SURVEY PAPER

LITERATURE SURVEY ON MULTIMEDIA DATA RETRIEVAL TECHNIQUES USING DATA MINING

By

D. SARAVANAN

Associate Professor, IFHE University, IBS Hyderabad, Telangana, India.

ABSTRACT

Data mining is a process of extracting facts from a given huge set of data. Of the available huge data set, multimedia is one which contains diverse data such as audio, video, image, text and motion, and such video data play a vital role in the field of video data mining. For extracting information from this huge content, we need special techniques. Because of numerous devices like cell phones, tablets, and other electronic devices available today, we can upload images or video data very easily. Today information comes in the form of electronic information instead of text information. Most of the information like news, entertainment, books, healthcare and weather forecasts are in the electronic form. Among this information the acquisition and storage of video data is an easy task, but retrieval of information from video data is challenging. This paper brings some of these issues and challenges involved in image extraction using data mining techniques.

Keywords: Data Mining, Image Extraction, Knowledge Extraction, Feature Extraction, Knowledge Discovery, Data Sets, Image Retrieval.

INTRODUCTION

Multimedia is any combination of text, art, sound, animation, and video delivered to you by computer or other electronic or digitally manipulated means [16]. Through the rapid growth of multimedia technology, multimedia content can be created, shared and distributed easily. The amount of available digital resources is continuously increasing, promoted by a growing interest of users and by the development of new technology for the ubiquitous enjoyment of digital resources. Multimedia information has become increasingly prevalent and it constitutes a significant component of the multimedia contents on the Internet. Information retrieval is based on multimedia components [10][11]. Since multimedia information can be represented in various forms, formats, and dimensions, searching such information is far more challenging than text-based search. For this huge content [2][8][9], it is necessary to remove the duplicated information, in order to reduce the search time. For this video data today, indexing and clustering are performed for efficient searching and retrieval [3][14]. While some basic forms of multimedia retrieval are available on the Internet, these tend to be inflexible and have significant limitations too. Today, most of the multimedia retrieval processes are done with the help of content based search [7] and numerous research articles have proposed these concepts.

Need of Literature Survey

Storing the multimedia data is an easy task, but retrieval is quite a challenging task. To overcome this difficulty today, most of the research works are carried in the field of video data mining. This paper brings various video data mining techniques.

1. Literature Survey

1.1 Title: Mining Spatiotemporal Video Patterns towards Robust Action Retrieval

Authors: Liujuan Cao, Rongrong Ji, Yue Gao, WeiLiu, and Qi Tian (2012)

1.1.1 Problem:

Coming to the video sharing websites like, YouTube,

MySpace, and Yahoo Video, nowadays, there is an increasing amount of user-contributed videos on the Web. To manage the ever growing videos, content-based accessing, browsing, search and analysis techniques emerge. More specifically, they have focused on the retrieval of actor-independent actions, i.e. the motion patterns were only taken care of rather than the visual appearances of the actor, and the near-duplicate action matching. Yet, it also differs from the traditional action recognition scenario such that, there is no predefined action category to ensure the scalability.

1.1.2 Finding

A spatiotemporal co-location video pattern mining is proposed in this approach with an application to robust action retrieval in YouTube videos. First, it introduces an attention shift scheme to detect and partition the focused human actions from YouTube videos, which is based upon the visual salience modeling together with both the face and body detectors. From the segmented spatiotemporal human action regions, the authors have extracted a 3D-SIFT detector. Then, they have quantized all the detected interest points from the reference YouTube videos into a vocabulary, based on which each individual interest point are assigned with a word identity. An APrior based frequent itemset mining scheme is then deployed over the spatiotemporal co-located words to discover the co-located video patterns. Finally, both the visual words and patterns are fused together and a boosting-based feature selection to output the final action descriptors has been leverage, which incorporates the ranking distortion of the conjunctive queries into the boosting objective.

1.1.3 Solution

In this paper, the authors have proposed a robust and discriminating action search paradigm, specialized for searching in the user contributed YouTube videos that typically has uncontrolled qualities. Their contributions are threefold: First, they have proposed an attention shift model for saliency-driven human action segmentation and partition. The second contribution is a spatiotemporal co-location video pattern mining paradigm, aiming for discovering more eigen word combinations to capture the motion patterns based on a Distance based Co-location pattern Mining. Finally, they have proposed a novel boosting-based discriminative feature resembling scheme, which incorporates the ranking distortions into the boosting objective to optimize the feature descriptor towards optimal action retrieval. Extensive evaluations have been conducted on a 60-hour YouTube video dataset as well as the KTH human motion benchmark with comparisons to the state-of- the-arts.

1.2 Title: Detection of Video Sequences using Compact Signatures

Authors: T. C. Hoad and J. Zobel (2006)

1.2.1 Problem

Existing methods for searching a video to identify the co derivatives have substantial limitations: they are sensitive to degradation of the video; they are expensive to compute; and many are limited to the comparison of whole clips, making them unsuitable for applications such as, monitoring of continuous streams. Most of the previously proposed search methods require direct comparison of the video features between the query clip and the data being searched, which is computationally expensive and sensitive to the changes that can occur during "lossy" processes such as, transcoding or analogue transmission. Existing cut-detection algorithms can reliably determine the position of a cut to within one frame (around 30 to 40 ms). Conceivably, this slight inaccuracy could cause difficulties when matching sequences that are recorded at different frame rates, as the intervals between cuts may be slightly different. Approximate matching techniques have been designed to avoid the problems caused by such inaccuracies.

1.2.2 Findings

Finding co-derivatives using this video signatures requires a search method that is capable of efficiently comparing the signatures to accurately locate similar sequences. The authors have described a novel technique for searching video signatures, based on an approximate string-matching method called local alignment. Once signatures have been calculated for the data being

searched, the query clip using one of the methods was described. This approximate string-matching technique can be used to align the query with segments of the data in the collection. Sections of video in the collection are ranked according to the similarity in the query clip to allow the user to quickly identify the most similar parts of the collection, as well as giving an indication of the degree of similarity. There are many difficulties in adapting local alignment to a video search. Dynamic programming—the method used for computing local alignment has been applied to a video search before, but previous adaptations of this technique have substantial limitations. Lienhart described one approach for using dynamic programming to locate co-derivative regions in video streams, but their proposal had two important weaknesses. First, the optimal alignment was found by computing a distance between every frame in the query and every frame in the data. This involved a computationally expensive vector distance calculation for each pair, making query evaluation costly: queries were reported to run in an approximate real time. Second, similarity between the clips was determined according to the number of exact matches between the frame representations. Since color features are altered by processes such as, transcoding and analogue transmission, exact matches are unlikely in the clips that have been degraded, and so the overall similarity of copies of a clip in different formats is likely to be low.

1.2.3 Solution

The authors have proposed new methods for the coderivative search of video, introducing four new techniques for producing a video signature data: the shot-length, color-shift, centroid-based, and combined methods. Each of these use different properties of the video to produce compact signatures. The most compact signature—the shot-length signature—is based on the pattern of edits in the video. It is fast to search and insensitive to the changes in bit rate and resolution. The color-shift signature captures the way in which the color in the frames changes over time, making it more robust than the existing feature-comparison methods. The centroid-based signature represents the movements of the centroids of luminance in the frames, capturing the motion in the clip using a novel and efficient motion-estimation algorithm. Finally, the combined signature uses evidence from both the centroid and color-shift signatures to produce a composite that has many of the advantages of each. They have also presented methods for searching these signatures, based on local alignment. This efficient algorithm is capable of accurately identifying the coderivative content, even when it comprises of only a small part of a long clip. The local alignment algorithm makes use of a scoring function to determine the similarity between the sequences of symbols; also have introduced several scoring systems that are applicable to the coderivative search and compared their effectiveness.

1.3 Title: Video Copy Detection by Fast Sequence Matching

Authors: Kwang-Ting Cheng, and Mei-Chen Yeh (2009)

1.3.1 Problem

Many existing approaches cast the task of CBCD into a traditional content-based key-frame retrieval framework, since both the tasks follow the query-by-example paradigm. However, CBCD aims at identifying the video copies instead of similar individual frames. For example, two videos of the same scene may be considered similar; however, they are not necessarily copies of each other (based on the definition described above) [12] Thus, methods that solely rely on frame-level similarities can suffer from high false positive rates. Since a video can be naturally represented as a sequence of frames, temporal constraints have been employed in the design of metrics that compare the similarities between the two videos. More specifically, videos are represented as strings of symbols and the edit distance between two symbol strings— defined as the minimal cost of any insertions, deletions, and substitutions of symbols needed to make two strings equal is used for measuring video similarity. Video matching methods based on such a metric have a number of merits. First, the ground distance used to compare the frame descriptors can be seamlessly integrated into the distance measurement. Second, the

temporal order is preserved during matching. Moreover, two similar videos that differ either in length, or in terms of other factors such as, differences in subsequences caused by incorrect key frame detection are likely to obtain a high similarity score based on such a metric.

1.3.2 Findings

1.3.2.1 Fast Matching

The second opportunity for acceleration is to filter the unnecessary alignments that would not possibly lead to successful matching. Suppose two sequences are unrelated; then, the best local alignment is no better than no alignment! Therefore, we can formulate an easier problem from the start: given two sequences, find those alignments that have a similarity exceeding a given threshold.

Inspired by FASTA, a fast algorithm used in bioinformatics for finding similar DNA and protein sequences, the authors first create a visual method called a dot plot. A dot plot puts a dot at (i, j) in an m by n matrix if the similarity between the descriptor i and descriptor j exceeds a specific threshold. Figure 3 shows an example of the dot plot. This plot can be easily constructed by using those inverted files built in the previous step. Note that, the dot plot is sparsed if two videos under comparison are either completely or partially unrelated.

The Smith-Waterman algorithm compares each frame of the query to every frame in the video database. Suppose the length of a query is m, and the size of the database (i.e. the number of frames) is N, the time complexity of the query would be O(mN). In this fast method, first the dot plots are constructed by retrieving the corresponding visual word and its video and frame IDs for each query frame. This step takes O(mL) using the vocabulary tree, where L is the depth of the tree. The complexity of deriving the local alignment from the dot plot depends on the number of dots in the plot. Suppose the size of the dot plot is m by n and it consists of r dots, identifying the diagonals requires a linear time O(m+n), and the remaining processes for examining diagonals would take O(r), overall. Since, in practice, the dots have been distributed sparsely and most dots can be eliminated in the initial filtration process, the overall runtime is generally more linear, rather than quadratic, with respect to the sequence lengths.

1.3.3 Solution

The edit distance is a powerful metric for measuring the dissimilarity between two video sequences and its variants can be used to effectively and efficiently identify the video segments that are locally aligned. In this paper, they have formulated the video copy detection problem as a local alignment problem between video sequences. A two-step method has been proposed to speed up the edit distance-based approaches which address the formulated problem. Results on the MUSCLE VCD benchmark and the MPEG-7 shape dataset demonstrate significant computational improvement without sacrificing the accuracy. One direction of the future research is to design a more effective feature descriptor. Frame representation is very crucial to the detection performance. Although they have showed that a semi-global descriptor provides a promising discriminative power, there is still room for the improvement in comparison with those representations based on local features. Since this method decomposes the representation and the indexing/matching process, any frame based representation could be easily incorporated into this framework. Another direction is to explore multiple sequence alignment techniques to find essential content within multiple relevant video streams. This could be a useful tool for creating a summary from huge volumes of near-duplicate videos on video sharing websites.

1.4 Title: Fast and Robust Short Video Clip Search for Copy Detection

Authors: Junsong Yuan, Ling-Yu Duan, Qi Tian, Surendra Ranganath, and Changsheng Xu (2004)

1.4.1 Problem

In this paper, the authors firstly attempt to broadly categorize most existing QVC work into 3 levels: concept based video retrieval, video title identification, and video copy detection. This 3-level categorization is expected to explicitly identify the typical applications, robust

requirements, likely features, and main challenges existing between the mature techniques and the hard performance requirements.

1.4.2 Findings

The authors have focused on the copy detection task, wherein the challenges are mainly due to the design of compact and robust low level features (i.e. An effective signature) and a kind of fast searching mechanism. In order to effectively and robustly characterize the video segments of variable lengths, they have designed a novel global visual feature (a fixed-size 144-d signature) combining the spatial-temporal and the color range information. Different from previous key frame based shot representation, the ambiguity of key frame selection and the difficulty of detecting gradual shot transition could be avoided. Experiments have shown the signature is also insensitive to color shifting and variations from video compression. As this feature can be extracted directly from the MPEG compressed domain, lower computational cost is required. In terms of fast searching, they have employed the active search algorithm. Combining the proposed signature and the active search, an efficient and robust solution for video copy detection has been achieved.

1.4.3 Solution

In this paper, the authors have presented a three-level QVC framework in terms of how to differentiate the diverse "similar" query requests. Although huge amounts of QVC research have been targeted in different aspects (e.g. feature extraction, similarity definition, fast search scheme and database organization), few work has tried to propose such a framework to explicitly identify different requirements and challenges based on rich applications. A closely related work has just tried to differentiate the meanings of "similar" at different temporal levels (i.e. frame, shot, scene and video) and discussed various strategies at those levels. According to the experimental observation and comparisons among different applications, we believe that a better interpretation of the term of "similar" is inherent to the user-oriented intentions. For example, in some circumstances, the retrieval of "similar" instances is to detect the exact duplicate or re-occurrences of the query clip. Sometimes, the "similar" instances may designate the re-edited versions of the original query. Besides, searching "similar" instances could also be the task of finding video segments sharing the same concept or having the same semantic meaning as that of the query. Different bottlenecks and emphasis exist at these different levels. Under the framework, we have provided an efficient and effective solution for video copy detection. Instead of the key frames-based video content representation, the proposed method treats the video segment as a whole, which is able to handle video clips of variable length (e.g. a sub-shot, a shot, or a group of shots). However, it does not require any explicit and exact shot boundary detection. The proposed OPD histogram has experimentally proved to be a useful complement to the CCD descriptor. Such an ordinal feature can also reflect a global distribution within a video segment by the accumulation of multiple frames. However, the temporal order of the frames within a video sequence has not yet been exploited sufficiently in OPD, and also in CCD. Although our signatures are useful for those applications irrespective of different shot order (such as the commercial detection), the lack of frame ordering information may make the signatures less distinguishable. The future work may include how to incorporate temporal information, how to represent the video content more robustly and how to further speed up the search process.

1.5 Title: A Simple but Effective Approach to Video Copy Detection

Authors: Gerhard Roth, Robert Lagani`ere, Patrick Lambert, Ilias Lakhmiri, and Tarik Janati (2010)

1.5.1 Problem

Detecting copies are an important and new topic that provides an alternative to watermarking for the copyright control and other applications. There are many possible solutions to this problem, so it is necessary to provide an evaluation framework. Given a query video, the copy detection task should find the sub-part of that video which matches a section of any of the test videos. The fact that

one-third of the query videos do not have any match in the test collection makes the problem more difficult, and means that the copy detection process must also return "no-match" in certain cases. If a match is found, the copy detection process should specify the match location and length in the query video, and in the test video collection. It should be pointed that, we cannot assume that there is at most one match to the query video in the test video collection, since this test video collection many contain duplicates. It is therefore important for the copy detection process to find all the matches for a given query video, if they exist, and not just the best match.

1.5.2 Findings

1.5.2.1Video Similarity Measurement

Assume we have a database of videos that have been preprocessed in this fashion to create a set of index files, and that we are given a query video. Once we have computed, the index feature counts for the query video by the same preprocessing, and the question is how to find the best match with the feature counts of the video database. This requires us to compare the SURF counts of a video query to the SURF counts for each entry in the video database. Now to actually compare a video query to a database entry, we must compare their sequences of the SURF feature counts for all possible matching video frames. First, we build a cross table where we compute the normalized L1 difference between the vector of 16 SURF feature counts for every query frame and every frame in the database video. When comparing two feature count vectors, the normalization process divides their L1 difference by the sum of the feature counts for the two vectors, and scales the result to be between 0 and 255. If there are K frames in the query video, and N video frames in the database video, then this takes O(KN) time. However, they have sub-sampled the database frames by a factor of at least ten (since typically N >> K), so the time to compute each normalized difference is very small. It is also clear that, the process of creating this cross-table of normalized differences of the SURF feature counts can be easily parallelized.

1.5.3 Solution

In terms of performance, this method has ranked in the middle of the TRECIVD submissions for copy detection. However, on the more common transformations (gamma and image quality change), it performs well. In terms of the localization, this approach is one of the best, as would be expected from a system that compares the video sequences for overlap. More importantly, they believe that this method satisfies many of their design goals. The authors are working on a number of improvements; implementing the process on GPU hardware, having a dynamic frame sub-sampling strategy, improving the decision scoring by a better combination of match length and average match value, and using simpler features (such as Harris or Fast corners).

Conclusion

Data mining refers to the extraction of knowledge or information from a huge data set, this play a very important role for knowledge extraction. Extracting knowledge from the image database is not an easy task, it consists of audio, motion, image, video and text information. Among this complex structure, it is really very difficult to extract the needed knowledge or information. Here the authors have tried to bring some of the existing information technique with their own advantages and disadvantages. The literature brings a clearer idea about what is the current scenario available in the image extraction. This study has brought a conclusion and yet an effective technique is needed in image extraction using data mining technique.

References

[1]. Liujuan Cao A, Rongrong Ji B.N., Yue Gao C, Wei Liu B., and Qi Tian, (2012). *"Mining Spatiotemporal video patterns towards robust action retrieval"*. Elsevier 17.

[2]. X. Wu, C.-W. Ngo, A. Hauptmann, and H.-K. Tan, (2009). "Real-Time Near-Duplicate Elimination for Web Video Search with Content and Context".

[3]. D. Saravanan, V. Somasundaram, (2014). "Matrix Based Sequential Indexing Technique for Video Data Mining". *Journal of Theoretical and Applied Information Technology*, Vol.67, No.3, pp.725-731.

[4]. T. C Hoad and J. Zobel, (2006). "Detection of Video Sequences using Compact Signatures". *ACM Transactions on Information Systems*, Vol.24, No.1, pp.1-50.

[5]. K. Aizawa, Y. Nakamura, and S. Satoh (Eds.), (2004). "Fast and Robust short Video clip search for copy Detection". *PCM 2004*, LNCS 3332, pp.479-488, Springer-Verlag Berlin Heidelberg.

[6]. Gerhard Roth, Robert Laganière, Patrick Lambert, Ilias Lakhmiri, and Tarik Janati, (2010). "A Simple but Effective Approach to Video Copy Detection". *Computer and Robot Vision (CRV), Canadian Conference*, pp.63-70.

[7]. X. Zhou, L. Chen, A. Bouguettaya, Y. Shu, X. Zhou, and J.A. Taylor, (2010). "Adaptive Subspace Symbolization for Content-Based Video Detection," *IEEE Trans. Knowledge and Data Eng.*, Vol.22, No.10, pp.1372-1387.

[8]. D. Saravanan, (2015). "Text information Reterival using Data mining Clustering Technique". *International Journal of Applied Engg. Research*, Vol.10, No.3, pp.7865-7873.

[9]. D. Saravanan, (2015). "Effective Multimedia Content Retrieval". *International Journal of Applied Environmental Sciences*, Vol.10, No.5, pp.1771-17783.

[10]. A. Natsev, R. Rastogi, and K. Shim, (1999). "WALRUS: A Similarity Retrieval Algorithm for Image Databases," in *Proc. ACM SIGMOD Int. Conf. Management of Data*, pp.395-406.

[11]. J. Li, J.Z. Wang, and G. Wiederhold, (2000). "IRM: Integrated Region Matching for Image Retrieval," in *Proc. of the 8th ACM Int. Conf. on Multimedia*, pp.147-156.

[12]. V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, (2004). "Region-based Image Retrieval using an Object Ontology and Relevance Feedback," in *Eurasip Journal on Applied Signal Processing*, Vol.2004, No.6, pp.886-901.

[13]. Mei-Chen Yeh and K. J.Tim Cheng (2009). "Video Copy Detection by Fast Sequence Matching". *Proc. 8th ACM International Conference on Image and Data Retrieval, CIVR 2009*, Greece.

[14]. A. Ronald Tony, D. Saravanan, (2015). Text Taxonomy using Data mining clustering system, *Asian Journal of Information Technology*, Vol.14(3), pp.97-104.

ABOUT THE AUTHOR

Associate Professor, IFHE University, IBS Hyderabad, Telangana, India.