Module
11

VIDEO INDEXING AND RETRIEVAL

Lesson
39

Video Content Representation

Instructional objectives

At the end of this lesson, the students should be able to :

1. Define a video segment
2. List the primitive features used for video concept extraction
3. Define evaluation function for video segments.
4. Analyze time – series video contents based on curve distribution
5. Derive connection types from content categories.
6. **Outline the bounding box principle for video structuring**

39.0 Introduction

In lesson-38, we had studied the basic aspects of image retrieval using keys extracted from compressed images and also studied how to extract the keys from video sequences. We also discussed the dominant motion estimation for the purpose of scene structuring. We now discuss the schemes of video content representation, starting with the definition of video segments and obtaining connection types corresponding to the content categories. We also introduce the bounding box principle, a tool necessary for video structuring.

39.1 Video Segments

A video segment is a sequence of video shots concatenated by scene transitions (e.g. fade in / out, wipe, dissolve, etc). A meaningful scene may be represented as
$V = \{ v_i, v_{i+1}, \ldots\ldots, v_{i+r-1} \}$, where v_i is the starting frame of a video sequence with frame number i, and r is the duration of this segment.

From a video segment, the frame changing can occur in combination of primitive feature (s), such as colour, size, shape, and/or high level features, such as action and timing, used to describe objects in the video frame. After the image processing, annotation or media conversion process, a sequence of raw video data can be transformed into a variety of attribute values of text, or numerical data types with temporal extension. A specific point of view (abbreviated as a view) in a video sequence can be represented by a special projection of these features. The evaluation value of a specific view generated by domain knowledge can be a single real number obtained from the combinations of relevant features in a single video frame and this evaluation value is application dependent. For example, the evaluation value can be a weighted sum of relevant features, or some other formula specified by the users and / or domain knowledge.

39.2 Evaluation function

Definition 1: An *evaluation function* of a video segment V according to a specific view with *q* features is defined as

$$E\left(V, A, T_s, T_e\right) = \Delta\left(T_i\right)$$

where Δ is the formula of relevant feature vector combination; T_i is the time interval from starting frame T_s to ending frame T_e; $\boldsymbol{A} = [\, f_1\ f_2 \ldots\ f_q]$ is a set of features for the specific video view. We use E(*t*) that stands for single evaluation value at time t for a specific video segment and point of view.

39.3 Video content segmentation and indexing

Before we provide a segmentation strategy, we first examine several typical curve distribution which occur in time – series video content

(a) Stepwise

(b) Peak

(c) Smooth change

(d) Irregular

Fig.39.1 Some typical curve distributions : Frame to frame differences shown against the frames, (a) stepwise, (b) peak, (c) smooth change and (d) irregular transitions.

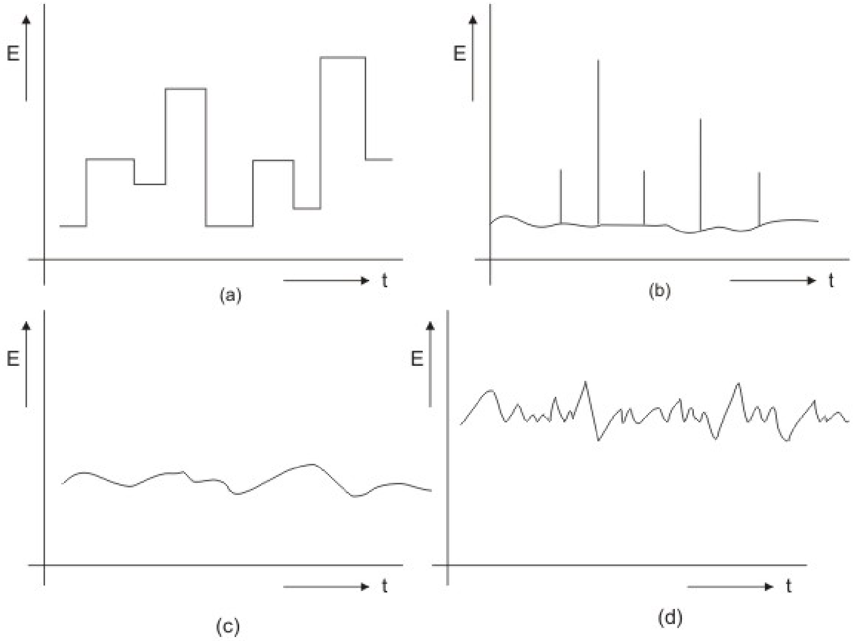

FIGURE 39.1 Some typical curve distributions; Frame to Frame differences shown against the frames, (a) stepwise (b) peak, (c) smooth change and (d) irregular

Fig.39.1 (a) shows the changing of semantic meaning in the video segment, or the variation time in the number of a certain object (e.g. cars on a street). Fig.39.1 (b) shows several large peaks, which represents the suddenly happened event (e.g. the frame difference in a cut detection process). Fig.39.1 (c) shows a situation of smooth changing (e.g. slow motion object or slow colour intensity change). Fig.39.1 (d) represents, the randomly distributed irregular curve (e.g. fast action)

According to the curve distribution, several kinds of curve features can be found. It is possible to classify the curve features into four categories – suddenly up edge, suddenly down edge, increase out of a range and decrease out of range in a curve.

We can derive seven connection types from these four categories: -

- **Connection Type – 1:** Large pulse when edge up and down happens in a short time period (eg 1/30 sec), as shown in fig.39.2.

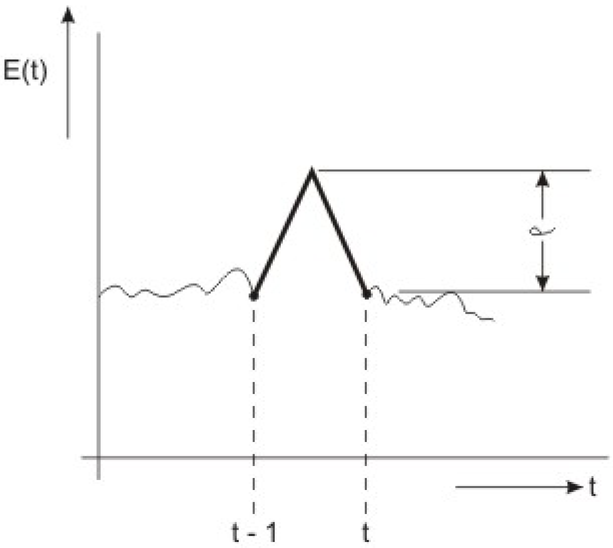

Fig. 39.2 Evaluation function vs. time for connection type 1.

- **Connection type 2,3:** Edge up/down as shown in fig.39.3:

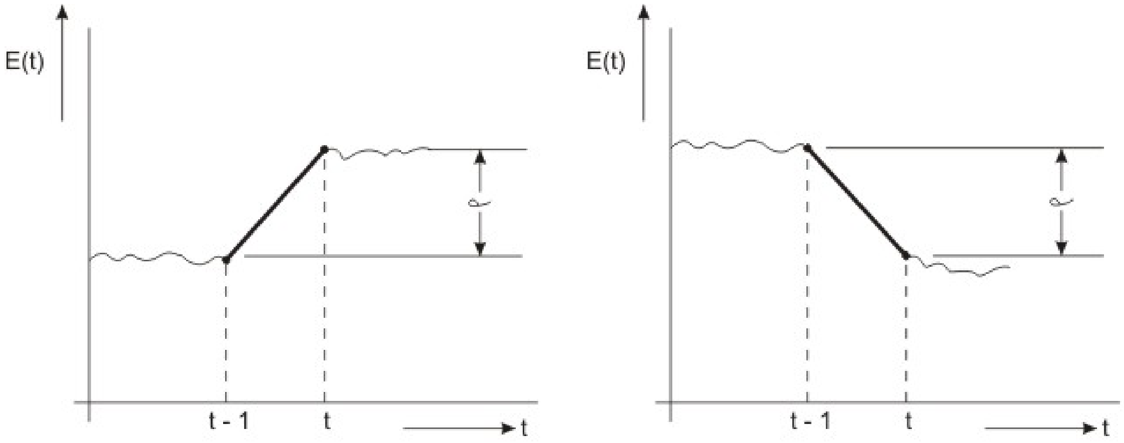

Fig. 39.3 Evaluation function vs. time for connection type 2 and type 3 (edge up/down).

$$E(t)-E(t-1)>\rho \text{ for edge-up}$$
$$E(t-1)-E(t)<-\rho \text{ for edge down}$$

where E (t-1) and e(t) are the evaluation function values at time t-1 and t, respectively.

- **Connection type 4, 5:** Increase / decrease, as shown in fig.39.4:

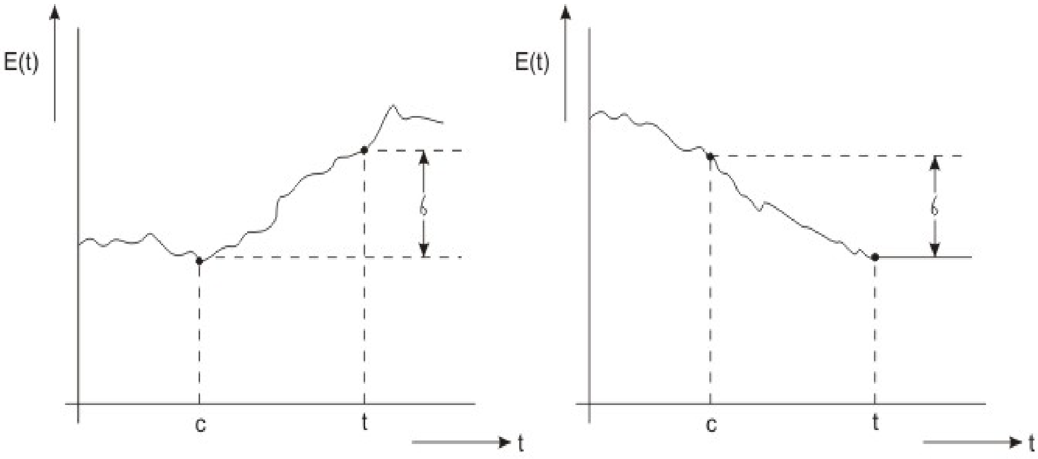

Fig. 39.4 Connection type 4 and type 5 (increase/decrease).

$$E(t)-\rho(c)>\sigma \text{ for increase}$$
$$E(t)-\rho(c)<-\sigma \text{ for decrease}$$

where, E (t) is the evaluation in function value at time t and P (C) is the value of current prominent point.

- **Connection type 6** : Long duration λ of steady situation, as shown in fig. 39.5.

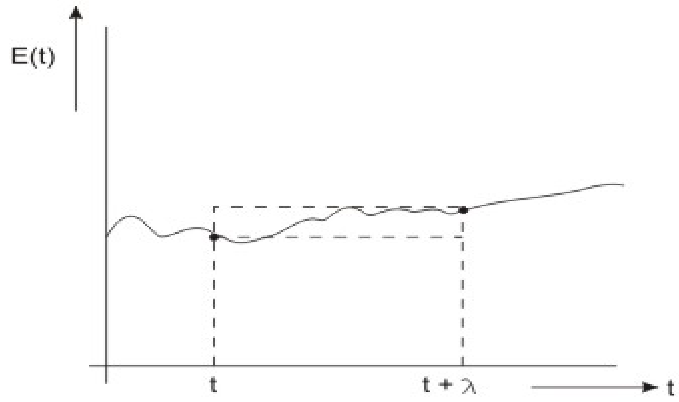

Fig. 39.5 Connection type 6.

Connection type – O is used for the unstable area, such as the starting and ending box.

39.4: Bounding Box Principle

Each video segment can be segmented into several structural units by a set of special evaluation values or semantic meanings.

In consequence of segmentation, the video subsequence between two special successive feature points can be separated and bounded by a bounding box. That is, the evaluation function of a video segment can be divided into a series of bounding boxes by special feature points. These special feature points are called prominent index points. Each segmented sequence is represented as a rectangular box with prominent point value and related information, like

- Box ID.
- Minimum and Max values
- Offset of duration (box length)
- Interbox connection type.

- Starting frame / time number
- Density information of box
- Previous and next subsequence linkages
- High level semantic meaning

39.5 Prominent index point

Definition-2: A *prominent index point* of evaluation function at time t is the point P that satisfies at least one of the following conditions:

- $|E_t - E_{t-1}| > \rho$ where, E_t is the current evaluation value at point t.
- $|E_t - P_c| > \sigma$ where, P_c is the current prominent point at time/frame C
- The time / frame difference between the previous prominent point and the current evaluation point is greater than a threshold λ

39.6 Video indexing

Chang and Lee used a B-tree as the index structure. For each new bounding box, inserting a new prominent point in the index tree is done by a searching index tree and adding the prominent point in a node. The related information about this bounding box is started in the storage space of a corresponding link list structure and can be accessed through a link list pointer accompanied with the prominent point in the leaf mode. By using the linked list, we can easily find the bounding boxes with similar prominent points and similar box shapes.