OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Real-Time Near-Duplicate Elimination for Web Video Search with Content and Context

A Srisivasulu .Y V Adisatyanarayana
Student Master of Computer Applications, Dr.Sgiet, Markapur, Ap, India.
Assoc.Professror, Dr.Sgiet, Ap, India.

ABSTRACT:- Near-duplicate video retrieval is becoming more and more important with the exponential growth of the Web. Though various approaches have been proposed to address this problem, they are mainly focusing on the retrieval accuracy while infeasible to query on Web scale video database in real time. To balance the speed and accuracy aspects, in this paper, we combine the contextual information from time duration, number of views, and thumbnail images with the content analysis derived from color and local points to achieve real-time near-duplicate elimination. The results of 24 popular queries retrieved from YouTube show that the proposed approach integrating content and context can reach real-time novelty re-ranking of web videos with extremely high efficiency, where the majority of duplicates can be rapidly detected and removed from the top rankings. The speedup of the proposed approach can reach 164 times faster than the effective hierarchical method, with just a slight loss of performance.

I. INTRODUCTION

The exponential growth of social media in Web2.0, the huge volume of videos being transmitted and **searched** on the Internet has increased rapidly. Users can capture videos by mobile phones, video camcorders, or directly obtain videos from the web, and then distribute them again with some changes. For example, users upload 65000 new videos each day on video sharing website YouTube and the daily video views were over 100 million in July 2006[1]. Among these huge volumes of videos, there exist large numbers of duplicate and near-duplicate videos.

Near-duplicate web videos are identical or approximately identical videos close to the exact duplicate of each other,which have similar time duration/length, but different in file formats, encoding parameters, photometric variations (color, lighting changes), editing operations (caption, logo and border insertion), and certain modifications (frames add/remove). A video is a duplicate of another, if it looks the same, corresponds to approximately the same scene, and does not contain new and important information.. A user would clearly identify the videos as “essentially the same.” Two videos do not have to be pixel-wise identical to be considered duplicates. A user searching for entertaining video content on the web, might not care about individual frames, but the overall content and subjective impression when filtering near-duplicate videos. Exact duplicate videos are a special case of near-duplicate videos, which are frequently returned by video search services.
Based on a sample of 24 popular queries from YouTube [2], Google Video [3] and Yahoo! Video [4], on average there are 27% redundant videos that are duplicate or nearly duplicate to the most popular version of a video in the search results [4]. As a consequence, users are often frustrated when they need to spend significant amount of time to find the videos of interest, having to go through different versions of duplicate or near-duplicate videos streamed over the Internet before arriving at an interesting video. An ideal solution would be to return a list which not only maximizes precision with respect to the query, but also novelty (or diversity) of the query topic. To avoid getting overwhelmed by a large number of repeating copies of the same video in any search, efficient near-duplicate video detection and elimination is essential for effective search, retrieval and browsing.

Due to the large variety of near-duplicate web videos ranging from simple formatting to complex editing, near-duplicate detection remains a challenging problem. Among existing content based approaches, many focus on the rapid identification of duplicate videos with global signatures, which are able to handle almost identical videos. However, duplicates with changes in color, lighting and editing artifacts can only be reliably detected through the use of more reliable local features. Local point based methods have demonstrated impressive performance in a wide range of vision-related tasks, and are particularly suitable for detecting near-duplicate web videos having complex variations. However, its potential is unfortunately underscored by matching and scalability issues.

II. LITERATURE SURVEY

A. Video Copy and Similarity Detection

Video copy and similarity detection has been actively studied for its potential in search [6], topic tracking [7] and copyright protection [8]. Various approaches using different features and matching algorithms have been proposed.

Among existing approaches, many emphasize the rapid identification of duplicate videos with compact and reliable global features. These features are generally referred to as signatures or fingerprints which summarize the global statistic of low-level features. Typical features include color, motion and ordinal signature [8], and prototype-based signature [5], [6].

.At a higher level of complexity, more difficult duplicates with changes in background, color, and lighting, require even more intricate and reliable features at region-level. Features such as color, texture and shape can be extracted at the keyframe level, which in turn could be further segmented into multiple region units. However, the issue of segmentation reliability and the granularity selection brings into question the effectiveness of these approaches. Recently, local interest points (keypoints) are shown to be useful for near-duplicate and copy detection [9], [10],.

Fundamentally, the task of near-duplicate detection involves the measurement of redundancy and novelty, which has been explored in text information retrieval [2],. The novelty detection approaches for documents and sentences mainly focus on vector space models and statistical language models to measure the degree of novelty expressed in words.

B. Context Analysis for Retrieval

Contextual information has been actively discussed from different viewpoints, ranging from the spatial, temporal, shape context, to pattern and topical context. Tags and locations are two commonly used context information for image retrieval [11]. Social links have attracted the attention which are used to study the user-to-user, and user-to-photo relations. The user context and social network context were also used to annotate images [12].

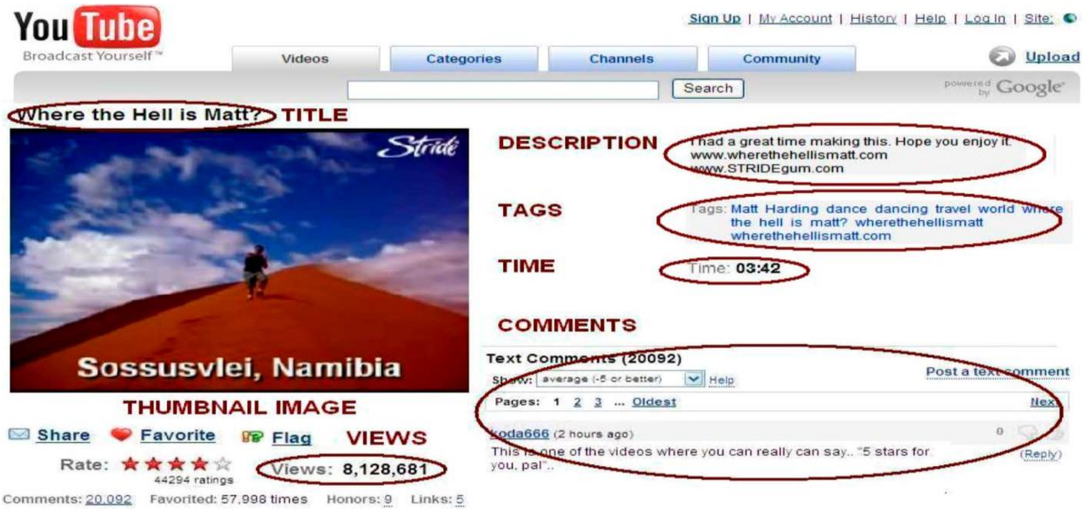
Most of the mentioned works are mainly based on the image sharing website Flickr. However, there has been little research exploring the context information for video sharing websites, such as YouTube. It remains unclear whether contextual resources are also effective for web videos. In particular, the integration of content and context formation for near-duplicate web video elimination has not been seriously addressed.

PRAPOSED WORK

This section details our proposed approach for real-time near-duplicate elimination.

A. Context Cues for Web Videos

One attractive aspect of social media is the abundant amount of context metadata associated with videos as shown in Fig. 1. The metadata includes different aspects of information such as tags, time durations, titles, thumbnail images, number of views, comments, usernames, and so on.



Where the Hell is Matt?

I had a great time making this. Hope you enjoy it.
www.wherethehellismatt.com
www.STRIDEgum.com

Tags: Matt Harding dance dancing travel world where the hell is matt? wherethehellismatt wherethehellismatt.com

Time: 03:42

Text Comments (20092)

Show: average (-5 or better)

koda666 (2 hours ago) 0

This is one of the videos where you can really can say.. "5 stars for you, pal"..

Rate: ★★★★☆ 44294 ratings

Views: 8,128,681

Comments: 20,092 Favorited: 57,998 times Honors: 9 Links: 5

Fig. 1. Rich context information is associated with web videos, e.g., Title, Tags, Thumbnail Image, Description, Time, Views, Comments, etc.

An interesting observation is that near-duplicate web videos more or less have similar *time duration*, with a difference of only a few s. As such, time duration could be a critical feature for efficient filtering of dissimilar videos. A potential risk, nevertheless, is when dealing with complex near-duplicate videos. These web videos could undergo conten t modification by dropping small parts of the videos or adding any arbitrary video segment at the beginning or end of the

videos. This results in near-duplicate videos of different lengths which could not be dealt with if using only time duration.

Thumbnail image is another critical context metadata associated with videos. In social media, thumbnail images are simply extracted from the middle of videos. These images give users a basic impression of the video content. In most cases, it is sufficient to categorize two videos as near-duplicates if:
a) their thumbnail images are near-duplicate and
b) their time durations are close to each other.

View count is an important indication of the popularity of videos. The larger number of views indicates that it is either a relatively important video or an original video. The popularity count of a video carries the implicit information about the intention and interest of the users. We can thus identify the common interest of the public at large.
Context information can accelerate the detection of near duplicates in two manners. First, context can be used jointly with content as a basis to perform near-duplicate decision.
Second context information can be used to uncover the set of dominant near-duplicate groups. In Section III-B, we explore the use of *time duration*, *thumbnail images* and *view counts* for real-time near-duplicate elimination. These contextual cues can be exploited in a complementary way. Time duration provides useful hint for rapid identification of dissimilar (novel) videos.

B. Proposed Framework

With the fact that contextual cue is cheaper to acquire from social media while content is expensive to process, a framework exploiting both cues is proposed as illustrated in Fig. 2. As a preprocessing step, time duration is used to rapidly but coarsely identify the preliminary groups of near-duplicate videos, where we term each group as a *dominant version*. For each dominant version, a seed video, which potentially is the original source from which other videos are derived, is selected. To be efficient, the selection is based on the color histograms of thumbnail images and their view counts. Since seed videos are potential sources, the final step of near duplicate detection is reduced to compare thumbnail images of candidate videos to the selected.

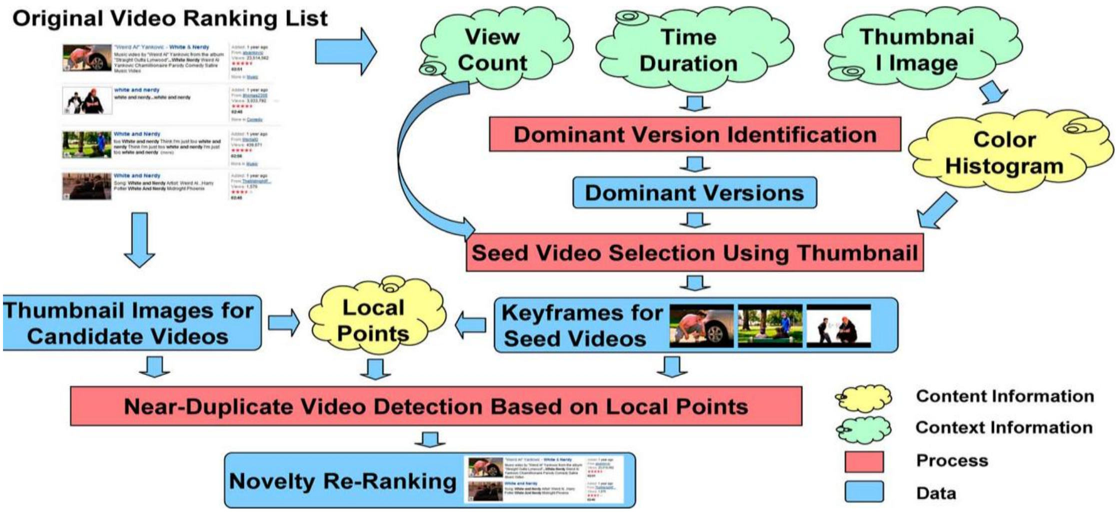

Original Video Ranking List
View Count
Time Duration
Thumbnail Image
Dominant Version Identification
Color Histogram
Dominant Versions
Seed Video Selection Using Thumbnail
Thumbnail Images for Candidate Videos
Local Points
Keyframes for Seed Videos
Near-Duplicate Video Detection Based on Local Points
Content Information
Context Information
Process
Data
Novelty Re-Ranking

Fig. 3. Framework of the real-time near-duplicate elimination with the Integration of content and context information.

The three main processes (dominant version identification, seed video selection, near-duplicate video elimination) are briefly described as below.

Dominant Version Identification: In an entertainment web video search, there usually exist a couple of dominant videos. Other videos in the search result are usually derived from them. For example, there are a short version and a full version in the query “The lion sleeps tonight.” This observation is evident in Table I where the most popular version takes up a large portion of the search result. For each query, dominant version identification is performed by analyzing the distribution of time duration

Seed Video Selection: Seed video selection is performed to pick one seed video for each dominant time. *Seed video* is defined as a potential source from which most near-duplicate videos derived. Videos falling within a specified time range are then filtered by matching the thumbnail images. Color histogram (CH) is extracted for every thumbnail image. Without delving into the content detail inside the videos, the color histogram of thumbnail images and the number of views are combined to select the seed videos. For each seed video, the representative key frames are extracted from the middle part of this video. We call these representative key frames as *prototype keyframes* since they are the potential “prototypes” from which other near-duplicates are transformed, either directly or indirectly.

. *Near-Duplicate Video Elimination:* According to the original ranking from the search engine, each video is compared with the seed videos and every novel video to see whether they are duplicates. The comparison is carried out in two ways, by using context and content information respectively. First, time duration information is treated as a filter to avoid the comparison between videos with considerably different length. Second, content featuresbased on local points extracted from the thumbnail images and prototype keyframes are matched. The first step speeds up detection by avoiding unnecessary comparison, while the second step keeps the expensive content processing as low as possible.

C. Seed Video Selection Using Thumbnail

Once the dominant time durations have been determined, the next step is to select one seed video for each dominant time duration. The objective is to pick one video that has the highest occurrences for each dominant version. Videos having the similar time duration do not necessarily mean they are near-duplicates, as in the case where videos with the same background music might have totally different visual contents, which is a common scenario for web videos. Instead of employing pure content based method, here, we combine the content and context to swiftly select the seed video.

The distance of the color histogram is computed based on the *Euclidean* distance, formulated as

$$d(H_i, H_j) = \sqrt{\sum_{k=1}^{m} (x_k - y_k)^2}$$

$$H_i = (x_1, \ldots, x_m), H_j = (y_1, \ldots, y_m).$$

where , and . The set of videos that fall into the largest distance bin thus forms the dominant near-duplicate group for the corresponding dominant time duration.

D. *Near-Duplicate Video Elimination Using Content*

Once the seed videos are selected, the elimination step willtake place. The objective of search result novelty re-ranking is to list all the novel videos while maintaining the relevance order. To combine query relevance and novelty, each video V_i is computed through a pairwise comparison between V_i and every seed video S_j and previously ranked novel video N_j . The redundancy is calculated by

$$R(V_i\ S_1, \ldots, S_k, N_1, \ldots, N_m) \\ = \mathrm{Max}(\max_{1 \le j \le m} R(V_i\ N_j), \max_{1 \le j \le k} R(V_i\ S_j)).$$

III. CONCLUSION

Social web provides a platform for users to produce, share, view, and comment videos. Huge number of web videos is uploaded each day. Among them, there exist a large portion of near-duplicate videos. . Different sets of context information are considered jointly to perform different tasks at various stages of redundancy elimination. Time duration information is used to filter out dissimilar videos. Experiments on 24 popular queries retrieved from YouTube showed that the proposed method that integrates the content and context information dramatically improve the detection efficiency. The speedup of performance is around 164 times faster than the effective hierarchical method proposed in [13].

In this paper, time duration and thumbnail image are two critical context features used to eliminate the near-duplicate web Authorized licensed videos. User-supplied titles, tags and other text description attached to web videos are usually inaccurate, ambiguous, and even erroneous for video sharing websites. However, among the noisy text information, there exist some useful cues worth exploring, such as episode number, named entities, which provide useful information for the novelty detection.

REFERENCES

[1]. X. Wu, A. G. Hauptmann, and C.-W. Ngo, “Practical elimination of near-duplicates from web video search,” in Proc. ACM Conf. Multimedia,

[2]. Augsburg, Germany, Sep. 2007, pp. 218–227.

[3]. YouTube, [Online]. Available: http://www.youtube.com, Available [3] Google Video, [Online]. Available: http://video.google.com, Available

[4]. Wikipedia, [Online]. Available: http://en.wikipedia.org/wiki/Youtube
Jaimes, “Conceptual Structures and Computational Methods for Indexing and Organization of Visual Information,” Ph.D

[5]. S. C. Cheung and A. Zakhor, “Fast similarity search and clustering of video sequences on the world-wide-web,” IEEE Trans. Circuits Syst.

[6]. Video Technol., vol. 7, no. 3, pp. 524–537, Jun. 2005

[7]. X. Wu, C.-W. Ngo, and Q. Li, “Threading and autodocumenting news videos,” IEEE Signal Process. Mag., vol. 23, no. 2, pp. 59–68, Mar. 2006.

[8]. Y. Ke, R. Sukthankar, and L. Huston, “Efficient near-duplicate detection and sub-image retrieval,” in Proc. ACM Conf. Multimedia, Oct. 2004, pp. 869–876

[9]. Jaimes, “Conceptual Structures and Computational Methods for Indexing and Organization of Visual Information,” Ph.D. dissertation, Columbia Univ., New York, 2003.

[10]. Joly, O. Buisson, and C. Frelicot, “Content-based copy retrieval using distortion-based probabilistic similarity search,” IEEE Trans.

[11]. Multimedia, vol. 9, no. 2, pp. 293–306, Feb. 2007.

[12]. L. Kennedy et al., “How Flickr helps us make sense of the world: Context and content in community-contributed media collections,” in Proc.ACM Conf. Multimedia, Augsburg, Germany, Sep. 2007, pp. 631–640.

[13]. B. Shevade, H. Sundaram, and L. Xie, “Modeling personal and socialnetwork context for event annotation in images,” in Proc. JCDL, Vancouver,BC, Canada, Jun. 2007, pp. 127–134.