EIGENNEWS: GENERATING AND DELIVERING PERSONALIZED NEWS VIDEOS

ABSTRACT

Next-generation news video consumption will be more personalized, device agnostic, and pooled from many different sources. The EigenNews system provides each viewer with a personalized newscast filled with stories that matter most to them. While personalized text-based news services already exist, personalized video news streams are still a critically missing technology. Our system records multiple channels, segments each news stream into individual stories, discovers links between the stories themselves and links between the stories and online articles, and generates a personalized playlist for each user. The dynamic mixing and aggregation of broadcast news videos from multiple sources greatly enriches the news watching experience by providing more comprehensive coverage and varying perspectives. Users can watch their playlists from the EigenNews website on the television, computer, tablet, or smartphone. Over time, the website collects click and view histories and uses this information to continuously optimize the personalization.

Index Terms— personalization, story segmentation, story linking, interactive videos, news aggregation

1 Introduction

News aggregation sites like Google News and Yahoo News have garnered significant attention in recent years, especially among younger consumers. One key reason for their popularity is the ability to personalize news. Users can customize what types of stories they want to read through a settings menu. Furthermore, these sites can progressively learn each user's preferences from their reading history to improve future selections.

EigenNews is a novel system that we have developed which personalizes broadcast news videos in the same spirit of personalize text-based news. Our system has three main advantages over traditional television news broadcasting:

- Personalization: We enable users to personalize and customize their individual playlists.
- Aggregation: We provide a broad selection by aggregating news videos from many different sources.
- Device-agnostic access: Our service can be viewed on televisions, computers, tablets, and smartphones.

To enable these advantages, the EigenNews system records news streams from multiple channels, segments the streams into individual stories, links the stories to each other and to online metadata, provides a story-level personalization that takes into account each user's viewing history and preferences, and streams the news videos through an interactive website that can be accessed on any device.

Fig. 1: Our EigenNews video studio. On the left is the server rack, which hosts the video recording equipment. The TV wall displays the different channels that are being recorded and analyzed. Living room is for conducting user studies on the benefits of dynamic mixing of news stories from multiple sources into a personalized playlist.

The rest of this paper is organized as follows. Sec. 2 highlights related work and identifies our novel contributions. Sec. 3 provides an overview of the recording architecture and EigenNews website. Then, Sec. 4 describes in detail our news video content analysis algorithms. Finally, the playlist personalization algorithm, which solves an elegant integer programming problem, is explained in Sec. 5.

2 Related Work

Following the success of personalized text news, personalized news video services have begun to appear. NoTube[1] has showcased personalized television news from Italian broadcaster RAI. Their system segments stories based on automatic speech recognition (ASR), finds named entities mentioned in the videos, and connects the named entities to online metadata. Watchup[2] is an iPad application that lets users choose stories from different online sources. Cronkite enables users explore more information about a story on a second screen

[1] http://notube.tv/showcases/personalised-news
[2] http://watchup.com

while watching the story on television [1].

The first problem in creating personalized news video playlists is generating accurately segmented stories. For story segmentation in broadcast news, many different audio-visual and text features have been studied [2, 3]. Hsu et al. [4] use a maximum entropy objective to select the most informative mid-level audio-visual features and also demonstrated an optimal fusion of the features. Further improvements using Support Vector Machines are shown in [5, 6]. Alternatively, Chaisorn et al. [7] employ a decision tree and Hidden Markov Model to locate the story boundaries. Many of the prior works focus solely on audio-visual features or solely on text.

Second, once the news videos are segmented, we want to match the videos to online news articles to generate rich metadata. Mori et al. [8] proposed a clustering-based approach to summarize and track events from a collection of news web pages. Lee et al. [9] extract and track important keywords in news articles. Hsu et. al. [10] use low-level visual features like color moments and Gabor textures to detect visual near-duplicates in news videos.

Lastly, generating personalized news video playlists has been considered in a few prior works [11, 12]. The problem is also closely related to personalized recommendation systems for online users, where two main technologies used are collaborative filtering [13] and content-based recommendation [14]. Recently, Liu et al. [15] showed how to model user preferences based on click histories to improve personalization.

Compared to the prior works, our novel contributions in the areas of story segmentation, story linking, and playlist personalization are summarized as follows:

- For story segmentation, we employ audio, video, and text features in a multi-modal segmentation algorithm.
- For story linking, we use visual and text analysis to accurately link stories. Despite viewpoint variations, transcoding artifacts, and station-specific clutter, robust image matching enables detection of similar footage generated by different news providers.
- For playlist personalization, we cast personalization as an integer programming problem that considers user preferences, viewing history, and story relationships in choosing an optimal set of stories for each user.

3 System Architecture

3.1 Recording and Processing Equipment

A video studio has been created to record the videos for the EigenNews system, as shown in Fig. 1. Our system currently records four channels from over-the-air (OTA) antennae (e.g. NBC, ABC, PBS) and 12 channels from cable boxes (e.g. CNN, HLTV, Bloomberg). The server rack contains three recording stations, where each recording station consists of a PC equipped with multiple TV tuners as depicted in Fig. 2. We use a combination of the Hauppaugge WinTV-HVR-950Q and WinTV-HVR-1950 for recording videos from the cable boxes and the HDHomeRun Dual for recording OTA programming. Each video is recorded as an MPEG-2 transport stream at a bitrate of 5 Mbit/s.

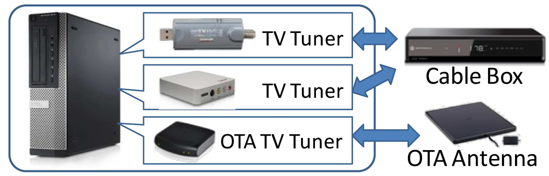

Fig. 2: A recording station consists of a PC and TV tuners that capture video from cable boxes and over-the-air antennae.

Sec. 4 will describe how we process the recorded videos to generate segmented news videos with rich metadata, which become the inputs to the playlist personalization algorithm discussed in Sec. 5.

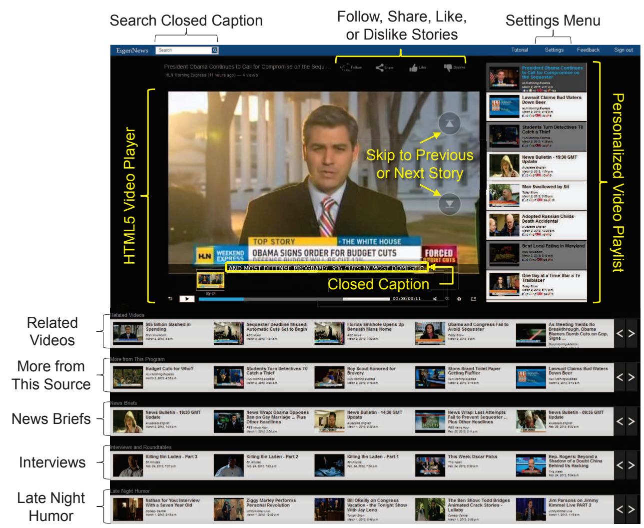

Fig. 3: EigenNews website for personalized news video.

First, a news video is segmented into individual stories. Then, each segmented story is linked to online articles as well as to other segmented stories. At the same time, the segmented videos are transcoded to appropriate bitrates for streaming through the EigenNews website. Since story linking and video transcoding are mutually independent, we perform these operations in parallel. The processing delays of 5.6 minutes for story segmentation and 8.6 minutes for story linking and transcoding have been measured.

3.2 User Interface

Fig. 3 shows a screenshot of the EigenNews website. When a user logs into the website, she/he will see a personalized playlist on the right, filled with stories that are intended to interest her/him. Story tiles in the playlist can also be easily reordered or removed. At the top, there is a search bar for querying the videos' closed captions and a settings menu for adjusting preferences toward different news programs and categories. In the player region, there are buttons for following stories related to the current story, sharing a link to the current story, skipping to the previous or next story and expressing like/dislike towards the current story. Beneath the player region, there are several menus for story exploration showing: (1) stories related to the current story from different news sources, (2) other stories from the same news source, (3) news briefs, (4) interviews, and (5) late night humor news clips.

All of the user's actions and viewing history are logged into a database. Our system analyzes the database to improve the personalized playlist generation, as explained in Sec. 5. The EigenNews website has been tested successfully in Chrome, Firefox, Safari, and Internet Explorer on different platforms and devices.

4 News Video Stream Analysis

For personalization, it is important to segment a recorded television news stream into individual stories and then predict each user's potential interest toward each story. In Sec. 4.1, our multi-modal story segmentation algorithm is presented. Then, we show in Sec. 4.2 how to match the segmented stories to each other and to online news articles.

4.1 Story Segmentation

Fig. 4 gives an overview of the proposed multi-modal segmentation algorithm. Due to different production rules across channels and the large diversity of news segments, multiple important cues in the news streams must be exploited to reliably segment the individual stories.

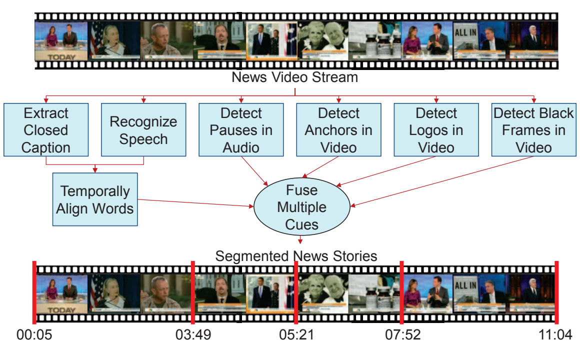

Fig. 4*:* Block diagram of the proposed news story segmentation algorithm.

4.1.1 Textual Cues

Some of the most important cues for story boundaries exist in the closed caption (CC). Often, >>> and >> markers are inserted to denote changes in stories or changes in speakers, respectively. Due to human errors, relying solely on these markers gives inaccurate segmentation results. Thus, we have additionally created a list of over 600 reliable transition phrases that occur at story boundaries, collected from several months of news programs across different channels. The porposed algorithm considers all occurrences of >>> markers and transition phrases as potential story boundaries.

There is a delay between the video and CC test that varies randomly even within the same newscast and can be as large as tens of seconds. To obtain accurate timestamps for the CC words, we perform automatic speech recognition (ASR) on the audio track using the CMU Sphinx Toolkit [16]. Words in CC and ASR are matched and aligned using a dynamic time warping algorithm [17].

4.1.2 Visual Cues

Visual boundaries provide transitions not discovered in CC alone and improve the accuracy of boundaries already found by CC. After experimenting on many visual cues, three very useful features were identified: (1) anchor frames, (2) logo frames, and (3) dark frames.

Typically, one or more anchorpersons appear between stories to provide a graceful transition. First, our system applies a Viola-Jones face detector [18] on all video frames. Then, a color histogram (CH) in HSV space is computed on an elliptical region on the face mainly containing facial interest points. After clustering these facial CHs over the entire newscast, dominant clusters are identified corresponding to the anchorperson(s) in the program. Finally, anchor shots are found from frames whose facial CH is close in L1 distance to the dominant cluster centroids, and the boundaries of the anchor shots become story boundaries.

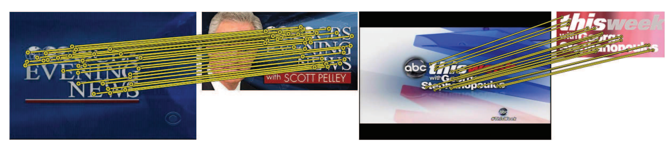

Fig. 5*:* Examples of matching video keyframes (left of each pair) to online logo images (right of each pair) using local visual features. Yellow lines denote post-RANSAC feature matches.

Many news programs insert the program logo between certain stories, although the logo appearance and position can vary unpredictably over time. We periodically crawl the news programs' websites for logo images. Then, the logos are automatically matched to the videos as follows: (1) Extract keyframes from the video at a rate of 1 keyframe/second. (2) Extract SURF[19] features from the logo images and video keyframes. (3) Compute a global signature [20] from the local features. (4) For each logo image, generated a ranked list of video keyframes by comparing the global signatures. (5) For the top 50 keyframes in the ranked list, perform pairwise comparison to the logo image with affine RANSAC [21]. A valid match is declared if the number of post-RANSAC feature matches exceeds $\tau_r = 20$. Fig. 5 shows two examples of online logo images that are successfully matched to video keyframes by this process.

Dark frames are frequently inserted at the boundaries of commercials and hence provide another valuable clue for segmentation. We convert all frames to grayscale, and we regard a frame as a dark frame if the mean is below μ_b and the standard deviation is below σ_b. Typical values are $\mu_b = 40$ and $\sigma_b = 10$ for graylevels in the range $[0, 255]$.

4.1.3 Auditory Cues

In addition to CC and visual cues, we detect significant pauses in the audio track. This audio cue captures the tendency of an anchorperson to pause momentarily or take a long breath before introducing a new story. A significant pause is defined as a duration longer than 0.3 seconds where the audio volume is below half the average volume of the news stream.

4.1.4 Segmentation Results

In EigenNews system, story segmentation is performed using multi-modal fusion of the introduced visual, auditory and textual cues. Table 1 shows the evaluation results of the proposed segmentation algorithm on TRECVID 2003 [22] dataset.

Table 1*:* Story segmentation Precision, Recall and F1 measures on TRECVID 2003 dataset

Dataset	Precision	Recall	F1
ABC News	55%	64%	59%
CNN Headline News	72%	79%	76%

4.2 Story Linking

In this section, we describe two methods for matching the segmented news videos to each other and to online articles, as depicted in Fig. 6. Matching segmented videos to each other enables discovery of related stories. Matching to online articles enables (1) extraction of informative titles for the segmented videos and (2) generation of hyperlinks to related online webpages which the user can read on the second screen.Three times every day, we crawl several major news websites (cnn.com, nbcnews.com, abcnews.com, nytimes.com) and save the URL, text, and images of all discovered articles. Sec. 4.2.1 describes lining with video keyframes. Then, Sec. 4.2.2 discusses a complementary approach of linking with text.

4.2.1 Visual Linking

Discovery of near-duplicate news images with low-level visual features such as color moments and Gabor textures has been considered in [10]. In contrast, we use local visual features for more robust matching in the presence of severe geometric and photometric distortions. Similar to Sec. 4.1.2, keyframes are sampled at 1 keyframe/second, local visual features are extracted, and global signatures are computed from the visual features. After the global signature comparisons produce a ranked list of keyframes, a shortlist of the top 50 keyframes is further verified geometrically with RANSAC, and valid matches must exceed a threshold of $\tau_r = 20$ post-RANSAC feature matches.

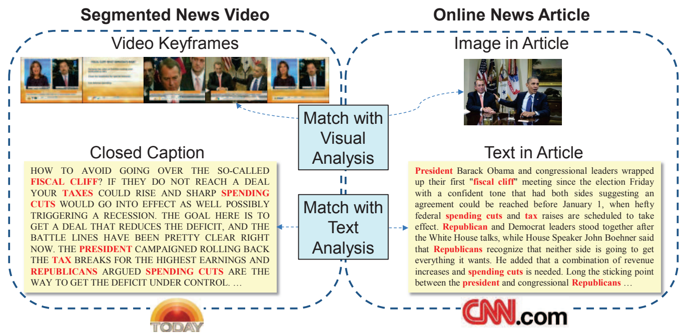

Fig. 6*:* Finding links between segmented news videos and online news articles, using visual analysis and text analysis.

4.2.2 Text-based Linking

As a complementary approach, we can match the CC of two stories or match the CC of a story to the text of an online article. First, bag-of-words histogram comparisons enable matching of text segments with similar distributions of words. Let $H_a(w)$ and $H_b(w)$ be the L1 normalized histograms of the words in the two documents, respectively. Then, using estimates of relative word frequencies $\{f(w)\}$, we compute a *tf-idf* histogram intersection score:

$$\begin{aligned} S\left(H_a, H_b\right) &= \sum_w \mathrm{idf}(w) \cdot \min\left(H_a(w), H_b(w)\right) \\ \mathrm{idf}(w) &= \log\left(\max_x f(x)\right) - \log\left(f(w)\right) \end{aligned}$$

If $S\left(H_a, H_b\right)$ exceeds a threshold $\tau_t = 2$, then a valid match is considered to be found.

Second, matching of named entities contained in the text segments provides additional links. We build a database of names using Wikipedia and Twitter and also employ the Stanford Named Entity Recognizer and Open Calais to identify all occurrences of names in the text segments. When two text segments share 5 or more names in common, a valid match is declared between the text segments.

5 Playlist Personalization

In the EigenNews system, personalization plays a central role in deciding which news stories appear in the playlist. In Sec. 5.1, we introduce an aggregation algorithm, which presents a limited number of stories, sorted into some preferred order. In Sec. 5.2, we demonstrate a method to generate a unique playlist by considering explicit preferences, personal information, and implicit preferences obtained by analyzing a user's history. Finally, in Sec. 5.3, we show that personalization is desirable for news video viewing and that EigenNews users preferred the personalized playlist compared to a traditional news broadcast or a randomly generated playlist.

5.1 Aggregation

The aggregation step selects and orders a limited number of news stories by considering both diversity and personaliza-

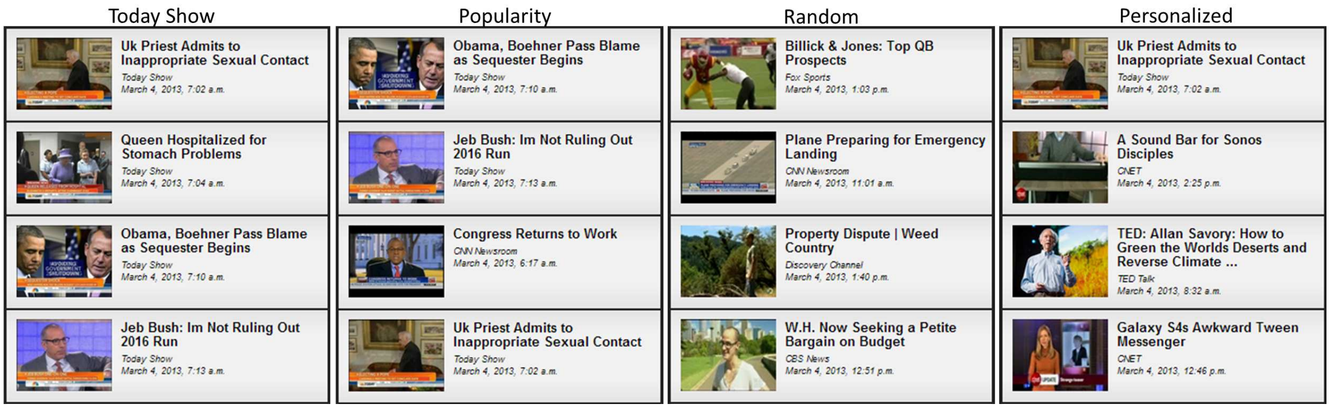

Fig. 7*:* Four different playlists. First, *Today Show* includes only stories from the latest episode of the Today Show. Second, *Popularity* in which news stories are sorted based on their number of related stories. Third, *Random* , where the news stories are chosen randomly from today's news videos. Fourth, *Personalized* playlist that presents a large selection of important stories from many sources the user enjoys.

tion. The following objectives are considered: (1) A good playlist provides a comprehensive view of today's news to the user. (2) A playlist should not contain duplicate or near-duplicate stories covered by many news sources. (3) The aggregation method must consider the user's time preferences. (4) The personalized playlist must accurately reflect implicit or explicit user preferences for certain programs or categories of news.

To achieve these goals, the problem is formalized as an integer linear programing optimization problem, or more specifically a maximum coverage problem [23], as follows:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} y_i && (1)\\ \text{subject to} \quad & Rx \geq y && (2)\\ & d^T x \leq t && (3) \end{aligned}$$

where n is the number of today's videos, For $i \in [1 \ldots n]$, $x_i \in \{0,1\}$ is 1 if i^{th} news story is selected and $y_i \in \{0,1\}$ is 1 if x_i is covered by a selected news story, which has been already selected. $R \in \{0,1\}^{n \times n}$ denotes an adjacency matrix, where 1 represents a link between news stories as mentioned in Sec. 4.2. Duration of the news story and time limitation are represented by d_i and t accordingly. The objective function maximizes the coverage of today's news. Solving this integer linear programing problem is sufficiently fast: 1 second computation for a pool of 30 stories.

5.2 Personalization Score

In addition to the weighted ($w_{coverage}$) news story coverage, personalization needs to consider user preferences. In the proposed framework, a personalized score is added to the objective function as presented below. This term aims to maximize the user's preference-based satisfaction score. Each news video is weighted by a personalized factor $c_i \in R, i \in [1 \ldots n]$.

$$\text{maximize} \quad w_{coverage} \sum_{i=1}^{n} y_i + c^T x \tag{4}$$

The constraints remain the same as in Eq. 2 and 3. Three different properties were considered in calculating the personalization score c. First, a user's preferences (s_{source}, $s_{category}$), described in Sec. 3.2, are explicitly set in the settings page by specifying weights for different video news sources, i.e, ABC, CNN, and news categories, i.e., sports, politics. Second, more recent news videos are preferred, therefore the broadcast time of each news video is considered in the personalization score by s_{time}. Third, the user's viewing history is an implicit way of expressing preferences. The $s_{history}$ represents the number of related news stories, which were watched previously by the user.

$$\begin{aligned} s_{time}(v) &= time_v - time_{current} \\ s_{history}(v) &= \sum_{w \in Videos} related(v, w) \end{aligned}$$

where, $v \in [1 \ldots n]$, and $Videos$ is a set containing indices of all videos. The function $related(v, w) \in \{0,1\}$ is 1 if videos v and w are linked. Stories are linked using the methods described in Sec. 4.2.

The final personalization score, c, is calculated as the weighted sum of the following scores:

$$\begin{aligned} c = \; & w_{source}.s_{source} + w_{category}.s_{category} \\ & + w_{time}.s_{time} + w_{history}.s_{history} \end{aligned} \tag{5}$$

The weights are selected arbitrarily considering preliminary tests.

An example comparing different playlists is shown in Fig. 7. First, on the left, *Today Show* refers to a playlist filled only with stories from an episode of the program. This playlist was pre-determined by one channel, so it lacks an interesting variety of stories from different channels. Second, *Popularity* news stories are shown, which represent the most important stories in a day by calculating the number of similar stories. Third, *Random* playlist is presented in which news stories are selected arbitrarily. Fourth, *Personalized* playlist is shown, which presents a rich selection of stories across channels, including a couple of tech talks, which is desirable to the selected user.

5.3 Subjective evaluation

Objective evaluation of generated playlist is rather difficult and inaccurate compared to a subjective evaluation. Therefore a rating test is designed to determine significant preferences between four generated playlists. The subjective test

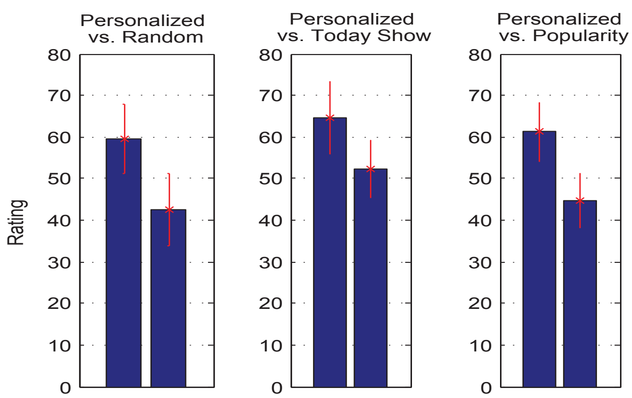

Fig. 8*:* Subjective evaluation. Comparing user ratings for random, Today Show and popularity playlists with the personalized playlist

contains six pairwise comparisons between personalized, randomly merged, news program and popularity playlists. Two playlists are shown in each comparison and the user rates both playlists on a linear scale marked by excellent, good, fair, poor and bad scores. The result of 21 users' ratings are compared and shown in Fig. 8. P-value is calculated for each subjective tests using a one-tailed paired two-sample Student's t-test. The calculated p-values for the personalized playlist compared to randomly merged, news program, and popularity playlists are 1.1%,1.3% and 0.2% respectively. The values are less than 5%, which shows that the personalized playlist is preferable to the other playlists with statistical significance.

6 Conclusion

The EigenNews system is a personalized news video system that is capable of generating personalized playlist of everyday's news for each user. To achieve precise story-level personalization, our system automatically records multiple channels, segments each news stream into individual stories, and discovers links among the stories and online metadata. Given segmented, linked, and annotated news stories, the personalization process can be elegantly cast as an integer programming problem, which has an efficient solution. As users watch more videos on the EigenNews website, their clicks and view histories help us further improve the personalization quality for subsequent sessions, and the set of all users' growing histories becomes a valuable resource for continued news video personalization research.

7 References

[1] K. Livingston, M. Dredze, K. J. Hammond, and L. Birnbaum, "Beyond broadcast," in *International Conference on Intelligent User Interfaces*, 2003, pp. 260–262.

[2] X. Gao and X. Tang, "Unsupervised and model-free news video segmentation," in *IEEE Workshop on Content-based Access of Image and Video Libraries*, 2001, pp. 58–64.

[3] Y. Zhai, A. Yilmaz, and M. Shah, "Story segmentation in news videos using visual and textual cues," in *ACM International Conference on Multimedia*, 2005, pp. 92–102.

[4] W. Hsu and S.-F. Chang, "A statistical framework for fusing mid-level perceptual features in news story segmentation," in *International Conference on Multimedia and Expo*, 2003, pp. 413–416.

[5] W. Hsu, *An Information-Theoretic Framework towards Large-Scale Video Structuring, Threading, and Retrieval*, Ph.D. thesis, Graduate School of Arts and Sciences, Columbia University, 2007.

[6] W. Hsu, S.-F. Chang, C.-W. Huang, L. Kennedy, C.-Y. Lin, and G. Iyengar, "Discovery and fusion of salient multi-modal features towards news story segmentation," in *SPIE Electronic Imaging*, 2004, pp. 244–258.

[7] T.-S. Chua, S.-F. Chang, L. Chaisorn, and W. Hsu, "Story boundary detection in large broadcast news video archives: techniques, experience and trends," in *ACM International Conference on Multimedia*, 2004, pp. 656–659.

[8] M. Mori, T. Miura, , and I. Shioya, "Topic detection and tracking for news web pages," in *IEEE/WIC/ACM International Conference on Web Intelligence*, 2006, pp. 338–342.

[9] S. Lee and H.-J. Kim, "News keyword extraction for topic tracking," in *International Conference on Networked Computing and Advanced Information Management*, 2008, pp. 554–559.

[10] W. H. Hsu and S.-F. Chang, "Topic tracking across broadcast news videos with visual duplicates and semantic concepts," in *International Conference on Image Processing*, 2006, pp. 141–144.

[11] B. Merialdo, K. T. Lee, D. Luparello, and J. Roudaire, "Automatic construction of personalized tv news programs," in *ACM International Conference on Multimedia*, 1999, pp. 323–331.

[12] H. Luo, J. Fan, and D. A. Keim, "Personalized news video recommendation," in *ACM International Conference on Multimedia*, 2008, pp. 1001–1002.

[13] A. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: scalable online collaborative filtering," in *International Conference on World Wide Web*, 2007, pp. 271–280.

[14] D. Billsus and M. J. Pazzani, "A hybrid user model for news story classification," in *International Conference on User Modeling*, 1999, pp. 99–108.

[15] J. Liu, P. Dolan, and L. R. Pedersen, "Personalized news recommendation based on click behavior," in *International Conference on Intelligent User Interfaces*, 2010, pp. 31–40.

[16] W. Walker, P. Lamere, P. P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel, "Sphinx-4: a flexible open source framework for speech recognition," Tech. Rep., 2004.

[17] A. G. Hauptmann and M. J. Witbrock, "Story segmentation and detection of commercials in broadcast news video," in *Advances in Digital Libraries Conference*, 1998, pp. 168–179.

[18] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[19] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[20] D. Chen, S. Tsai, V. Chandrasekhar, G. Takacs, R. Vedantham, R. Grzeszczuk, and B. Girod, "Residual enhanced visual vector as a compact signature for mobile visual search," *Signal Processing*, 2012.

[21] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[22] "Trecvid: Trec video retrieval evaluation, http://www-nlpir.nist.gov/projects/trecvid," .

[23] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions-I," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, Dec. 1978.