AN INDEXING AND SEARCHING STRUCTURE FOR MULTIMEDIA DATABASE SYSTEMS

Shu-Ching Chen, Srinivas Sista, Mei-Ling Shyu, and R. L. Kashyap*

School of Electrical and Computer Engineering

Purdue University, West Lafayette, IN 47907-1285, U.S.A.

{shuching, sista, shyu, kashyap}@ecn.purdue.edu

ABSTRACT

In this paper, we present a database searching structure that incorporates image processing techniques to model multimedia data. A *Simultaneous Partition and Class Parameter Estimation (SPCPE)* algorithm that considers the problem of video frame segmentation as a joint estimation of the partition and class parameter variables has been developed and implemented to identify objects and their corresponding spatial relations. Based on the obtained object information, a *web spatial model (WSM)* is constructed. A *WSM* is a multimedia database searching structure to model the temporal and spatial relations of semantic objects so that multimedia database queries related to the objects' temporal and spatial relations on the images or video frames can be answered efficiently.

Key Words: multimedia database systems, indexing, database searching, video segmentation

1. INTRODUCTION

As more information sources become available in multimedia systems, the knowledge embedded in images or videos, especially spatial knowledge, should be captured by the data structure as much as possible. For this purpose, an unsupervised video segmentation method, the *Simultaneous Partition and Class Parameter Estimation (SPCPE)* algorithm, and a multimedia database searching structure called *Web spatial model (WSM)* are incorporated together. The objective of the *SPCPE* algorithm is to obtain objects in each video frame and their corresponding spatial relations [3, 4]; while the objective of the *WSM* is to model the spatial relations among objects, each covered by a bounding

* To whom correspondence should be directed (Tel: 765-494-0723).

This work has been partially supported by National Science Foundation under contract IRI 9619812.

box. The basic twenty-seven spatial relations introduced in [1, 2] are used in the *WSM* to model the objects' spatial relations. Based on the object information provided by the video segmentation method, the *WSM* can structure the temporal and spatial relations of semantic objects so that the multimedia database queries that involve objects' temporal and spatial relations on the images or video frames can be answered efficiently.

2. VIDEO FRAME SEGMENTATION

The problem of video frame segmentation is posed as a joint estimation of the partition and class parameter variables. The *SPCPE* algorithm starts with an arbitrary partition and computes the corresponding class parameters. From these class parameters and the data, a new partition is estimated. Both the partition and the class parameters are iteratively refined until there is no further change in them.

Suppose we have two classes. Let the partition variable be $\boldsymbol{c} = \{\boldsymbol{c}_1, \boldsymbol{c}_2\}$. Suppose we use a family of 2D polynomial functions (Equation 1) to describe the class k, where the error has a Gaussian distribution with mean zero and variance ρ_k for class k. Suppose we collect all the pixel values y_{ij} belonging to class k into a vector $\boldsymbol{y}_k$ and convert Equation 1 into a vector-matrix form as shown in Equation 2. Each row of the matrix $\boldsymbol{\Phi}$ is given by $(1,\ i,\ j,\ ij)$ and a_k is the vector of parameters $(a_{k0}, \cdots, a_{k3})^T$. The parameter estimates of class k, $\hat{a}_k$, can be computed directly using Equation 3.

$$y_{ij} = a_{k0} + a_{k1} i + a_{k2} j + a_{k3} ij, \quad \forall (i,j)\ y_{ij} \in \boldsymbol{c}_k, \quad k = 1, 2 \tag{1}$$

$$\boldsymbol{y}_k = \boldsymbol{\Phi} a_k, \quad k = 1, 2. \tag{2}$$

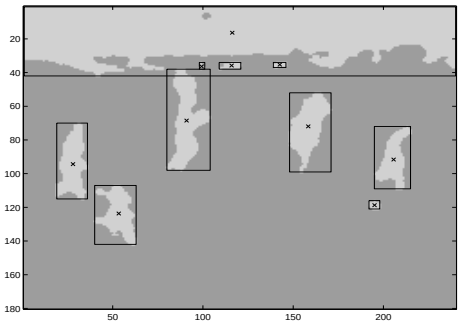
$$\hat{a}_k = (\boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^T \boldsymbol{y}_k, \quad k = 1, 2. \tag{3}$$

Let the parameters to be estimated be denoted by $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ where the parameters of class k are $\boldsymbol{\theta}_k = (a_{k0}, \cdots, a_{k3})^T$. The joint estimation can be simplified to Equation 4.

$$\begin{aligned} (\hat{\boldsymbol{c}}, \hat{\boldsymbol{\theta}}) &= \operatorname*{Arg\,min}_{(\boldsymbol{c}, \boldsymbol{\theta})} J(\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ J(\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= \sum_{y_{ij} \in \boldsymbol{c}_1} -\ln p_1(y_{ij}; \boldsymbol{\theta}_1) + \sum_{y_{ij} \in \boldsymbol{c}_2} -\ln p_2(y_{ij}; \boldsymbol{\theta}_2). \end{aligned} \tag{4}$$

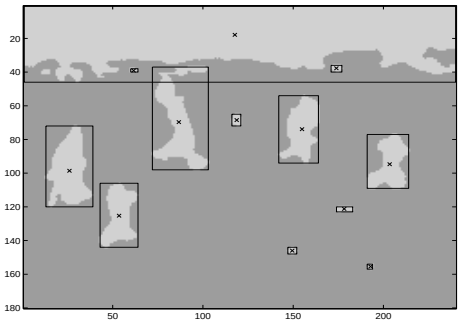
The video segmentation method is applied to an example soccer video. From the results on frames 1 through 60, a few frames – 1, 6 and 12 – are shown in Figure 1 along with the original frames adjacent to them. The centroid of each segment is marked with an 'x' and the segment is shown with a bounding box around it.

(a) Frame 1



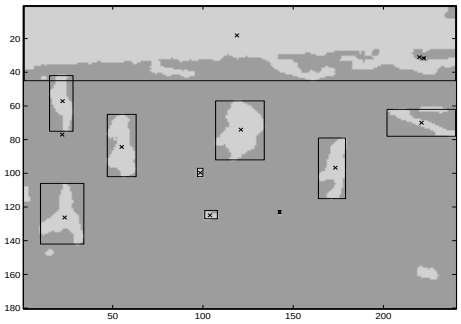
(b) Partition of Frame 1

(c) Frame 6



(d) Partition of Frame 6

(e) Frame 12



(f) Partition of Frame 12

Figure 1: Figures (a),(c),(e) are the original Frames 1,6,12 (on the left) and (b),(d),(f) show their corresponding segments (on the right). The centroid of each segment is marked with an 'x' and the segment is shown with a bounding box around it.

3. WEB SPATIAL MODEL (WSM)

There are three types of nodes – a *spatial node*, *intermediate node*, and *semantic object node* – which are connected by the *connection link* and the *ordered link*. Each type of node forms a layer in a WSM (as shown in Figure 2). The link is used to represent the connections and the relations between the nodes, and the object information structured in a WSM is provided by the video segmentation method introduced in the previous section. The types of nodes and links are defined as follows.

- *spatial node*: The twenty-seven spatial relations are represented by each spatial node. These nodes are the root nodes in the web spatial structure. They have no incoming link and can have more than one outgoing link to their children nodes. For example, the root node with number 1 or 10, which represents the centroid of the semantic object, is in the same region as that of the target semantic object or on the left of the semantic object, respectively.

- *connection link*: The *connection link* connects a *spatial* node and an *intermediate* node.

- *intermediate node*: These nodes are used to connect the *spatial* node and the semantic object node. Each *intermediate* node has only one incoming link from the root node and has two outgoing links to connect the *semantic object* nodes.

- *ordered link*: The *ordered link* connects an *intermediate* node and a *semantic object* node. The links are numbered by **1** and **2**. The links with number **1** and number **2** point to the *semantic object* and the *target semantic object*, respectively.

- *semantic object node*: These nodes represent the semantic objects. They are the leaf nodes of the web spatial structure.

The information stored in each node is defined in the following definition:

Definition 1: Let O be a set of n semantic objects such that O = $(o_1, o_2, \ldots, o_n)$. Each intermediate node is associated with a pair that consists of two semantic object nodes o_i, o_j $\forall i, j (1 \leq i \leq n, 1 \leq j \leq n, i \neq j)$. The spatial relation S to this pair is o_i S o_j. R = {(m_1,(sf_1,ef_1)), (m_1,(sf_1,ef_1)), ...} is a set of pairs for each intermediate node. Associated with each (m_k,(sf_k,ef_k)), $\forall k, (1 \leq k \leq n)$, is a single image frame for an image media stream m_k or a range of video frames for video stream m_k that goes from frame number sf_k to ef_k. For the image media stream, $sf_k = ef_k$.

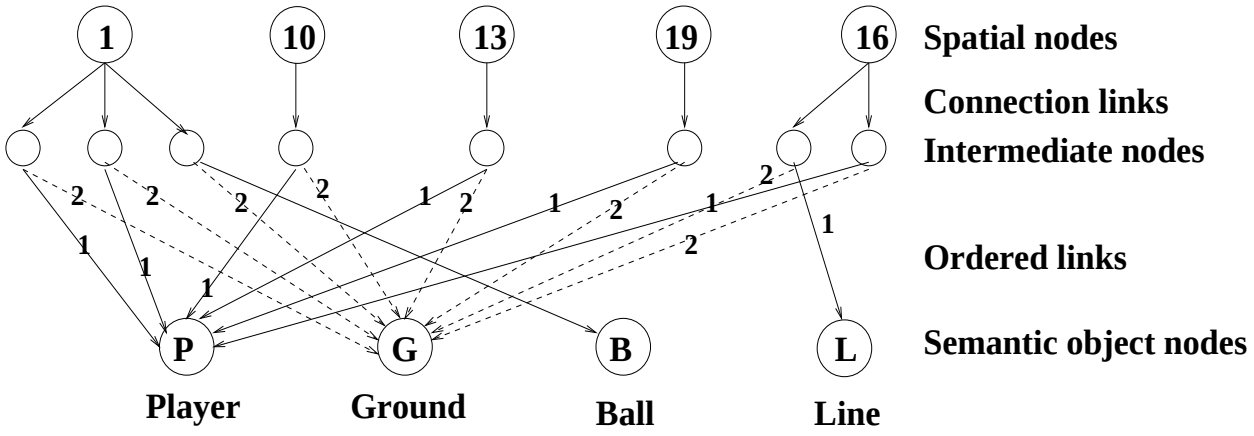


Figure 2: Web spatial relation model for semantic objects. Ordered links with number 1 (arrows) and number 2 (dashed arrows) point to the semantic objects and target semantic object node **(Ground)**.

4. MULTIMEDIA SPATIAL DATABASE QUERIES USING WSM

WSM can help to answer spatial multimedia database queries. Figure 2 is a WSM to model spatial relations in Figure 1. For simplicity, Figure 2 shows only the case when the **Ground** is selected as the target semantic object. Also, the segment for the sign boards is not included. The cases when another semantic object is chosen as the target semantic object are not shown here. Following is an example showing how to use WSM to help spatial database queries.

- **Query:** *Find the video clip beginning with a player on the left of the soccer field (ground) followed by the ball appearing in the center of the ground, and then the ball disappearing and the goal line appearing on the right of the ground.*

In this query, first, we want to find a player on the left of the ground; root node with number 10 (represented *left*) of WSM is identified. The only intermediate node is checked. This intermediate node has ordered links pointing to the *Player* and *Ground* semantic object nodes. The order links pointing to *Player* and *Ground* have order number 1 and 2, respectively. This tells us that the *Player* is at the left of the target semantic object *Ground*. The corresponding frame numbers stored in this intermediate node can help us to find the query video clip that matches the first query criterion. The same mechanism is applied for the second and the third query criteria.

5. CONCLUSIONS

In this paper, a database searching structure called *WSM* that incorporates an image processing technique (*SPCPE*) is proposed to efficiently answer the multimedia database queries related to the temporal and spatial relations of the objects on the images or video frames. WSM has non-unique parents and allows parallel searches and concurrent browsing paths. If the spatial relations are structured in WSM, then the burden of on-line processing of the raw image or video data for the database queries involving the spatial relations is reduced.

6. REFERENCES

[1] Shu-Ching Chen and R. L. Kashyap, "Augmented Transition Networks as Semantic Models for Multimedia Presentations, Multimedia Database Searching, and Multimedia Browsing," Technical Report TR-ECE 98-15, School of Electrical and Computer Engineering, Purdue University, December 1998.

[2] Shu-Ching Chen and R. L. Kashyap, "A Spatio-Temporal Semantic Model for Multimedia Presentations and Multimedia Database Systems," accepted for publication in *IEEE Transactions on Knowledge and Data Engineering.*

[3] R. L. Kashyap and S. Sista, "Unsupervised Classification and Choice of Classes: Bayesian Approach," Technical Report TR-ECE 98-12, School of Electrical and Computer Engineering, Purdue University, July 1998.

[4] S. Sista and R. L. Kashyap, "Unsupervised video segmentation and object tracking," to appear in IEEE Int'l Conf. on Image Processing, 1999.