V

Video Data Mining

JungHwan Oh
University of Texas at Arlington, USA

JeongKyu Lee
University of Texas at Arlington, USA

Sae Hwang
University of Texas at Arlington, USA

INTRODUCTION

Data mining, which is defined as the process of extracting previously unknown knowledge and detecting interesting patterns from a massive set of data, has been an active research area. As a result, several commercial products and research prototypes are available nowadays. However, most of these studies have focused on corporate data — typically in an alpha-numeric database, and relatively less work has been pursued for the mining of multimedia data (Zaïane, Han, & Zhu, 2000). Digital multimedia differs from previous forms of combined media in that the bits representing texts, images, audios, and videos can be treated as data by computer programs (Simoff, Djeraba, & Zaïane, 2002). One facet of these diverse data in terms of underlying models and formats is that they are synchronized and integrated hence, can be treated as integrated data records. The collection of such integral data records constitutes a multimedia data set. The challenge of extracting meaningful patterns from such data sets has lead to research and development in the area of multimedia data mining. This is a challenging field due to the non-structured nature of multimedia data. Such ubiquitous data is required in many applications such as financial, medical, advertising and Command, Control, Communications and Intelligence (C3I) (Thuraisingham, Clifton, Maurer, & Ceruti, 2001). Multimedia databases are widespread and multimedia data sets are extremely large. There are tools for managing and searching within such collections, but the need for tools to extract hidden and useful knowledge embedded within multimedia data is becoming critical for many decision-making applications.

BACKGROUND

Multimedia data mining has been performed for different types of multimedia data: image, audio and video. Let us first consider image processing before discussing image and video data mining since they are related. Image processing has been around for some time. Images include maps, geological structures, biological structures, and many other entities. We have image processing applications in various domains including medical imaging for cancer detection, and processing satellite images for space and intelligence applications. Image processing has dealt with areas such as detecting abnormal patterns that deviate from the norm, and retrieving images by content (Thuraisingham, Clifton, Maurer, & Ceruti, 2001). The questions here are: *what* is image data mining and *how* does it differ from image processing? We can say that while image processing is focused on manipulating and analyzing images, image data mining is about finding useful patterns. Therefore, image data mining deals with making associations between different images from large image databases. One area of researches for image data mining is to detect unusual features. Its approach is to develop templates that generate several rules about the images, and apply the data mining tools to see if unusual patterns can be obtained. Note that detecting unusual patterns is not the only outcome of image mining; that is just the beginning. Since image data mining is an immature technology, researchers are continuing to develop techniques to classify, cluster, and associate images (Goh, Chang, & Cheng, 2001; Li, Li, Zhu, & Ogihara, 2002; Hsu, Dai, & Lee, 2003; Yanai, 2003; Müller & Pun, 2004). Image data mining is an area with applications in numerous domains including space, medicine, intelligence, and geoscience.

Mining video data is even more complicated than mining still image data. One can regard a video as a collection of related still images, but a video is a lot more than just an image collection. Video data management has been the subject of many studies. The important areas include developing query and retrieval techniques for video databases (Aref, Hammad, Catlin, Ilyas, Ghanem, Elmagarmid, & Marzouk, 2003). The question we ask yet again is what is the difference between video information retrieval and video mining? There is no clear-cut answer for this question yet. To be consistent with our terminology, we can say that finding correlations and patterns previously unknown from large video databases is video data mining.

MAIN THRUST

Even though we define video data mining as finding correlations and patterns previously unknown, the current status of video data mining remains mainly at the pre-processing stage, in which the preliminary issues such as video clustering, and video classification are being examined and studied for the actual mining. Only a very limited number of papers about finding any patterns from videos can be found. We discuss video clustering, video classification and pattern finding as follows.

Video Clustering

Clustering is a useful technique for the discovery of some knowledge from a dataset. It maps a data item into one of several clusters, where clusters are natural groupings for data items based on similarity metrics or probability density models (Mitra & Acharya, 2003). Clustering pertains to unsupervised learning, when data with class labels are not available. Clustering consists of partitioning data into homogeneous granules or groups, based on some objective function that maximizes the inter-cluster distances, while simultaneously minimizing the intra-cluster distances. Video clustering has some differences with conventional clustering algorithms. As mentioned earlier, due to the unstructured nature of video data, preprocessing of video data by using image processing or computer vision techniques is required to get structured format features. Another difference in video clustering is that the time factor should be considered while the video data is processed. Since video is a synchronized data of audio and visual data in terms of time, it is very important to consider the time factor. Traditional clustering algorithms can be categorized into two main types: partitional and hierarchical clustering (2003). Partitional clustering algorithms (i.e., *K-means* and *EM*) divide the patterns into a set of spherical clusters, while minimizing the objective function. Here the number of clusters is predefined. Hierarchical algorithms, on the other hand, can again be grouped as agglomerative and divisive. Here no assumption is made about the shape or number of clusters, and validity index is used to determine termination.

Two of the most popular partitional clustering algorithms are *K-means* and *Expectation Miximization (EM)*. In K-means, the initial centroids are selected, and each data item is classified to a cluster with the smallest distance. Based on the previous results, the cluster centroids are updated, and all corresponding data items are re-clustered until there is no centroid change. It is easily implemented, and provides a firm foundation of variances through the clusters. We can find the papers using the K-means algorithm for video clustering in the literature (Ngo, Pong, & Zhang, 2001). EM is a popular iterative refinement algorithm that belongs to the model-based clustering. It differs from the conventional K-means clustering algorithm in that each data point belongs to a cluster according to some weight or probability of membership. In other words, there are no strict boundaries between clusters. New means are computed based on weighted measures. It provides a statistical model for the data and is capable of handling the associated uncertainties. We can find the papers using the EM algorithm for video clustering in the literature (Lu, & Tan, 2002; Frey, & Jojic, 2003).

Hierarchical clustering methods create hierarchical nested partitions of the dataset, using a tree-structured dendogram and some termination criterion. Every cluster node contains child clusters; sibling clusters partition the points covered by their common parent. Such an approach allows exploring data on different levels of granularity. Hierarchical clustering methods are categorized into agglomerative (bottom-up) and divisive (top-down). An agglomerative clustering starts with one-point (singleton) clusters and recursively merges two or more of the most appropriate clusters. Divisive clustering starts with one cluster of all data points and recursively splits the most appropriate cluster. The process continues until a stopping criterion is achieved. The advantages of hierarchical clustering include: embedded flexibility regarding the level of granularity, ease of handling of any forms of similarity or distance, and applicability to any attribute types. The disadvantages of hierarchical clustering are vagueness of termination criteria, and the fact that most hierarchical algorithms do not revisit constructed (intermediate) clusters for the purpose of their improvement. Hierarchical clustering is used in video clustering because it is easy to handle the similarity of extracted features from video, and it can represent the depth and granularity by the level of tree (Okamoto, Yasugi, Babaguchi, & Kitahashi, 2002).

Video Classification

While clustering is an unsupervised learning method, classification is a way to categorize or assign class labels to a pattern set under the supervision. Decision boundaries are generated to discriminate between patterns belonging to different classes. The data set is initially partitioned into training and test sets, and the classifier is trained on the former. A framework to enable semantic video classification and indexing in a specific video domain (medical video) was proposed (Fan, Luo, & Lin, 2003). VideoGraph, a tool for finding similar scenes in a video, was studied (Pan & Faloutsos, 2001). A method for classification of different kinds of videos that uses the output of a concise video summarization technique that forms a list of key frames was presented (Lu, Drew, & Au, 2001).

Pattern Finding

A general framework for video data mining was proposed to address the issue of how to extract previously unknown knowledge and detect interesting patterns (Oh, Lee, Kote, & Bandi, 2003). In the work, they develop how to segment the incoming raw video stream into meaningful pieces, and how to extract and represent some feature (i.e., motion) for characterizing the segmented pieces. Then, the motion in a video sequence is expressed as an accumulation of quantized pixel differences among all frames in the video segment. As a result, the accumulated motions of the segment are represented as a two dimensional matrix. Further, a method to capture the location of motions occurring in a segment is developed using the same matrix. How to cluster those segmented pieces using the features (the amount and the location of motion) extracted by the matrix above is studied. Also, an algorithm is investigated to determine whether a segment has normal or abnormal events by clustering and modeling normal events, which occur most frequently. In addition to deciding normal or abnormal, the algorithm computes a Degree of Abnormality of a segment, which represents the distance of a segment from the existing segments in relation to normal events.

A fast dominant motion extraction scheme called Integral Template Match (ITM), and a set of qualitative and quantitative description schemes were proposed (Lan, Ma, & Zhang, 2003). A video database management framework and its strategies for video content structure and events mining were introduced (Zhu, Aref, Fan, Catlin, & Elmagarmid, 2003). The methods of extracting editing rules from video stream were proposed by introducing a new data mining technique (Matsuo, Amano, & Uehara, 2002).

FUTURE TRENDS

As mentioned above, there have been a number of attempts to apply clustering methods to video data. However, these classical clustering techniques only create clusters but do not explain why a cluster has been established (Perner, 2002). The conceptual clustering method builds clusters and explains why a set of objects confirms a cluster. Thus, conceptual clustering is a type of learning by observations, and it is a way of summarizing data in an understandable manner. In contrast to hierarchical clustering methods, conceptual clustering methods build the classification hierarchy based on merging two groups. The algorithm properties are flexible in order to dynamically fit the hierarchy to the data. This allows incremental incorporation of new instances into the existing hierarchy and updating this hierarchy according to the new instance. A concept hierarchy is a directed graph in which the root node represents the set of all input instances and the terminal nodes represent individual instances. Internal nodes stand for the sets of instances attached to them and represent a super-concept. The super-concept can be represented by a generalized representation of this set of instances such as the prototype, the method or a user-selected instance. We can find a work applying this conceptual mining to image domain (Perner, 1998). However, we cannot find any papers related to conceptual clustering for video data. Since it is important to understand what a created cluster means semantically, we need to study how to apply conceptual clustering to video data.

In fact, one of the most important techniques for data mining is association-rule mining since it is most efficient way to find unknown patterns and knowledge. Therefore, we need to investigate how to apply association-rule mining to video data. For association-rule mining, we need a set of transactions (*D*), a set of the literals (or items, *I*), and an itemset (*X*) (Zhang & Zhang, 2002). After video is segmented into the basic units such as shots, scenes, and events, each segmented unit can be modeled as a transaction, and the features from a unit can be considered as the items contained in the transaction. In this way, video association mining can be transformed into problems of association mining in traditional transactional databases. A work using some associations among video shots to create a video summary was proposed (Zhu & Wu, 2003). But they did not come up with the concepts of transaction and itemset. We need to further investigate the optimal unit for the concept of transaction, and the possible items in a transaction of video to characterize it.

Also, we need to study whether video can be considered as time-series data. It looks positive since the behavior of a time-series data item is very similar to that of video data. A time-series data item has a value at any given time, and the value is changing over time. Similarly a feature of video has a value at any given time, and the value is changing over time. If video can be considered as time-series data, we can get the advantages of the techniques already developed for time-series data mining. When the similarity between data items is computed, the ordinary distance metrics, such as Euclidean distance, may not be suitable, because of its high dimensionality and time factor. In order to address this problem, alternative ways are used to get a more accurate measure of similarity; for example, Dynamic Time Warping and Longest Common Subsequences.

Although most of data mining techniques are in batch processing, including video data mining as well as conventional data mining, some applications need to be processed in real time or near real time. For example, the anomaly detection system in a surveillance video using data mining should be processed in real time. Therefore, we also need to examine online video data mining.

CONCLUSION

Data mining describes a class of applications that look for hidden knowledge or patterns in large amounts of data. Most of data mining research has been dedicated to alpha-numeric databases, and relatively less work has been done for the multimedia data mining. The current status and the challenges of video data mining which is a very premature field of multimedia data mining, are discussed in this paper. The issues discussed should be dealt with in order to obtain valuable information from vast amounts of video data.

REFERENCES

Aref, W., Hammad, M., Catlin, A.C., Ilyas, I., Ghanem, T., Elmagarmid, A., & Marzouk, M. (2003). Video query processing in the VDBMS testbed for video database research. *Proceeding of 1st ACM International Workshop on Multimedia Database*, 25-32.

Fan, J., Luo, H., & Lin, X. (2003). Semantic video classification by integrating flexible mixture model with adaptive EM algorithm. *Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*, 9-16.

Frey, B.J., & Jojic, N. (2003). Transformation-invariant clustering using the EM algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(1), 1-17.

Goh, K.S., Chang, E., & Cheng, K.T. (2001). SVM binary classifier ensembles for image classification. *Proceedings of the 10th International Conference on Information and Knowledge Management*, 395-402.

Hsu, W., Dai, J., & Lee, M.L. (2003). Mining viewpoint patterns in image databases. *Proceeding of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 553-558.

Lan, D.J., Ma, Y.F., & Zhang, H.J. (2003). A novel motion-based representation for video mining. *Proceedings of 2003 IEEE International Conference on Multimedia and Expo, (ICME '03), Vol.3*, 469-472.

Li, T., Li, Q., Zhu, S., & Ogihara, M. (2002). A survey on wavelet applications in data mining. *ACM SIGKDD Explorations Newsletter*, *4*(2), 49-68.

Lu, C., Drew, M.S., & Au, J. (2001). Classification of summarized videos using hidden Markov models on compressed chromaticity signatures. *Proceeding of 9th ACM International Conference on Multimedia*, 479-482.

Lu, H., & Tan, Y.P. (2002). On model-based clustering of video scenes using scenelets. *Proceedings of 2002 IEEE International Conference on Multimedia and Expo*, *Vol.1*, 301-304.

Matsuo, Y., Amano, M., & Uehara, K. (2002). Mining video editing rules in video streams. *Proceeding of the 10th ACM International Conference on Multimedia*, 255-258.

Mitra, S., & Acharya, T. (2003). *Data mining: Multimedia, soft computing, and bioinformatics*. John Wiley & Sons, Inc.

Müller, H., & Pun, T. (2004). Learning from user behavior in image retrieval: Application of market basket analysis. *International Journal of Computer Vision*, *56*(1/2), 65-77.

Ngo, C.W., Pong, T.C., & Zhang, H.J. (2001). On clustering and retrieval of video shots. *Proceeding of 9th ACM International Conference on Multimedia*, 51-60.

Oh, J., Lee, J., Kote, S., & Bandi, B. (2003). Multimedia data mining framework for raw video sequences. *Mining Multimedia and Complex Data, Lecture Notes in Artificial Intelligence, Vol. 2797* (pp. 18-35). Springer Verlag.

Okamoto, H., Yasugi, Y., Babaguchi, N., & Kitahashi, T. (2002). Video clustering using spatio-temporal image with fixed length. *Proceedings of 2002 IEEE International Conference on Multimedia and Expo*, *Vol. 1*, 53-56.

Pan, J.U., & Faloutsos, C. (2001). VideoGraph: A new tool for video mining and classification. *Proceedings of the 1st ACM/IEEE-CS Joint Conference in Digital Libraries*, 116-117.

Perner, P. (2002). *Data mining on multimedia data*. Springer.

Perner, P. (1998). Using CBR learning for the low-level and high-level unit of a image interpretation system. In S. Singh (Ed.), *International Conference on Advances Pattern Recognition ICAPR98* (pp. 45-54). London: Springer Verlag.

Simoff, S.J., Djeraba, C., & Zaïane, O.R. (2002). MDM/KDD2002: Multimedia data mining between promises and problems. *ACM SIGKDD Explorations Newsletter, 4*(2), 118-121.

Thuraisingham, B., Clifton, C., Maurer, J., & Ceruti, M.G. (2001). Real-Time data mining of multimedia objects. *Proceedings of 4th IEEE International Symposium on Object-Oriented Real-Time distributed Computing, ISORC-2001*, 360-365.

Yanai, K. (2003). Managing images: Generic image classification using visual knowledge on the Web. *Proceeding of the 11th ACM International Conference on Multimedia*, 167-176.

Zaïane, O.R., Han, J., & Zhu, H. (2000). Mining recurrent Items in multimedia with progressive resolution refinement. *Proceedings of 16th International Conference on Data Engineering*, 461-470.

Zhang, C., & Zhang, S. (2002). *Association rule mining*. Springer.

Zhu, X., Aref, W.G., Fan, J., Catlin, A.C., & Elmagarmid, A.K. (2003). Medical video mining for efficient database indexing, management and access. *Proceedings of the 19th International Conference on Data Engineering (ICDE '03)*, 569-580.

Zhu, X., & Wu, X. (2003). Sequential association mining for video summarization. *Proceedings of 2003 IEEE International Conference on Multimedia and Expo, (ICME '03), Vol. 3*, 333-336.

KEY TERMS

Classification: A method of categorizing or assigning class labels to a pattern set under the supervision.

Clustering: A process of mapping a data item into one of several clusters, where clusters are natural groupings for data items based on similarity metrics or probability density models.

Concept Hierarchy: A directed graph in which the root node represents the set of all input instances and the terminal nodes represent individual instances.

Conceptual Clustering: A type of learning by observations and a way of summarizing data in an understandable manner.

Degree of Abnormality: A probability that represents to what extent a segment is distant to the existing segments in relation with normal events.

Dendogram: It is a 'tree-like' diagram that summaries the process of clustering. Similar cases are joined by links whose position in the diagram is determined by the level of similarity between the cases.

Digital Multimedia: The bits that represent texts, images, audios, and videos, and are treated as data by computer programs.

Image Data Mining: A process of finding unusual patterns, and making associations between different images from large image databases. One could mine for associations between images, cluster images, classify images, as well as detect unusual patterns.

Image Processing: A research area detecting abnormal patterns that deviate from the norm, and retrieving images by content.

Video Data Mining: A process of finding correlations and patterns previously unknown from large video databases.