ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

A Survey on Video Event Retrieval using Visual State Binary Embedding Model

Kanchan S. Deshmukh.

M.E Student, Department of Computer Engineering, DYPCOE, Akurdi, SPPU, Pune, India

ABSTRACT: Since from last decade, analysis of video content has gained growing research interest in domain of computer vision and multimedia. In video content analysis, retrieval of event in unconstrained scenarios vital research problem because of large scale unstructured visual information from the video descriptions. There are number of methods and models designed for video event retrieval, but suffered from the various limitations such as scalability, processing speed and efficiency. Designing the efficient, scalable and fast model for video event retrieval by considering visual approach, semantic approach and relevance feedback approach. At first, designing the VSBE model in order to encode the video frames which are containing the important semantic data in binary matrices. This helps to achieve the fast event retrieval under unconstrained scenarios. The approach needs limited key frames from the training event videos for the functioning of hash training so that complexity of computation will be less during training process. Additionally,
applying the pairwise constraints those are generated from the visual states for stretching the events local properties as semantic level in order ensure the accuracy. In second contribution, extending the VSBE model called Extended VSBE (EVSBE) in order address the problem of end user satisfaction and out of event videos by using algorithm of log based relevance feedback. The performance will be evaluated in terms of precision, recall, accuracy and training time.

KEYWORDS: VSBE, EVSBE, MED, CBIR, SVM, NPRF, HASHING

I. INTRODUCTION

Within the final decade, Video pursuits are viewed to be difficult patterns in video streams, which commonly built a nice variety of semantics comparable to more than a few objects, human movements, and scenes. Special from multimedia event detection (MED) that tries to learn reliable classifiers to mechanically notice pre-outlined routine in unknown movies, occasion retrieval, when given a question video, objectives to search for semantically principal movies from enormous video repositories. Unluckily, this lacks effective and scalable strategies to deal at high velocity with big scale video datasets. In multimedia content material evaluation duties, a video can be represented either as a flat vector through feature aggregation, or as a chain of feature vectors. Nevertheless, a flat video vector could lose the interior structure of the video itself. Alternatively, representing a video as a series of function vectors is deemed to broaden the computational complexity significantly, especially once need an efficient search inside significant scale video datasets. In lots of real-time functions the technique of binary embedding, which is customarily referred to as hashing, has been extensively adopted to encode high dimensional characteristic vectors into compact binary codes, leading to quick computation through XOR operators in the Hamming space to approximate the space between feature vectors, as a consequence reaching scalable understanding retrieval. Until now a form of hashing models had been proposed and greatly applied to the near-replica content material search and visible monitoring. Nonetheless, there are a couple of disorders when making use of binary embedding approaches for video occasion evaluation. On one hand, most hashing methods are more commonly designed on the visual stage alternatively than the semantic degree, with an outcome that a semantic hole could exist between the visible representation and occasion description. However, the transformation from the real number area into the binary house may just purpose severe knowledge loss, in particular the lack of the spatial and temporal understanding describing tricky patterns in videos. In order to facilitate quick event retrieval as well as hold as much discriminative know-how as viable, propose an efficient and scalable model of visual state binary embedding (VSBE). On this model, outline a novel metric to evaluate the representativeness of each and every key body in a given video by way of due to the fact that its three significance measures on the video-level, event-level and global-level respectively. A quantity of totally representative key frames are then chosen to sketch the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

semantic cues. The visual expertise contained in these selected key frames can dis- close the main semantics of the movies, and they are descriptive to the corresponding events. Such style of visible understanding from the training video corpus will probably be used to kind the semantic constraints for finding out the binary embedding services. Within the retrieval segment, each and every query video is first represented as a binary matrix, the place every row stands for a key body. Content based image retrieval (CBIR) has received much attention in the last decade, which is motivated by the need to efficiently handle the rapidly growing amount of multimedia data. Content based image retrieval is the technologies that retrieve images from a very large data base by their low level visual features such as colour, texture and shape. It covers versatile areas, such as image segmentation, image feature extraction, representation, mapping of features to semantics, storage and indexing, image similarity distance measurement and retrieval making CBIR system development a challenging task.

Hashing: To accelerate information in large multimedia databases, one of the most promising approach is to embed the high dimensional data into binary code, which is often called hashing. Recently many hashing algorithms have been proposed and these are mainly used in the nearest neighbor search of image data. In addition, cross-media hashing methods have been proposed for heterogeneous media retrieval In many multimedia retrieval applications, binary embedding approaches are mainly designed at the visual rather than the semantic level, so it is difficult to embed the complex structures into binary codes for videos. Until now there has been a very limited number of hashing methods specifically designed for unconstrained video event retrieval.

Video Summarization: Video summarization aims to find the most representative shots and form a condensed synopsis of the whole video, which contains the most important details of the original video. Both unsupervised and supervised approaches have been proposed in the last few years. Inspired by these ideas, it is reasonable to select the most event-descriptive segments from the videos to train the indexing models, which has two advantages: first, it has lower time complexity and memory cost; second, it removes some noisy segments that adversely affect the model training. To our best knowledge, the proposed VSBE model in this paper is the first attempt to adopt the video summarization technique for large-scale video event analysis.

II. LITERATURE SURVEY

2.1 Paper name Scalable Video Event Retrieval by Visual State Binary Embedding [1]
Authors: Litao Yu, Zi Huang, Jiewei Cao, and Heng Tao Shen
Description: In this authors proposed a completely approach to event detection .Here authors proposed a new model called VSBE model which is used to find out which type of event is. Here they first divides the video into keyframes and then extract key frames from them .And then it selects a a conversion of each key frame into binary form and then it compares each frame with each training data sets frames and try to find out which kind of video it is.

2.2 Paper Name: Temporal sequence modeling for video event detection [2]
Authors: Y. Cheng, Q. Fan, S. Pankanti, and A. Choudhary
Description: In this authors proposed a completely unique approach for event detection in video by temporal sequence modeling. Exploiting temporal info has lain at the core of the many approaches for video analysis (i.e., action, activity and event recognition). not alike earlier everything doing temporal exhibiting at linguistics event level, to model temporal dependencies within the knowledge at sub event level while not victimization event annotations. This frees model from ground truth and addresses many limitations in previous work on temporal modeling. Supported this concept, represent a video by a sequence of visual words learnt from the video, and apply the Sequence Memorizer to capture long-range dependencies in an exceedingly temporal context within the visual sequence.

2.3 Paper Name: Video co-summarization: Video summarization by visual co-occurrence [3]
Authors: W.S.Chu,Y. Song, and A.Jaimes
Description: In this authors present a video co-summarization, a completely unique perspective to video summarization that exploits visual co-occurrence across multiple videos. actuated by the observation those vital visual ideas tend to look repeatedly across videos of an equivalent topic, to summarize a video by finding shots that co-occur most often across videos collected employing a topic keyword. The most technical challenge is coping with the exiguity of co-occurring patterns, out of tons of two probably thousands of immaterial shots in videos being thought of. To wear down this challenge, developed a maximal Biclique Finding (MBF) rule that's optimized to search out sparsely

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

co-occurring pat- terns, discarding less co-occurring patterns though they're dominant in one video. Rule is parallelizable with closed-form updates, so will simply proportion to handle an outsized range of videos at the same time. Incontestable the effectiveness of approach on motion capture and self-compiled YouTube datasets.

2.4 Paper Name: Iterative multi view hashing for cross media indexing [4]
Authors: Y.Hu, Z.Jin, H.Ren, D.Cai, and X.He
Description: In this authors study the cross media categorization drawback by learning the discriminative hashing functions to map the multi-view information into a shared playing house. Not solely pregnant within-view similarity is needed to be preserved, additionally incorporate the between-view correlations into the encryption theme, wherever mapping the similar points close and push apart the dissimilar ones. To the present finish, a completely unique hashing rule referred to as unvaried Multi-View Hashing (IMVH) by taking this data into consideration at the same time. To unravel this joint improvement drawback with efficiency, additional develop an unvaried theme to take care of it by employing a lot of versatile division model.

2.5 Paper Name: Caffe: An open source convolutional architecture for fast feature embedding [5]
Authors: Y. Jia
Description: Caffe provides transmission scientists and practitioners with a clean and modifiable framework for progressive deep learning algorithms and a set of reference models. The framework may be a BSD-licensed C++ library with Python and MATLAB bindings for coaching and deploying general purpose convolutional neural networks and alternative deep models expeditiously on artifact architectures. Caffe fits business and web scale media wants by CUDA GPU computation, process over forty million pictures daily on a singleK40 or Titan GPU (2.5 ms per image). By separating model illustration from actual implementation. Caffe permits experimentation and seamless switch among platforms for simple development and preparation from prototyping machines to cloud environments.

III. PROPOSED SYSTEM

Recent years, the process of multimedia based event recognition and retrieval increasing researchers attention because of extensive growth of videos generated by users over Internet as well as various applications such as consumer content management, web video indexing etc. The process of multimedia event retrieval is nothing but complex events recognition automatically from the set of unconstrained videos. This process is very challenging and complex to achieve. There are different solutions introduced in literature so far. Many efforts are made on evaluating the efficacy of low-level features for event recognition and retrieval. But as events are often characterized by similarity in semantics rather than visual appearance, therefore the recent methods are presenting solutions by using high-level semantic concepts in order to assist in the retrieval of events. From the available methods, the most popular method to achieve the scalable event information retrieval from the large video datasets is learning binary embedding called as hashing functions. These methods are utilized for near duplicate search in multimedia. But the limitation of these methods is that are based on visual level approach for near duplicate retrieval rather than semantic level approach. Recently this problem was solved by visual state binary embedding (VSBE) model design for fast and efficient video event retrieval which is based on both visual level and semantic level techniques. The problem with this method is that it does not deal with out of event videos in order to satisfy the end users requirement.

There are number quantity of methods and models suitable for video event retrieval, but experienced the various limitations like scalability, processing speed and efficiency. Designing the efficient, scalable and fast model for video event retrieval by considering visual approach, semantic approach and relevance feedback approach. At first, designing the VSBE model as a way to encode the recording frames which are containing quite semantic data in binary matrices. This helps to offer the fast event retrieval under unconstrained scenarios. The approach needs limited key frames in the training event videos to the functioning of hash training to ensure that complexity of computation will probably be less during training process. Additionally utilizing the pairwise constraints those are generated from the visual states for stretching the events local properties as semantic level to be able ensure that the accuracy. In second contribution, extending the VSBE model called Extended VSBE (EVSBE) so as address the situation of consumer satisfaction and out of event videos by making use of algorithm of log based relevance feedback. The performance is going to be evaluated with regards to precision, recall, accuracy and training time.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

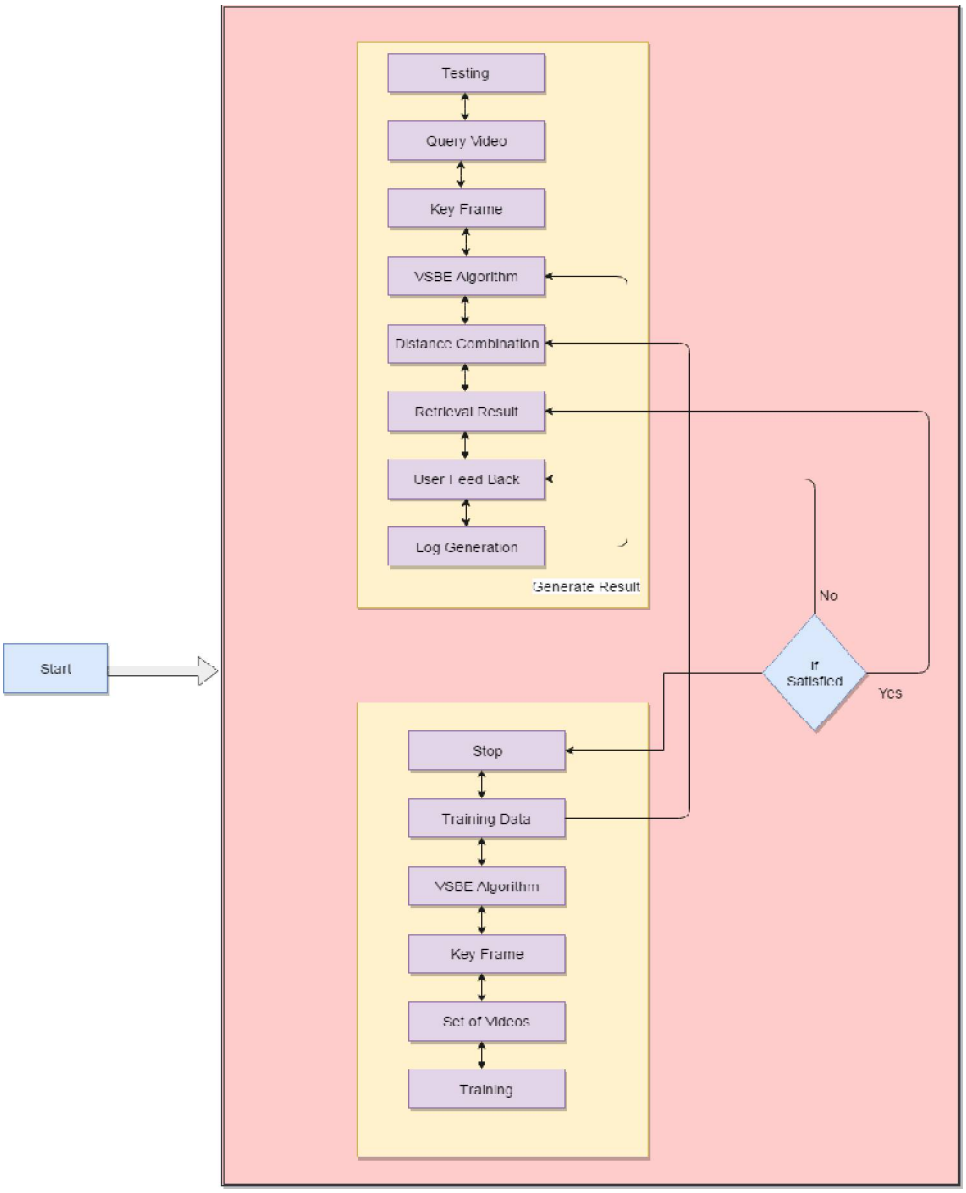

Figure 4.1 Flow Diagram of Proposed System

IV. ALGORITHM

5.1 The algorithm of VSBE.

Input: The selected key frame feature matrix X, constraint matrices U+ and U, hash bit r, enforcement parameter γ, and balance parameters α and β.

Output: Local optimal hash mapping matrix of W, the bias vector b, and the visual states matrix Y.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Randomly initialize W;
Randomly initialize Y and orthogonalize it;
Compute the affinity matrix according to (8);
Compute matrix L based on (9);
repeat
Update W based on (14);

ComputeV = L + λ (U^{+T} U^{+} --- U^{-T} U^{-} + α (XP --- I) ^{T} (XP --- I) + P^{T} P);

Compute Y by eigen decomposition of V;
until Convergence;
Compute b by calculating the median numbers of each
column of Y ;
Return W, b and Y.

V. RESULT

Preparetraining set by applying above all steps on all steps on training videos to extract features. Apply distance computation by using two inputs test query video features and training video features. Present the retrieval events result based on distance matching.

VI. FUTURE WORK

Prepare training set by applying above all steps on all training videos to extract features. Apply distance computation by using two inputs test query video features and training videos features. Present the retrieval events results based on distance matching.

VII. CONCLUSION

Here studied the binary embedding model that is VSBE forfor scalable event retrieval with content based video retrieval in large unconstrained video databases. First, evaluated the representative ability of the key frames from the event-relevant videos, and select the top ranked frames to sketch visual states. Then constructed the pair-wise constraints as prior knowledge to embed the visual states into binary codes at the semantic level. Finally, proposed the VSBE algorithm and its iterative solution. The experimental results on the challenging TRECVID MED dataset demonstrated that proposed VSBE model can both accelerate the training procedure, and boost retrieval accuracy. Also noticed that although proposed VSBE model simultaneously considers the static and dynamic properties of the videos, the information loss of the video representation is still severe after binary embedding, especially when there are a large number of null videos (i.e., irrelevant to any pre-defined events) in the testing set. Another issue is that the VSBE model is flexible when there are new event categories, i.e., the model cannot be incrementally trained. Here, proposing a new method for relevance feedback. It is combination of two existing techniques of relevance feedback scheme: query point movement and query expansion. Taking advantage of irrelevant images and advantages of both traditional techniques, method gives better results. By combining both techniques of query modification that are query point movement and query expansion, these two approaches can benefit from irrelevant examples. This method does not require complex computations, but offers very significant improvements in accuracy compared to traditional techniques.

REFERENCES

[1]Litao Yu, Zi Huang, Jiewei Cao, and Heng Tao Shen, Scalable Video Event Retrieval by Visual State Binary Embedding, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 18, NO. 8, AUGUST 2016
[2] Y. Cheng, Q. Fan, S. Pankanti, and A. Choudhary, Temporal sequence modeling for video event detection, in Proc. IEEE Conput. Vis. Pattern Recog., Jun. 2014, pp. 22352242.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

[3] W.-S. Chu,Y. Song, andA. Jaimes, Video co-summarization: Video summarization by visual co-occurrence, in Proc. IEEE Comput. Vis. Pattern Recog., Jun. 2015, pp. 35843592.
[4] Y. Hu, Z. Jin, H. Ren, D. Cai, and X. He, Iterative multi-view hashing for cross media indexing, in Proc. 22nd ACM Int. Conf. Multimedia, 2014, pp. 527536.
[5] Y. Jia, Caffe: An open source convolutional architecture for fast feature embedding, 2013. [Online]. Available: http://cae.berkeleyvision.org.
[6] X. Li, C. Shen, A. Dick, and A. van den Hengel, Learning compact binary codes for visual tracking, in Proc. IEEE Comput. Vis. Pattern Recog., Jun. 2013, pp. 24192426.
[7] L. Liu, L. Shao, and P. Rockett, Boosted key-frame selection and correlated pyramidal motion-feature representation for human action recognition, Pattern Recog., vol. 46, no. 7, pp. 18101818, 2013.
[8] W. Liu, J.Wang, R. Ji, Y.-G. Jiang, and S.-F. Chang, Supervised hashing with kernels, in Proc. IEEE Comput. Vis. Pattern Recog., Jun. 2012,pp. 20742081.
[9] S. Lu, Z.Wang, T. Mei, G. Guan, and D. D. Feng, A bag-of-importance model with locality-constrained coding based feature learning for video summarization, IEEE Trans. ultimedia, vol. 16, no. 6, pp. 14971509, Oct. 2014.
[10] X. Zhu, Z. Huang, H. T. Shen, and X. Zhao, "Linear cross-modal hashing for efficient multimedia search," in *Proc. 21st ACM Int. Conf. Multimedia*, 2013, pp. 143–152.
[11] J. Revaud, M. Douze, C. Schmid, and H. J´egou, "Event retrieval in large video collections with circulant temporal encoding," in *Proc. IEEE Comput.Vis. Pattern Recog.*, Jun. 2013, pp. 2459–2466.