Video Searching Using Free-Hand Sketches

|  |  |  |
| --- | --- | --- |
| Decena, Kervin Levi  BS Computer Science  kedecena@up.edu.ph | Doria, Ariel B.  BS Computer Science  [abdoria@up.edu.ph](mailto:abdoria@up.edu.ph) | Monserrat, Toni-Jan Keith  tjmonsi@gmail.com |

ABSTRACT

Searching videos using descriptive text as input is a common video-searching method on the internet. Though common, the method is quite inefficient since not all the needed information found can be filtered through descriptive text. On this paper, a method on comparing free-hand sketches to find similar scenes from a pool of video is proposed. On this method, each frame from the videos will be processed using edge detection algorithm and Binary Coherent Edge Descriptor (BiCE) and compared to the input sketches. Indicating threshold in similarity between successive frames will apply to optimize searching. To test the proposed method, ten videos with almost five-minute duration will be used as experiment videos and twenty sketches as experiment input, wherein ten sketches are expected to be found on the video. If half of the results are correct, then it can be concluded that the method is useful while user preference result will dictate if the new technique have potential use in internet video searching.

INTRODUCTION

Most search engines over the internet use text input for searching a particular video. However, descriptive text searching does not encompass all objects and scenarios contained in the video clips. Descriptive texts or metadata are usually limited to title, description, author, and duration of the video. Aside from limited information, another problem on descriptive texts is that authors can write inaccurate, irrelevant, or incomplete descriptions on the video. This scenario is common on the internet as most of the titles and descriptions of online videos do not describe the actual objects and events that can be found inside the video clips. In effect, search engines give faulty results forcing the users to find alternative ways or strategies to find videos that they are looking for.

Watching and scanning each result is an alternative way that the user can do to make sure that the video result are relevant to the needed information. However, this strategy is tedious and time-consuming. Watching every video from search engine results is not efficient and calls for a better solution to solve the problem. Aside from being inefficient, the scenario also shows that the purpose of a search engine is not absolutely achieved. Therefore, a supplementary application is needed to ensure the relevance and correctness of video search results. A need for accompanying application arises that will completely specify and get the needed information from the user to serve as the basis of searching scene from the videos.

We proposed a solution that uses sketch images as inputs for searching videos. With this type of input, user can easily express the scene being sought. This application will use local file storage as video database and not the internet platform. Frame of videos in the database will be processed using an edge detection algorithm and BiCE algorithm to get the edges of objects found on the frame. Optimizing the successive frames will be apply since the consecutive frames contains the same object. To find the needed image, the input sketch will be transformed into a binary image and will be compared to each processed frame from the videos. With this proposed method, the application will increase the assurance that the search results contain the object or scenarios that the user is looking for. In addition, this video search application will be optimal in terms of running time and memory space.

RELATED WORKS

Many works have established methods on searching videos without focusing the information provided by the video’s metadata. Deng and Manjunath used the primary features, such as color, texture and motion for indexing videos. Their system consists of automatic video partition, feature extraction, video search and retrieval [4]. Chang et al. used visual cues in obtaining accurate video object information [2]. In the second paper of Chang, Chen, and Sundaram, they developed the first online video search engine that supports automatic object-based indexing and spatio-temporal queries. They included client-server architecture in which the client can send sketch over the network while the query server is responsible for finding the features in the pool of videos [3].

In terms of fast searching algorithm, Yuan et al. used index-structure method to search short video clips by employing multi-resolution kd-tree to complete exact K-Nearest Neighbor Query and range query. The purpose of the study is to the precisely and quickly search out short video segments having similar content [7]. On the other hand, Yuan et al. (2004) used the Query by video clip to search video by including feature extraction, similarity measure, database organization and search or query scheme. They employed active search algorithm and applied signature in video frames to robustly find videos [8].

In 2014, Benjamin W. et al implemented Juxtapoze, an application that allows user to sketch digitally while the application suggests clipart that match the sketch in real-time. The application uses the concept of Binary Coherent Edge descriptors (BiCE) [1]. BiCE was introduced by Zitnick C. in his paper published on 2010. BiCE gives an output of binarized edge histogram for each patch in the image and encodes the position, orientation and local linear length of edges in the patch to identify the presence of edges [9].

Aside from BiCE, another image processing technique to be used in this research is edge detection. Maini, R. and Aggarwal, H. compared five edge detection techniques: Sobel, Roberts Cross, Prewitt’s, Laplacian of Gaussian, and Canny. Among these techniques, Canny Edge Detector produces the best result [5]. In support, according to Srinivas, P., Malathilatha, Y. L., and Prasad, D. M. (2013), Canny is an efficient edge detection technique. On their research, they also tested five edge detection algorithms for a traffic monitoring system. Their results prove the effectiveness of Canny Edge Detector as the system counted the vehicles with increased accuracy when Canny is used [6].

In the proposed application, we will use the processed sketch image as point of comparison to locate videos that contain the objects and scenes needed by the user. Unlike the previous studies, the proposed application will use both the BiCE and edge detection. Also, another difference of the proposed application is that it searches through the user’s local storage. Thus, this application is offline and will be a standalone application. Lastly, the application should be more effective and efficient compared to the existing video searching methods.

# METHODOLOGY

In this research, a method on comparing sketches to find similar scenes from a pool of video is proposed. With the use of sketches as inputs, the user can completely express the desired object he/she wants to find over the video collection. Since the title of a video clip does not always describe the scenes found in the video, the user will lessen the time to browse and watch the prospected video clip to make sure that the certain video contains the desired scene.

On the proposed application, the user will sketch using the panel provided by the application. The user can input regular or irregular shapes, lines and points over the plane. With this kind of input, basis for comparison will be more specific and detailed. Texture and color of an object will not be included since the edges and lines are the only basis needed for comparing images. Form of an object almost distinguishes it from the other object so edges is mainly the comparison of basis of this study. After inputting the free-hand sketches, the application will ask the directory of the video pool. All the video files under the selected directory will undergo the process of searching scenes that resemble the input sketch. Using this technique, the user can easily visualize the directory where she/he wants to include in the searching process. It also lessen the uploading time of the needed video in the searching environment.

On the process of searching, the videos will be treated as sequences of images. These images will undergo edge detection before it can be compared to the input sketch. Canny edge detector will be the chosen algorithm as it is proven to be effective and efficient. After undergoing edge detection, BiCE algorithm will be used. The technique will divide the image into patches using an overlying grid and the edges on each patch will be gotten. Using this technique, edges at the very detailed object can be retrieved. The edges found will serve as the comparison entity between the user’s sketch and the objects that can be found in the extracted image.

To minimize the time of comparison, optimization of consecutive frames with great similarity will also be considered. There will be a threshold that dictates whether a frame is similar to its subsequent frame. If the subsequent frame is found to be almost the same as the prior frame, the subsequent frame will not be processed and compared. In this method, not only the time of comparison will be lessen but also the memory space allocated because it will avoid multiple detection of similar consecutive frames that match the sketch.

All the images that resemble the input sketch will be noted and the process proceeds from one frame to the next, and from one video to another. The process will take multithreaded to consume smaller amount of time and space. While analyzing all the videos, the results which contains the similar frames will be displayed. Each of the results will include the title of the video where it is found, and the time the frame appears on its video. The application will not wait for the program to finish searching every video before it display the result. The results shall be shown as soon as it is detected.

The rationale behind this proposed solution is to search videos with optimal amount of time and space while maintaining accuracy.

# EVALUATION

We can evaluate the accuracy of the application via experimental tests and via user satisfaction rating. To know if the sketch-based video searching is better than text-based searching in terms of accuracy and time of searching a certain object in the video, the experiment consists of ten videos with almost five-minute duration each stored in a specific folder. The user satisfaction rating will need 30 participants from the host university who will try the application and test its degree of usefulness and speed in video searching.

In measuring the accuracy, there will be twenty sketches as input, wherein ten sketches are expected to be found on the videos. These sketches consist of shapes, lines and points. The sketch that will be checked are manually done and if half of the input sketch correctly specified the time in the video where it can be found, then it can be considered as reliable and accurate.

In measuring the time, the start of measure will start when the queries are sent and ends when the all the results are displayed. It will be compared to the time of searching using the metadata and descriptive text.

User satisfaction rating will be implemented in considering the users’ preference in searching the video over video collection. Thirty users will try the application. They will be asked which searching method is more convenient to use in finding the interest data over a video collection. If at least half of the users are satisfied with the application, then it can be considered that the proposed solution may have potential use in online video searching.

TIMELINE

**January-** *Data Gathering*. Existing algorithms and programs to be used in the application will be gathered. Preparation of the machines to be used and acquiring software to be use in the application.

**February-** *Implementation of code*. Programs and applications needed for the study will be implemented using the prospective programming language. Applying the algorithm in to series of code and applying local testing for the input and output derivation.

**March-** *Testing the application.* Using experiment and user study method, the accomplished application will be tested. Measuring the metric to test the accuracy and time in using the application based on the evaluation method used.

**April-** *Assessment of the result.* Interpreting the results and finding new information will be done. Comparing the result with previous method and improving the new proposed method will be made.

**May-** *Concluding results.* Finalizing the new information gathered from the study. Paper integrations of the results and findings in the implementation study.

# REFERENCES

1. Benjamin, W., Chandrasegaran, S., Ramanujan, D., Elmqvist, N., Vishwanathan, S. V. N., & Ramani, K. (2014, April). Juxtapoze: supporting serendipity and creative expression in clipart compositions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 341-350). ACM. Retrieved from https://engineering.purdue.edu/cdesign/wp/wp-content/uploads/2014/01/juxtapoze.pdf
2. Chang, S. F., Chen, W., Meng, H. J., Sundaram, H., & Zhong, D. (1997, November). VideoQ: an automated content based video search system using visual cues. In *Proceedings of the fifth ACM international conference on Multimedia* (pp. 313-324). Retrieved from http://www.ee.columbia.edu/ln/dvmm/publications/97/acmpaper.pdf
3. Chang, S. F., Chen, W., & Sundaram, H. (1998, October). VideoQ: a fully automated video retrieval system using motion sketches. In *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on*(pp. 270-271). IEEE. Retrieved from http://ame2.asu.edu/faculty/hs/pubs/wacv98.pdf
4. Deng, Y., & Manjunath, B. S. (1997, October). Content-based search of video using color, texture, and motion. In *Image Processing, 1997. Proceedings., International Conference on* (Vol. 2, pp. 534-537). IEEE. Retrieved from https://vision.ece.ucsb.edu/sites/vision.ece.ucsb.edu/files/publications/97ICIPCBS.pdf
5. Maini, R., & Aggarwal, H. (2009). Study and comparison of various image edge detection techniques. *International journal of image processing (IJIP)*, *3*(1), 1-11. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.301.927&rep=rep1&type=pdf
6. Srinivas, P., Malathilatha, Y. L., & Prasad, D. M. (2013). Image Processing Edge Detection Technique used for Traffic Control Problem. *IJCSIT) International Journal of Computer Science and Information Technologies*, *4*(1), 17-20. Retrieved from http://www.ijcsit.com/docs/Volume%204/Vol4Issue1/ijcsit2013040105.pdf
7. Yuan, J., Duan, L.Y., Tian, Q., & Xu, C. (2004, October). Fast and robust short video clip search using an index structure. In *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval* (pp. 61-68). ACM. Retrieved from http://users.eecs.northwestern.edu/~jyu410/index\_files/papers/p61-Yuan\_Duan\_Tian\_MIR04.pdf
8. Yuan, J., Duan, L.Y., Tian, Q., Ranganath, S., & Xu, C. (2005). Fast and robust short video clip search for copy detection. In *Advances in Multimedia Information Processing-PCM 2004* (pp. 479-488). Springer Berlin Heidelberg. Retrieved from http://www3.ntu.edu.sg/home/jsyuan/index\_files/papers/Yuan\_Duan\_Tian\_PCM04.pdf
9. Zitnick, C. L. (2010). Binary coherent edge descriptors. In *Computer Vision–ECCV 2010* (pp. 170-182). Springer Berlin Heidelberg.