable of prompt responding the query by motion in an effective manner. Experimental results show the effectiveness of our proposed approach.

6. REFERENCES

Figure 3: Layered segmentation of multiple moving objects

Figure 4: Translational motion: tracking of a ball (left). Its smoothed version by wavelet analysis (right)

Figure 5: Rotational motion: rotational trajectory of left upperarm and its angle value