Mining Video Associations for Efficient Database Management

Xingquan Zhu and Xindong Wu
Department of Computer Science, University of Vermont
Burlington, Vermont 05405, USA
{xqzhu, xwu}@cs.uvm.edu

Abstract

To support more efficient video database management, this paper explores the concept of video association mining, with which the association patterns are characterized by sequentially associated video shots and their cluster information. Given a continuous video sequence *V,* the video shot segmentation mechanism is first introduced to parse it into discrete shots. We then cluster shots into visually distinct groups and construct a shot cluster sequence by using the class label of each shot. An association mining scheme is designed to mine sequentially associated clusters from the sequence. Those detected associations will convey valuable knowledge for video database management. The experimental results demonstrate the effectiveness of our design.

1 Introduction

Since the 1990s, data mining has been a very active area for research and applications. Many successful techniques have been implemented through academic research and industrial applications [Agrawal and Srikant, 1994][Agrawal and Srikant, 1995][Wu, 1995] [Han and Kamber, 2000]. However, these approaches deal with various databases (like transaction datasets) in which the relationship between data items is explicitly given. Video and image databases are different from these databases. The most distinct feature of video and image databases is that the relationship between any two of their items cannot be explicitly (or precisely) figured out. This inherent complexity of the multimedia data has suggested that mining knowledge from multimedia materials is even harder than from general databases [Zhu *et al,* 2003][Thuraisingham, 2001][Zaiane, *et al.,* 1998]. Generally, there are two types of video mining techniques: (1) special pattern detection [Zhu *et al,* 2003], which detects some predefined special patterns; and (2) video clustering and classification [Oh *et al.,* 2002][Pan and Faloutsos, 2002], which clusters and classifies video units into different categories.

Different from these two types of video mining schemes, we address a new research area of video mining, video association mining, in this paper, where associations from video units are used to explore video knowledge. We will present a definition for video associations, and design a video association mining algorithm. As shown in Fig. 1, we first segment a video sequence into shots and cluster shots into groups. Then, we assemble the class information of each shot to form a shot cluster sequence. We mine sequential associations from the sequence to find clusters with strong correlations and take strongly correlated clusters as video associations.

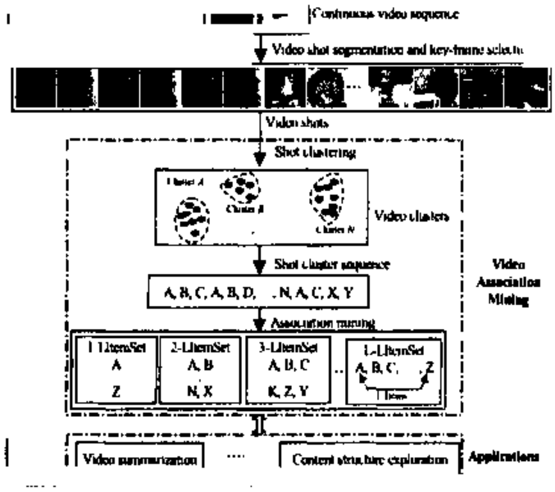
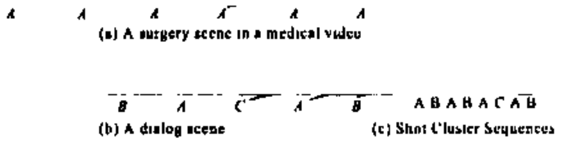


Fig. 1 Video association mining -- System architecture



(a) A surgery scene in a medical video

(b) A dialog scene

(c) Shot Cluster Sequences

Fig. 2 Some typical video scenes and video data transformation

2 Video Association Mining

Generally, most videos from our daily life are edited by editors, where various kinds of shots are packed as scenes to convey video scenarios, as shown in Fig. 2. There are two typical video scenes: (1) scenes that consist of visually similar shots, as demonstrated in Fig. 2(a); and (2) scenes that consist of visually distinct shots, as shown in Fig. 2 (b). In the first type of scenes, most video shots are visually similar. Take Fig. 2(a) as an example, if we denote each of the shots by "A", all shots form a sequence "AAAAAAA", and the self-coherence of "A" indicates an association of itself. We name this type of association as an i*ntra-association,* i.e., all items in the association are the same. In the second type of scenes, sequential associations exist too. In Fig.2(b), if we denote the actor by "A", the actress by "B" and the shot containing both of them by "C", all shots form a sequence "ABABACAB". The co-occurrence of "A" and "B" implies an association. We name this type of association as an *inter-association,* i.e., items in the association are different.

Based on the above observations, wc define a video association as a sequential pattern with $\{X_1..X_i..X_L;\ X_i' < X_j'$ for any $i < j\}$, where *X,* is a video item (see definition below), *L* denotes the length of the association. *X′* denotes the temporal order of *X„* and $X_i' < X_j'$ indicates that *X,* happens before X_j For an *inter-association,* $X_1 \cap .. \cap X_L = \emptyset$; and for an *intra-association,* $X_1 \cap .. \cap X_L = X_k,\ k = \forall[1,L]$. For simplicity, we use *{X}* as the abbreviation of an association.

Due to the fact that the temporal information in a video sequence plays an important role in conveying the video content, we integrate traditional association measures *{support* and *confidence)* and video temporal information to evaluate video associations. Some definitions are given as follows:

- An *item* is a basic unit, which denotes a shot cluster.
- Given a shot cluster sequence, the *temporal distance (TD)* between two items is the number of shots between them. E.g, given sequence "ABDEC", the temporal distance of "AB" is *TD(AB)=0,* and for "AC" is *TD(AC)* =3.
- An *L-ItemAssociation* is a video association that consists of *L* sequential items. E.g., "ABC" is a 3-ItemAssociation.
- The *L-ItemSet* is an aggregation of all *L-ItemAssociations,* with each of its members being an *L-ItemAssociation.*
- Given a *temporal distance threshold (TDT) TDT=T,* the *temporal support {TS)* of an association *{X1...X$_L$}* is defined as the number of times that this association appears sequentially in the sequence. In addition, each time this association appears, the temporal distance between any two neighboring items of the association should be no more than *T* shots. That is, given any *X,* and X_{i+1} $(i \geq 1, i+1 \leq L)$, $TD(X_i, X_{i+1}) \leq T$.
- Given *TDT=T,* the *confidence* of an association $\{X_1..X_i..X_L\}$ is defined as the ratio between the temporal support of *{X}* when *TDT=T* and the number of maximal possible occurrences of the association {X}. For an inter-association, the maximal possible occurrence of the association is determined by the number of occurrences of the item with the minimal support. Its confidence is defined by Eq.(l). For an intra-association, all items are the same, the maximal possible occurrence of the association is determined by the support of the item and the association length (L), as defined by Eq.(2), where *(x)* indicates the maximal integer which is not larger than *x.*

$$Conf\{X\}_{TDT=T} = TS\{X\}_{TDT=T} / Min\{TS(X_1),..,TS(X_L)\} \quad (1)$$

$$Conf\{X\}_{TDT=T} = TS\{X\}_{TDT=T} / \langle TS(X_i)/L \rangle, i = \forall[1,L] \quad (2)$$

- The *L-LItemSet* is an aggregation of all *L-ItemAssociations* that each of their temporal support is no less than a given threshold.

2.1 Association Mining Algorithm

Our video mining algorithm consists of the following phases:

1. Transform Phase. Given video *V,* this step transforms *V* from continuous frames into a sequence dataset *D.*

2. L-LItemSet Phase. In this phase, we mine both intra-associations and inter-associations from D. We first find the L-ItemSet, and then use L-ItemSet and a user-specified threshold to find L-LItcmSet. We will iteratively execute this phase until no more non-empty L-LItcmSet can be found.

3. Collection & Pruning Phase. This phase prunes and selects valuable associations from all *L-LitemSets.*

Since Phase 3 is quite obvious, we focus on Phases 1 and 2. Meanwhile, due to the fact that all items in intra-associations are the same, mining this type of associations is relatively easy, wc hereby introduce mechanisms on mining inter-associations only.

In the transform phase, we adopt some video processing techniques to segment a video sequence into shots, and execute shot clustering to explore relationships among shots. To detect video shots, we use our existing algorithm in [Zhu *et al.,* 2003]. For the sake of simplicity, we select the 10th frame of each shot as its key-frame. After the shot segmentation, we adopt a modified split-and-merge clustering algorithm [Horowitz and Pavlidis, 1974] to cluster video shots into groups where visually similar shots are first merged into groups and the groups with large visual variances are split into two clusters. After shot clustering, each shot will receive a class label, we sequentially aggregate the class information of each shot by its original temporal order to form a shot cluster sequence *D.* As shown in Fig. 2, each icon image denotes one shot and the letter below it indicates its class label, and the acquired shot cluster sequences are given in Fig. 2 (c). Table 2 gives an example of the video association mining, where the first column presents the video shot cluster sequence.

In the L-LItcmSet phase, we use the large ItemSet from the previous pass to generate the candidate ItemSet and then measure their temporal support by making a pass over the database A as shown in Fig. 3. At the end of the pass, the support of each candidate is used to determine the L-LItemSet. The candidate generation is similar to the method in [Agrawal and Srikant, 1995]. It takes the set of all *k-1 -ItemAssociations* in L_{k-1} and all their items as input, and works as shown in Fig. 4. Take the 3-LItemSet L_3 in the fourth column of Table 2 as an example. If L_3 is given as the input, we will get the ItemSet shown in the fifth column after the join. After pruning out sequences whose subsequences are not in L_3, the sequences shown in the sixth column arc left. E.g., {ABDC} is pruned out because its subsequence {BDC} is not in L_3.

```
(1)  Given user specified TDT and TS thresholds.
(2)  I_1={1-ItemSet}; L_1={1-LItemSet}; // Find {1-LItemSet} by
     using {1-ItemSet} and the user-specified thresholds.
(3)  For ( k = 2; L_{k-1} ≠ Ø; k ++ )
     Begin:
(4)    I_K=New candidates generated from L_{k-1} (see Fig. 5).
(5)      For each k-ItemAssociation in I_k, we count its temporal
         support by considering the user's specification with TDT.
(6)    L_k=Candidates in I_k with the minimum temporal support.
     End
```

Fig. 3 The L-LItemSet Phase of the mining algorithm

```
(1).  Join the items of associations in L_{k-1}
(2).  Insert the join results into I_k
      select {p.Item_1,.., p.Item_{k-1}, q.Item_{k-1}}
      from   {L_{k-1}.p, L_{k-1}.q, p ≠ q}
      where  {p.Item_1=q.Item_1 & ..& p.Item_{k-2}=q.Item_{k-2}}
(3).  Delete any member x ∈ I_k such that some
      {k-1-ItemAssociation} of x is not in L_{k-1}.
```

Fig. 4 Candidate generation

In Fig. 3 and Fig. 4, L_k denotes the *k-LItemSet* and I_k the *k-ItemSet.*

3 Experimental Results and Discussion

Traditional video database systems use video shots as the units to index and manage video data where the visual similarities among shots are used to construct the index structure. Unfortunately, a single shot which is separated from its context has less capability to convey semantics; Moreover, the index considering only visual similarities ignores the temporal information among shots. Consequently, the constructed cluster nodes may contain shots that have considerable variances both in semantics and visual content and hereby do not make much sense to human perception. Accordingly, a semantic video database management framework has been presented [Zhu *et al,* 2003] where video semantic units (events, scenes or other scenario information) are used to construct a database index. To facilitate this goal, one of the most important tasks is to detect the video semantic units. In this section, our experimental results will demonstrate that the proposed video association mining technique could be used to explore semantic units for the management of video database systems.

To evaluate the ability of video associations in addressing local event and scenario information and figure out the relationship between *TDT* and the mined associations, we set *TDT* with different values *T (T=l. 3, 5, 7, 9* and °°) and assess the associations. For each acquired association, we scan the datasct to check whether all items in the association belong to the same scene each time when the association appears. We define the *Scene Coverage (SC)* of an association as the ratio between the frequency of the association's items belonging to the same scene and the frequency of the association's appearance. The higher the *SC,* the better the association addresses the scene and event information. On the other side, with an adopted temporal support, the smaller the *TDT,* the less is the number of mined associations. This indicates that the *TDT* also acts as a factor for pruning associations. We hereby define the *Pruning Rate (PR)* as the ratio between the number of associations when *TDT* is *T* (7=1, *3, 5, 7, 9)* and 00. We have performed our experiments with 5 news videos and 20 medical videos (about 700 minutes), and the results arc given in Table 2.

Table 1 demonstrates that when the *TDT* increases, the mined associations become worse in addressing the local scenario and event information. One reason for this declination is that the clustering process may cluster semantically unrelated shots into one group, and consequently, when evaluating the *scene coverage,* the items of an association would come from different events. However, with relatively small *TDT* values, this type of errors can be reduced, because items with a small temporal distance would more likely belong to one semantic unit. On the other hand, Table 1 also indicates that the smaller the *TDT,* the less is the number of mined associations, but the better arc the mined associations in addressing event and scenario information. Depending on the user's objectives of association mining, a balance between the number of associations and the *SC* is necessary to select a reasonable value for *TDT.*

Table I. Video association mining results

TDT	1	3	5	7	9	∞
SC	0.893	0.868	0.755	0.623	0.553	0.342
PR	0.264	0.377	0.516	0.658	0.761	1.0

4 Conclusions

To facilitate video database management, we have explored a new research area of video data mining. A video association mining algorithm has been proposed. Given video *V,* we first transform it from sequential frames to a relational dataset by shot segmentation, clustering, and constructing a shot cluster sequence. The video mining scheme mines sequentially associated video items from this sequence. In addition to using the traditional association measures, we have integrated temporal features among video shots into the video association evaluation. The experimental results have demonstrated the ability of our mined associations in addressing semantic information for video database management.

Acknowledgments

This research has been supported by the U.S. Army Research Laboratory and the U.S. Army Research Office under grant number DAAD19-02-1-0178.

References

[Agrawal and Srikant, 1994] R. Agrawal and R. Srikant, Fast algorithm for mining association rules. *Proc. of VLDB, 1994.*

[Agrawal and Srikant, 1995] R. Agrawal and R. Srikant, Mining sequential patterns. *Proc. of 11th ICDE Conference, pp.3-14, 1995.*

[Han and Kamber, 2000] J. Han and M. Kambcr, Data Mining: Concepts and Techniques, *Morgan Kaufmann, 2000.*

[Horowitz and Pavlidis, 1974] S. Horowitz, T. Pavlidis, Picture segmentation by a directed split-and-mcrgc procedure. *Proc. of Int. Joint Conf. on Pattern Recognition, pp. 424—433, 1974.*

[Oh and Bandi, 2002] J. Oh and B. Bandi, Multimedia data mining framework for raw video sequence. *Proc. ofMDM/KDD, 2002.*

[Pan and Faloutsos, 2002] J. Pan and C. Faloutsos, GeoPlot: Spatial data mining on video libraries. *Proc. of CI KM, 2002.*

[Thuraisingham, 2001] B. Thuraisingham, Managing and mining multimedia database. *CRC Press, 2001.*

[Wu, 1995] Xindong Wu, Knowledge acquisition from databases. *Ablex Publishing Corp., USA, 1995.*

[Zaiane *et al.,* 1998] O. Zaiane, J. Han, Z. Li, S. Chee and J. Chiang, MultimediaMiner: a system prototype for multimedia data mining. *Proc. ofACMSIGMOD, 1998.*

[Zhu *et al.,* 2003] X. Zhu, W. Aref, J. Fan, A. Catlin, and A. Elmagarmid, Medical video mining for efficient database indexing, management and access. *Proc. of ICDE, Mar., 2003.*

Table 2. An example of video association mining, where $\{x\}^C$ indicates an association, X denotes the items of the association, *S* and *C* indicate the temporal support and confidence respectively. The sequential order in the first column is from left to right, top to bottom.

Shot cluster sequence of *V*	2-LItemSet		3-LItemSet (*TDT*=∞)	Candidate 4-ItemSet (after join)	4-ItemSet (after pruning)	4-LItemSet (*TDT*=∞)
	(*TDT*=1)	(*TDT*=∞)				
A B C B C D C A C B C A B D A B C D B E	$\{AB\}_4^{1.0}$, $\{AC\}_3^{0.8}$, $\{AD\}_1^{0.5}$, $\{BA\}_2^{0.5}$, $\{BB\}_1^{0.5}$, $\{BC\}_4^{0.7}$, $\{BD\}_1^{1.0}$, $\{CB\}_4^{0.7}$, $\{CC\}_2^{0.7}$, $\{CD\}_2^{0.7}$, $\{DB\}_2^{0.7}$	$\{AB\}_4^{1.0}$, $\{AC\}_3^{0.8}$, $\{AD\}_3^{1.0}$, $\{BA\}_3^{0.8}$, $\{BB\}_3^{1.0}$, $\{BC\}_4^{0.7}$, $\{BD\}_3^{1.0}$, $\{CB\}_4^{0.7}$, $\{CC\}_1^{1.0}$, $\{CD\}_3^{1.0}$, $\{DB\}_1^{1.0}$	$\{ABC\}_3^{0.8}$, $\{ABD\}_3^{1.0}$ $\{ACB\}_3^{0.8}$, $\{ACD\}_3^{1.0}$ $\{BCD\}_3^{1.0}$, $\{CDB\}_3^{1.0}$	$\{ABCD\}$ $\{ABDC\}$ $\{ACBD\}$ $\{ACDB\}$	$\{ABCD\}$ $\{ACDB\}$	$\{ABCD\}_3^{1.0}$