Integration of Visual Temporal and Textual Distribution Information for News Video Mining

Prof Shivamurthy R C, Tauseef Ahmed S S
Department of Computer Science, Akshaya Institute of Technology, Tumkur
mail4tauseef@gmail.com, shivamurthyrc@gmail.com,

ABSTRACT- News web videos exhibit several characteristics, including a limited number of features, noisy text information, and error in near-duplicate key frames (NDK) detection. In this paper, a novel framework is proposed to better group the associated web videos to events. First, the data preprocessing stage performs feature selection and tag relevance learning. Next, multiple correspondence analysis is applied to explore the correlations between terms and events with the assistance of visual information. Finally, a probabilistic model is proposed for news web video event mining, where both visual temporal information and textual distribution information are integrated. . Co-occurrence and visual near duplicate feature trajectory induced from NDKs are combined to calculate the similarity between NDKs and events. It is needed urgently the advanced technologies for organizing, analyzing, representing, indexing, filtering, retrieving and mining the vast amount of videos to retrieve specific information based on video content effectively, and to provide better ways for entertainment and multimedia applications.

Keywords - Co-occurrence, multiple correspondence analysis (MCA), near-duplicate key frames (NDK), news web video event mining, trajectory.

I. INTRODUCTION

Advances in the media and entertainment industries, including streaming audio and digital TV, present new challenges for managing and accessing large audio-visual collections. Search engines and video sharing websites such as YouTube, YouKu, Google, make it convenient for the users to access relevant news web videos. News wires like CNN, BBC, and CCTV also publish news videos. 2014data show that over 100 h of videos are uploaded to YouTube every minute, and six billion hours of videos are watched each month on YouTube. These facts demonstrate a new challenge for the users to grasp the major events available from searching video databases. To address this need, news web video event mining approaches have been developed. When the users search a topic, most want to know: what happened, why it happened, and how it happened. Major event mining can facilitate more effective news web video browsing and a better understanding of the entire topic through the relationships among events. For example, sample search results of the topic ―London terror attack‖ from YouTube are demonstrated in Fig. 1. The results are mainly ranked by text relevance or popularity, which means that a thumbnail image and its corresponding sparse set of tags are not sufficient to help users understand the main content of the topic. The users have to browse many news web videos in the returned list and even watch most of the videos. This is not only time consuming, but also difficult if thousands of news web videos are returned by a search engine. This situation calls for effective approaches to automatically group relevant news web videos into events, and then mine the relationships among them Visual information suffers from semantic gap and user subjectivity problems, and textual information can be noisy, ambiguous, and sometimes incomplete. Therefore, using either visual or textual information alone for news web video event mining usually yields unsatisfactory results. Such challenges motivate us to utilize both visual and textual features for news web video event mining, with the aim to overcome their shortcomings, while leveraging their advantages.

For visual information, machine learning of visual and text features to perform supervised annotation and some important shots are frequently inserted into videos or reports as a reminder or support of viewpoints, which carry useful information. Because of the unique role of near-duplicate key frames (NDK) in the news search, topic detection and tracking (TDT) and copyright infringement detection, these duplicate shots/key frames are clustered to form different groups according to visual content.

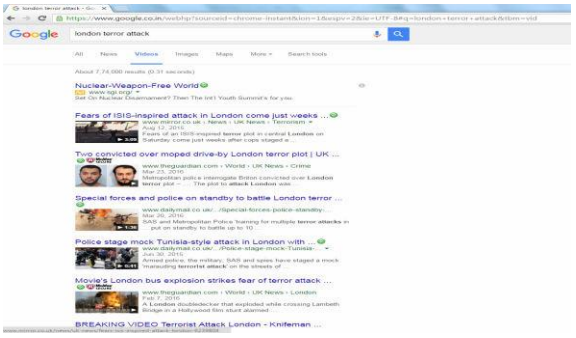

Fig. 1. Search results of ―London terror attack‖ in YouTube on April, 2016.

Visual near-duplicate feature trajectory is an extension of the textual feature trajectory, which models the visual feature distribution along the timeline in a 2-D space with one dimension as time, and the other as the feature weight. In this paper, the visual features are referred to the NDKs. It

can evaluate the importance of an NDK varying over time, which will be used to cluster those NDKs belonging to the same event with different viewpoints. In another way, NDK-within-video information can enhance the robustness of the visual near-duplicate feature trajectory, while it is affected by the NDK editing/detection problem. Therefore, we explore integrating NDK-within-video information and visual near-duplicate feature trajectory, as visual temporal information, to cluster more NDKs belonging to the same event. In contrast with broadcast news videos, where there is textual information from speech transcripts for event mining, news web videos generally have much less textual information (such as titles and tags). This is because news web videos are usually uploaded by users who do not make rich titles and tags for the videos. In addition, these titles and tags are typically noisy, ambiguous, and incomplete. Even worse, the users may include irrelevant hot terms (words) simply to attract attention. Because of the different backgrounds or habits, people use a wide variety of terms to describe the same video. This characteristic would impact the textual distribution information (i.e., the term weight of each word in each NDK) in multiple correspondence analysis (MCA). Therefore, we explore using NDK neighbors to enhance the weights of the terms, which would better represent the high level semantic information of NDKs. In this paper, a novel framework is proposed that integrates the visual temporal information and textual distribution information for news web video event mining. After feature selection and tag relevance learning by neighbor voting, MCA is explored to extract the NDK-level event similarity with the assistance of textual information. Next, both co-occurrence information and visual near-duplicate feature trajectory induced from NDKs are used to detect the similarity between NDKs and events. Finally, in order to integrate visual and textual information for event mining, a hybrid probabilistic model is proposed. Although some of the techniques adopted in this paper, such as NDK, MCA, and visual feature trajectory have been used in previous work, here we bring these techniques together to complement each other. The main novelty and contributions of this paper are as follows.

1) To address the uncontrolled user tagging, ambiguity, and Over personalization, both the MCA similarity measure and neighbor stabilization process (visual neighbor information) are integrated to generate the textual distribution information, which accurately and efficiently learns tag relevance by the visual-content relationship between NDKs to improve the robustness of the textual information. Moreover, it can better explore the degree of correlation between different terms and events.

2) The visual near-duplicate feature trajectory, i.e., the time distribution information of an NDK, is integrated with the NDK-within-video information (co-occurrence) as the visual temporal information to cluster more NDKs belonging to the same event, which obtains robust and accurate features to improve web event mining.

3) A novel unified probabilistic framework is proposed to integrate the visual temporal and textual distribution information for news web video event mining.\

II. RELATED WORK

Definitions

Topic is a seminal event or activity, along with all directly related events and activities. Therefore, we can infer that atopic consists of events and activities. A ―hot topic‖ in is defined as a topic that appears frequently over a period of time. Generally, a ―hot topic‖ has the following characteristics:

(1) It appears in many news stories on one or more news channels;

(2) It has a strong continuity, which means that many different events relevant to the topic are also reported; and

(3) Its popularity changes over time. An event in topic detection and tracking (TDT) is defined as something that happens at a specific time and place, along withal the necessary preconditions and unavoidable consequences. Such an event might be a donation, a game, or a concert performance. An activity in TDT is the connected series of events with a common focus or purpose, which happens in specific places during a given time period. For instance, inactivity may be a campaign, a survey, or an earthquake relief.NDK is a group of key frames that are visually similar, but appear different because of the variations introduced during the acquisition time, lens setting, lighting condition, and editing operation. NDKs have been used in the real-world applications such as TDT.

Topic Detection and Tracking

TDT automatically structures online news articles into topics. TDT detects new topics and tracks known events in text news streams, and hence, many studies inset focused on text data. Studies have been conducted in the multimedia field. The topics were tracked with visual duplicates, which resulted in the concept of NDK. The novelty and redundancy detection was explored in, in which visual duplicates and speech transcripts are integrated to measure the similarity of cross-lingual news stories. With the assistance of NDK constraints, news stories were clustered into topics by constraint-based co clustering. News stories from different TV channels were linked by textual correlation and key frame matching. The retrieval of NDKs plays an important role in measuring the video-clip similarity and tracking video shots of multilingual sources. The system in, segmented news videos into stories and constructed the dependencies among stories as a graph structure. The interface Media Walker supported in browsing the

development of news topics. With the textual correlation and key frame matching, topic clusters were grouped in and news stories from different TV channels were linked in .A video log management model was proposed in, which is comprised of automatic log annotation and user-oriented log search. Sports video semantic event detection was explored in, which is based on the analysis and alignment of the webcast text and broadcast video. Experiments demonstrated that the incorporation of webcast text into sports video analysis significantly facilitates the sports video semantic event detection. Topic discovery was deployed by constructing the duality between stories and textual visual concepts through bipartite graphs. Visual-text time-dependent alignment was explored in to summarize the topics. Text co-occurrence and visual feature trajectory we refused for news web video event mining.

Feature Trajectory

Feature trajectory is a statistical measure for the information retrieval. It evaluates the importance of a feature which varies over time. The characteristics of word trajectory were analyzed in to identify the important and less-important, periodic and periodic words, from the perspective of time-series word signal. A parameter-free probabilistic model was proposed to analyze the time-varying features and to detect the burst events from text streams. The idea of mining hot terms by the timeline analysis was presented in. Hot topics were further extracted using multidimensional sentence modeling grounded on hot terms. GoogleTrends was used to predict the milestone events of a topic.

Multiple Correspondence Analysis (MCA)

Multiple Correspondence Analysis (MCA) can measure the correlations among multiple variables, which is able to capture the correlation among nominal variables. The feature-value pairs and classes can be projected into a 2-D space constructed by the first and second principal components, because of the fact that over 95% of the total variance can be captured by the first two principal coordinates. The function of MCA is to use the textual distribution information to mine the correlation among NDKs and events. However, the characteristics of the textual information (including noisy, ambiguous, incomplete, synonyms, polysemy, and Multilanguage) of news web videos make news web video event mining a challenge problem. The property of the textual information would affect the efficiency of MCA. Therefore, we explore using NDK neighbor information to get more robust and effective semantic relationships among NDKs. MCA and co-occurrence were applied to NDK-level news web video event mining through linear integration. We have extended the study in several ways. First, in the current work, the visual near-duplicate feature trajectory is integrated with the NDK-within-video information (co-occurrence) as the visual temporal information to cluster more NDKs belonging to the same event. Second, for the textual distribution information, we explore using the associations among NDKs to find more related terms (term group) of each NDK. Then we try to make use of the semantic relationships between term groups and events to mine the similarity between an NDK and an event. Third, as the input of MCA, the indicator matrix between text terms and NDKs is in the form of binary values. In this current work, a weighted matrix is adopted to model the distribution of each term in NDKs more accurately.

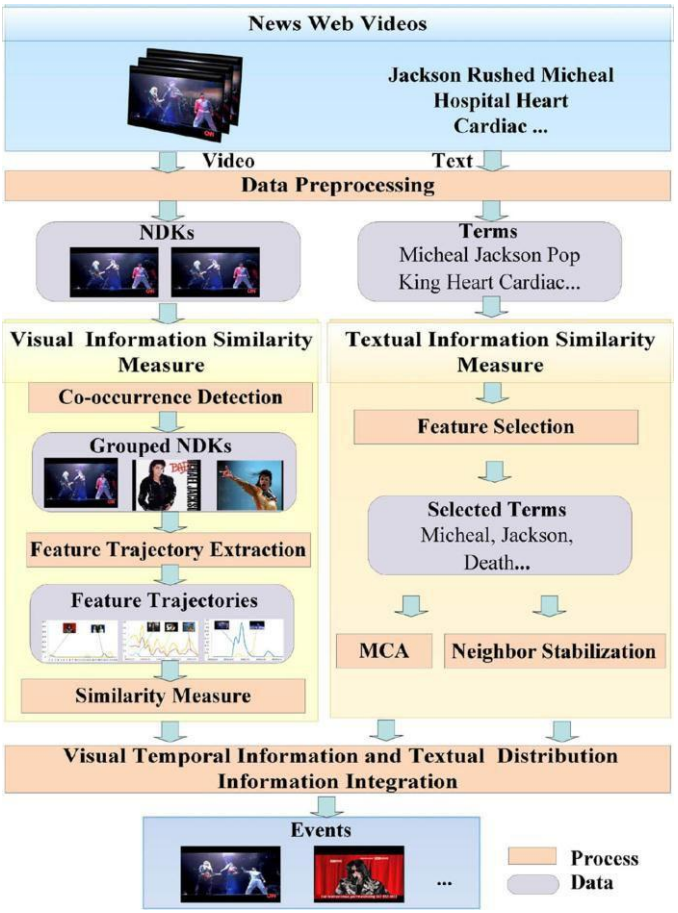

Fig. 2. Proposed framework for news web video event mining.

III. NEWS WEB VIDEO EVENT MINING

This consists of four stages: data preprocessing, textual information similarity measure, visual information similarity measure, and hybrid probabilistic model for integration (see Fig. 2). After getting the correlation between terms and events through visual neighbor information, MCA is used to calculate the similarity between each NDK and event through textual distribution information in NDKs. For the visual information, NDK-within video information and visual near-duplicate feature trajectory are fused to improve the robustness and accuracy of the visual features. Finally, the content-based visual temporal information and textual

distribution information are integrated through the proposed probabilistic model. In this study, event mining is achieved within each topic. The input of our framework is the news web videos returned from a user query. After NDK detection and grouping, a series of NDKs and their corresponding terms can be obtained. As a result, the similarity between each NDK and each event is calculated, and every NDK is assigned to the event with the largest similarity. The output is the classified events.

Data Preprocessing

After shot boundary detection, the middle frame in each videos hot is extracted as the key frame for the shot. Each video can be represented by a series of key frames. Then, NDK detection method is utilized to detect the NDKs among videos. Local points are detected with Harris–Laplace and described by SIFT. The detected NDKs are further grouped to form clusters by transitive closure. For each topic, (1) is applied to calculate the probability of an NDK belonging to each event

$$P(\text{NDK}p, Eq) = |\text{NDK}p \_ Eq| / |\text{NDK}p| \quad (1)$$

where $|\text{NDK}p \_ Eq|$ is the number of common videos between NDKp and event Eq . $|\text{NDK}p|$ is the number of videos whose key frames are in NDKp. Eq is the qth event. Finally, each NDK is marked as the event label which has the largest probability. Ground truth is manually determined according to the search results from Wikipedia and Google. Terms extracted from titles and tags are treated as textual features. Because of the noisy user-supplied tag information, text words are pruned using methods, including word stemming and special character removal.

Textual Information Similarity Measure

Feature selection and visual neighbor information are used to enhance the weights of the representative terms. The less important terms are neglected. MCA is then applied to calculate the MCA-based transaction weights, targeted to bridge the gap between an NDK and terms. In order to apply MCA, each feature in the training data set is discredited into several partitions (i.e., feature-value pairs), and the same partition ranges are used to discredited the testing data set. As a result, the similarity between each feature-value pair and an event is calculated. Finally, the weights between each NDK and all events are calculated by summing the weights of the feature-value pairs along all features.

Visual Information Similarity Measure

NDK-within-video information is first used to measure the similarities among NDKs. Second, the visual feature trajectories induced from NDKs are used to find the highly relevant NDKs as the time distribution feature. Because of the complementary characteristics of the NDK-within-video information and the time distribution information, both are utilized to measure the similarity between an NDK and an event.

Hybrid Probabilistic Model for Integration

A hybrid probabilistic model is proposed for better video event mining, which integrates the visual and textual information. Ultimately, every NDK is grouped to the event with the largest similarity value.

IV. TEXTUAL INFORMATION SIMILARITY MEASURES

Generally, news web videos use titles and tags to describe their content. The features extracted from terms are treated as textual information. There are numerous frequently accompanied terms from titles and tags, which convey useful information. Hence, we propose to mine the textual distribution information as a part of our proposed framework.

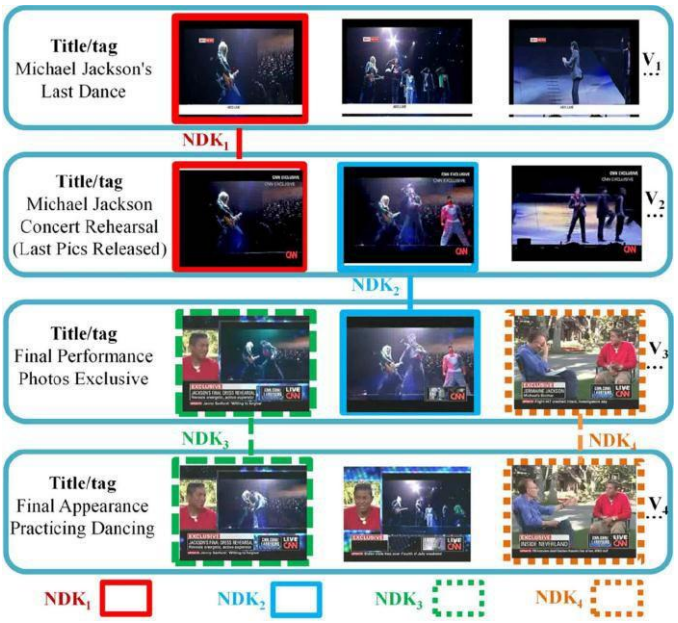

Fig. 3. Each video is described by its title/tag, which is taken as textual information. The features of each term extracted from this textual information are taken as the tag-based features.

Feature Selection and Neighbor Stabilization

For feature selection, the chi-squared statistics with respect tithe classes are adopted to evaluate the importance of the terms with WEKA. Noisy terms would impact the accuracy of MCA. Therefore, word pruning is an unavoidable problem. All terms are ranked in descending order, and then the gaps between neighbors in this sorted list are calculated. Significant terms are ranked above the largest gap based on our empirical study. Finally, the largest gap is used as the

threshold, where all the terms with smaller weights are filtered.

Multiple Correspondence Analysis with Correlation Information

MCA is composed of two steps: training and testing. In order to illustrate the principle of MCA, the training process is taken as an example. First, all the features are combined to Forman indicator matrix with NDKs as the instances (rows), terms and event labels as the categories of variables (columns) (settable I), where in the testing process, it does not need to label the NDKs. After calculating the tiff value of each term distributed in each NDK (textual distribution information), this indicator matrix can be represented in a 2-D table NT (NDKs versus Terms)

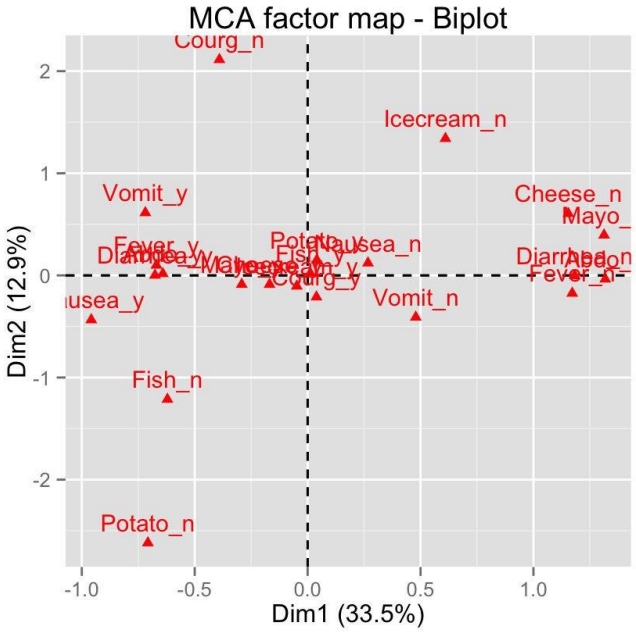

Fig. 4. Geometrical representation of MCA.

Textual Similarity Measures

We calculate the similarities between each NDK and all events using

$$\text{Sims}(\text{NT}_k, E_r) = \gamma \times \text{TW}_{k,r} + (1 - \gamma) \times \text{Sim}_T(\text{NDK}_k, E_r).$$

V. VISUAL INFORMATION SIMILARITY

It is indeed easier to know whether two videos are similar or not rather than to obtain their true labels, because the space of potential labels is very large.

News web videos have fewer textual features than text documents, and these features are often noisy, ambiguous, and incomplete. Thus, video content information compensates for the textual information. The NDKs are taken as the visual ‑terms, ‑which contain rich information. Since NDKs provide a strong cue to link event-relevant videos across sources, languages, and times, an NDK supports key frame redundancy. The features derived from NDKs are treated as visual information, such as NDK-within-video information and visual near duplicate feature trajectory.

Burstyn Period Detection

A web search engine usually returns a large number of search results, mainly according to the text relevance. Some results may not be relevant. For example, when searching ‑Michael Jackson Dead,‖ most users want to view news web videos from the accident on June 25, 2009. However, because of the keyword ‑Michael Jackson,‖ a large number of news web videos about ‑Songs of Michael Jackson‖ will also be listed. Actually, each topic appears frequently within a certain period. To discover the events from the search results, it is essential to locate the burst period according to the video upload time. The method is used to locate the burst region of each topic as follows:

$$\text{Raj} = [\text{tj} - \text{w}, \text{tj} + \text{w}] |\text{Vj}| \geq \alpha\_\text{n} \text{ k=1}(|\text{Vk}|) \text{ n}$$

NDK inside Video Information

Important visual shots are frequently inserted into the relevant videos. These NDKs usually carry useful video content information and can be used to group videos of similar themes into events. There are numerous frequently accompanied NDKs that convey useful information. As shown in Fig. 3, each NDK is composed of a series of key frames, such as NDK1=_*V*1 1, *V*2 1 _ and NDK2=_*V*2 2, *V*3 2 _.

Both NDK1 and NDK2 are appeared in video V2.

Data preprocessing is the first stage. Multiple Correspondence Analysis (MCA) is then applied to explore the correlation between terms and classes, targeting for bridging the gap between NDKs and high-level semantic concepts. Next, co-occurrence information is used to detect the similarity between NDKs and classes using the NDK-within-video information.

Visual Near-duplicate Feature Trajectory The visual near-duplicate feature trajectory models the visual feature distribution along the timeline in a 2-D space (one dimension as time and the other as feature weights). The importance increases proportionally to the number of videos appearing in a certain time-period. Usually, a few representative NDKs will often appear in a period of time, but rarely in other time periods. These NDKs with similar trends imply their consistency. Hence, the relevance of these visual feature trajectories can be clustered to form events.

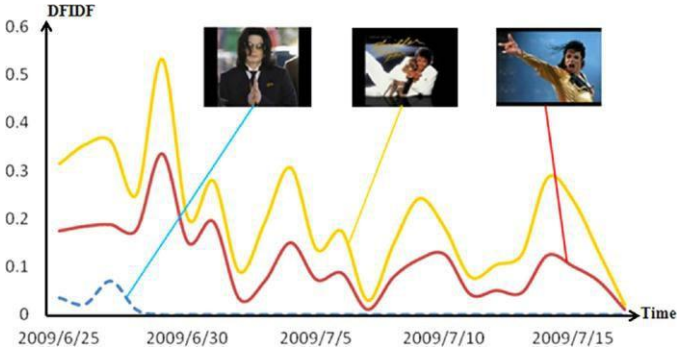

Fig. 5. Different NDKs on the same event could have similar visual near duplicate feature trajectories. For example, the key frames of ―Jackson is dancing in an MTV‖ and ―Cover of MTV‖ share similar visual near-duplicate feature trajectory distributions over time. They belong to the event ―A tribute of Michael Jackson dead.‖ In contrast, the key frame of ―Jackson is praying‖ shows different feature distributions, which belongs to the event ―Sadness of Michael Jackson dead.‖

Visual Similarity Matches

Different NDKs on the same event could have similar visual feature trajectories. For example, both the events ―A tribute of Michael Jackson dead‖ and ―Sadness of Michael Jackson dead‖ belong to the same topic (i.e., ―Michael Jackson dead‖).As shown in Fig. 5, the NDK of ―Jackson is dancing in an MTV‖ and ―Cover of MTV‖ share similar visual near-plicate feature trajectory distributions over time, since they belong tithe same event ―A tribute of Michael Jackson dead.‖ Incontrast, the key frame of ―Jackson is praying‖ shows a different feature distribution, while it belongs to another event ―Sadness of Michael Jackson dead.‖ Similar scenes might have multiple trajectories because of the NDK detection error, video editing, or other reasons. Some NDKs belonging to one group may be falsely detected to form several separated clusters. Moreover, anew NDKs are missed and falsely treated as non-NDK. Here the current visual feature trajectories deviate from the ideal one. Fig. 6shows an example of several NDKs on ―Jackson is dancing in the last rehearsal.‖ Unfortunately, they to the same event ―Last Rehearsal‖ with similar scenes, but demonstrate different trends. Since co-occurrence can further group NDKs with relevant visual content, we believe that it can enhance the robust and consistent characteristics of the visual near duplicate feature

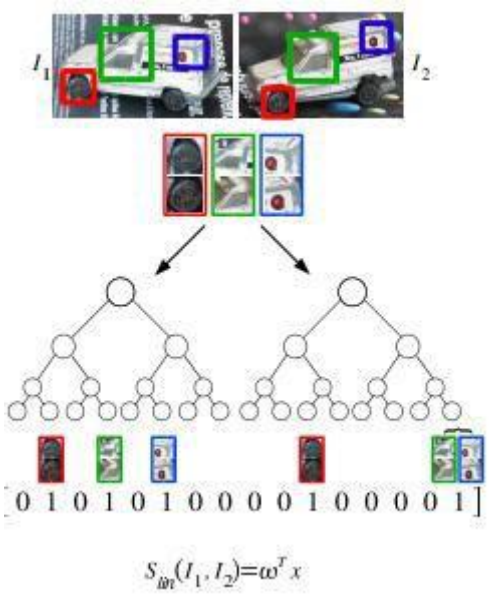

Fig. 6. Example of similar scenes but with multiple trajectories.

trajectories and, thus, improve the performance of news web video event mining.

VI. CONCLUSION

In view of the unique characteristics of news web videos, such as the limited number of features, the unavoidable error in NDK detection, and noisy text information, news web video event mining has been a challenging task. In this paper, a novel hybrid probabilistic framework is proposed for news web video event mining, which integrates the textual and visual information, and aims to solve not only noisy and limited textual information but also the unavoidable video editing and NDK detection problems. Next, a visual neighbor information extraction method proposed to deal with the well-known ambiguity and overly personalized problems. Meanwhile, the purgative textual information helps to bridge the gap between NDKs and the high-level semantic concepts. Moreover, both the textual and visual features with relatively low frequencies are considered a useful information in ourexperiments.

REFERENCES

[1]Statistics. (2014). [Online]. Available: http://www.youtube.com/yt/Press/statistics.html

[2] J. Yuan, Y.-L. Zhao, H. Luan, M. Wang, and T.-S. Chua, ―Memory recall based video search: Finding videos you have seen before based on your memory,‖ *ACM Trans. Multimedia Comput., Commun. Appl.*,
vol. 10, no. 2, pp. 1–21, 2014.

[3] X. Wu, C.-W. Ngo, and Q. Li, ―Threading and autodocumenting newsvideos: A promising solution to rapidly browse news topics,‖ *IEEE SignalProcess. Mag.*, vol. 23, no. 2, pp. 59–68, Mar. 2006.

[4] Y. Ke, R. Sukthankar, and L. Huston, ―Efficient near-duplicate detectionand sub-image retrieval,‖ in *Proc. ACM Int. Conf. Multimedia*, 2004,pp. 869–876.
[5] C. Zhang, X. Wu, M.-L. Shyu, and Q. Peng, ―A novel web video event mining framework with the integration of correlation and co-occurrenceinformation,‖ *J. Comput. Sci. Technol.*, vol. 28, no. 5, pp. 788–796, 2013.
[6] K.-Y. Chen, L. Luesukprasert, S.-C. T. Chou, ―Hot topic extraction based on timeline analysis and multidimensional sentence modeling,‖ *IEEETrans. Knowl. Data Eng.*, vol. 19, no. 8, pp. 1016–1025, Aug. 2007.
[7] C.-W. Ngo, W.-L. Zhao, and Y.-G. Jiang, ―Fast tracking of near-duplicatekey frames in broadcast domain with transitivity propagation,‖ in *Proc.14th ACM Int. Conf. Multimedia*, 2006, pp. 845–854.
[8] X. Wu, C.-W. Ngo, and A. G. Hauptmann, ―Multimodal news story clusteringwith pairwise visual near-duplicate constraint,‖ *IEEE Trans. Multimedia*,vol. 10, no. 2, pp. 188–199, Feb. 2008. [9] J. Cao, C.-W. Ngo, Y.-D. Zhang, and J.-T. Li, ―Tracking web video topics:Discovery, visualization, and monitoring,‖ *IEEE Trans. Circuits Syst.Video Technol.*, vol. 21, no. 12, pp. 1835– 1846, Dec. 2011.