Near-Duplicate Segments based news web video event mining

Chengde Zhang [a], Dianting Liu [b], Xiao Wu [a,*], Guiru Zhao [c], Mei-Ling Shyu [b], Qiang Peng [a]

[a] School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China
[b] Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146, USA
[c] China Earthquake Networks Center, Beijing 100000, China

ARTICLE INFO

ABSTRACT

News web videos uploaded by general users usually include lots of post-processing effects (editing, inserted logo, etc.), which bring noise and affect the similarity comparison for news web video event mining. In this paper, a framework based on the concept of Near-Duplicate Segments (NDSs) which effectively integrates spatial and temporal information is proposed. After each video being divided into segments, those segments from different videos but sharing similar visual content are clustered into groups. Each group is named as an NDS, which infers the latent content relations among videos. The spatial-temporal local features are extracted and used to represent each video segment, which could effectively capture the main content of news web videos and omit the noise such as the disturbance/influence from video editing. Finally, the visual information is integrated with the textual information. The experiment demonstrates that our proposed framework is more effective than several existing methods with a significant improvement.

© 2015 Elsevier B.V. All rights reserved.
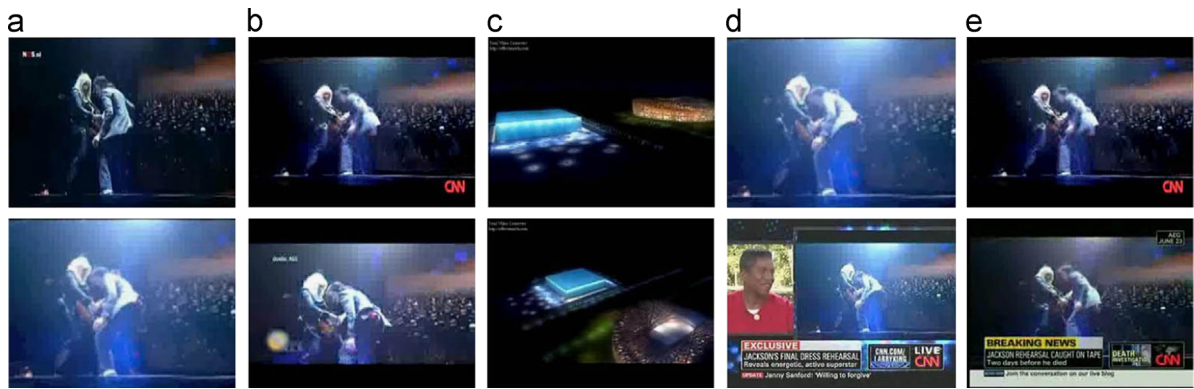
1. Introduction

During the last decade, with the advanced techniques of digital recording and storage devices and the fast development of the Internet platform, it has become much easier for the users to collect and distribute news web videos online. For example, roughly 100 h of new videos are uploaded to YouTube every minute, and more than 1 billion unique users visit YouTube each month [1]. However, when searching a hot topic online, the website may return thousands of videos that are not relevant, and the topic of interest will not be directly observable by simply watching those videos one by one. This distracts the users from the gist of the event and forces them to painstakingly browse the returned videos for an overview of the topic of interest. Based on the above observations, an event mining solution should provide the users a quick overview for each topic and a reduced browsing time.

News web videos are composed of visual and textual information. For visual information, many current event mining algorithms adopt keyframe-based approaches. The detection of Near-Duplicate Keyframes (NDKs) [2,3] plays an important role in measuring video clip similarities and tracking video shots from multi-lingual sources. NDKs have been abundantly used in many applications [4,5]. NDKs can accurately group similar keyframes with differences in video capturing, viewpoints, colors, contrasts, and camera. Fig. 1(a), (b), and (c) shows three examples of different pairs of NDKs in terms of lighting, viewpoints, and camera lens. However, when the keyframes in a video were post-processed for different purposes, multimedia publishers add various post-productions to the same source video (e.g., text/visual effects) as shown in Fig. 1(d) and (e). The post-productions would significantly change the original

* Corresponding author.
E-mail address: wuxiaohk@home.swjtu.edu.cn (X. Wu).

Fig. 1. (a), (b) and (c) are three different pairs of NDKs with differences in lighting, viewpoint, and camera lens. (d) and (e) are examples of added diverse post-productions, which are not successfully grouped into the NDKs of (a) and (b).

visual content of the frames, and therefore bring great challenges to the similarity detection among keyframes. We notice that the post-productions keep a relatively static appearance on the screen, so that hopefully using non-static information (e.g., spatial-temporal features) of post-edited videos could help extract event-related content. Inspired by NDKs [2], we define a new concept called Near-Duplicate Segments (NDSs) to represent the latent connection among videos. NDSs can effectively capture the main content of news web videos and omit the disturbance/influence from video editing. Therefore, NDSs are used to group content-related video segments from different videos. An NDS plays the same role as the hot terms in the text mining field.

In addition to the visual features, textual features (e.g., title, tags) also provide content-related information of the videos. For textual information, different people would use different words to describe the same object, due to their different backgrounds in education, culture, or country. Motivated by the fact that noisy terms (titles/tags) would impact the accuracy of textual information, an Adaptive Association Rule Mining (AARM) approach was developed to mine more semantic related terms [5–8]. Moreover, the results show that AARM can further enhance the robustness of the distribution characteristics of terms in NDSs. Finally, visual and textual information are integrated for news web video event mining.

The rest of this paper is organized as follows. Section 2 gives a brief overview of related work. The details of the NDSs construction are discussed in Section 3. The progress of news web video event mining is presented in Section 4. The empirical study is presented in Section 5. Finally, Section 6 concludes our work.

2. Related work

2.1. Near-duplicate video detection and retrieval

Near-duplicate video detection and retrieval has been a hot research topic and it has attracted quite a lot of attention from the community. Extensive researches [9–15] have been conducted in the last decade. Comprehensive survey on this topic can be found in [10,11,13]. A comparative study of methods for video copy detection was introduced in [13]. Different state-of-the-art techniques,

using various kinds of descriptors and voting functions, were described: global video descriptors based on spatial and temporal features; local descriptors based on spatial, temporal as well as spatio-temporal information. Latest improvements and progress in near-duplicate video retrieval, as well as related topics including low-level feature extraction, signature generation, and high-dimensional indexing, have been investigated in [10]. A rich body of research has been devoted, among which local point based methods [2,3,9,10,12,13,15,16] have become increasingly popular in this field, due to their advantages in handling geometric and photometric variations.

Near-duplicate web video detection was originally explored in [9], which combines the global signature derived from color histogram and local point based pairwise comparison among keyframes. To balance the speed and accuracy aspects, the contextual information from time duration, number of views, and thumbnail images with the content analysis derived from color and local points were further integrated to achieve real-time near-duplicate elimination [16]. Based on the intuition that near-duplicate videos should preserve strong information correlation in spite of intensive content changes, a new method was proposed in [17], which exploits video information correlation instead of content similarity. Canonical Correlation Analysis (CCA) for video information correlation discovery was deployed for heavily changed near-duplicate videos. Effective retrieval with stronger tolerance was achieved by replacing video-content similarity measures with information correlation analysis. To efficiently search for similar videos by local features, an effective near-duplicate retrieval approach was proposed in [15], which integrates enhanced weak geometric constraint (E-WGC), Hough Transform (HT), and reverse entropy (RE). Three aspects were considered: indexing of local visual features, fast pruning of false matches at frame-level, and localization of Near-Duplicate Segments at video-level. Experiments demonstrate the effectiveness and efficiency of the proposed method, which achieved the state-of-the-art performance in near-duplicate video retrieval. A novel method was proposed in [14] to address the efficiency and scalability issues for near-duplicate web video retrieval, a compact spatio-temporal feature was introduced to represent videos and an efficient data