Integration of Hurricane Wind Analysis and Multimedia Semantic Content Analysis for Public Outreach

Guy Ravitz, Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL 33124, USA
{*ravitz,shyu*}*@miami.edu*

Mark D. Powell
NOAA/AOML
Hurricane Research Division
4301 Rickenbacker Cswy
Miami, FL 33149, USA
Mark.Powell@noaa.gov

Abstract

*Natural disasters, such as hurricanes, could have an enormous impact on society. The level of the public's preparedness could make a significant difference in the severity of casualty and damage inflicted by such storms. We present a prototype system to reach out to the public and improve their awareness of the potential dangers involved with such weather events. This web-based system aggregates H*Wind storm track and wind fields data along with relevant videos extracted from YouTube and displays it to the user using Google Earth. A content-based concept detection algorithm is used to extract the videos, which may describe the impact of the storm in relevant geographic locations. Using Hurricane Ike as a case study, the result demonstrates how some of the information collected and displayed by the system could have increased the awareness of the public and potentially helped prepare them better to the devastating storm.*

Keywords: Concept Detection, Web 2.0, Severe Weather Warning System, Public Outreach

1. Introduction

A plethora of sources, such as the newspaper, television, radio, and Internet websites, which provide severe weather analysis and warnings, are available for the public nowadays. However, property damage and loss of life due to poor preparedness during hurricanes, amongst other, is still very large. According to [2], the 2005 atlantic hurricane season is directly responsible for about 1731 deaths and over $\$100$ billion in property damage in the United States alone. Hurricane Katrina, of that year, was blamed for about 1500 direct deaths and about $\$81$ billion in property damage. According to [21], the 2008 Atlantic hurricane season was blamed for about 789 deaths and over $\$41$ billion in property damage. Hurricane Ike was held responsible for about $\$28$ billion in property damage and 103 deaths. These numbers show us that there is still a big gap between the weather forecasters, warning systems currently at place, and the general public. In [17], we showed that the Saffir-Simpson (SS) scale, which is currently used to communicate the disaster potential of hurricanes in the Western Hemisphere, is not successful in communicating the destructive potential of a hurricane, and introduced the kinetic energy based wind and surge destructive potential (SDP) scale to improve the public's risk perception. For example, the SS scale indicated 'only' a category 2 storm for Hurricane Ike, while the SDP scale rating was much higher.

Web 2.0 [5, 14] and the success of social networks such as FaceBook and MySpace have demonstrated the incredible influence people may have on their peers and colleges. Many personal experiences are shared today on YouTube [24] via home and professional videos, and we believe that using this great data source may help our goal of increasing the impact of severe weather warning on the society. Furthermore, the content on YouTube is user-generated and therefore, videos found on YouTube may be less politically influenced than national or local news broadcasts. We believe that videos showing the high waves and surge which were available hours prior to Hurricane Ike's landfall could have helped increase the public's risk perception and potentially led to more evacuations.

This paper is organized as follows. In Section 2, we discuss some relevant existing work. In Section 3, the proposed framework and the prototype system that we have developed are presented. Some preliminary results and analysis of our prototype system are provided in Section 4. Finally, the conclusion and future work are given in Section 5.

2. YouTube and Multimedia Content Analysis

In 2007, Boll [3] suggested that multimedia research and Web 2.0 have a lot in common and that currently Web 2.0 is not benefiting enough from multimedia research. This article is based on [5], where panelists from Google, Yahoo! Research, and University of California at Berkeley's School of Information met and discussed this topic. Some of the potential benefits mentioned in this article were: a simple shot segmentation based video bookmark generation, speaker recognition, shot comparison for finding similar or identical videos in the database, a content based grouping, video content classification, and semantic concept identification combining content and context.

The author also added that the content of multimedia in Web 2.0 applications such as YouTube and Flicker is different in structure and quality from the clean room data that is currently used by multimedia researches, and called multimedia researchers to use content that meets the everyday users in their research [3]. For example, in [9], Kim et al. presented a face tracking and recognition system, which was tested on both known benchmark data and real world YouTube data. The authors reported that the recognition performance on the benchmark data sets was about 100%, while the performance using the YouTube videos was much lower (around 70%). Indeed, in recent years, the Web 2.0 research area enjoyed increasing content based analysis research activities.

A large amount of attention has been dedicated to analyzing structures and usage of Web 2.0 applications [4, 6, 15, 19]. In [4], the authors performed the analysis of video popularity characteristics of two large Web 2.0 systems, namely YouTube and Daum Videos (a similar system to YouTube in Korea), to illustrate how different such user generated content (UGC) systems are from other systems such as video on demand (VOD). This report highlighted some potential issues of such applications, such as poor search and recommendation engines, content aliasing, and illegal uploads.

Recent research of multimedia content analysis integration into Web 2.0 systems concentrated in two main areas, namely duplicate and/or relevant video detection [8, 13, 23] and automatic generation of uniform metadata [1, 20]. As identified in [4], one major issue with UGC systems is the fact that many videos in the database are similar if not identical. In [23], Wu et al. presented a framework which combined contextual information, such as the duration and the number of views, with the content based information, such as color and local point, to achieve real-time near-duplicate elimination. The authors showed that based on 24 popular queries of databases like YouTube, Google Video, and Yahoo Video, there were 27% redundant videos on average that were a duplicate or near duplicate of the most popular version of the videos in the search results.

The existence of reliable metadata makes tasks such as retrieval and classification of videos or images much easier with much lower complexity. However, according to Begeja and Van Vleck [1], the metadata which accompanies self-described videos, such as the ones in YouTube, are unreliable. In their work, the authors proposed a framework which automatically generated reliable metadata that could then be used for contextual advertising in IP Television (IPTV) environments. A processing unit called the Analyzer would receive the content (videos) and would run several algorithms to generate the metadata for the content like speech recognition, face detection, major keywords, etc. These metadata are then sent out as a separate stream.

The popularity of Web 2.0 and the great challenges of using UGC data make it important to ensure that content analysis frameworks perform well with UGC data by either modifying existing work or generating new more robust frameworks. This integration has become a major interest of the TRECVID [18] society as well. In recent years, this society has encouraged the participants to use UGC data from sources such as Flicker and YouTube.

In our previous work [10, 11, 12], we have proposed several concept detection frameworks whose goal has been to find semantic meanings of different video shots based on the content of the audio and video (low-level features). In this paper, we present a framework that is designed to detect several concepts (high-level semantics) from UGC data extracted from YouTube based on the information extracted from H*Wind [16] and uses that to integrate videos and H*Wind output into a Web-based framework which uses the Google Earth API. The potential usability of such an integration is demonstrated via a prototype of the proposed system.

3. The Proposed Framework

As we look to increase public severe weather awareness and preparedness, the fact that Web 2.0 applications seem to be closer to the users and are better at involving and integrating users so that they participate more, [3] makes Web 2.0 a perfect candidate to achieve this goal and to bridge the current gap between weather forecasting and the general public.

In this paper, a framework is proposed that integrates video data collected from YouTube and analyzes the data collected from H*Wind to increase public awareness during the times of severe weather event threats. The structure of the proposed framework can be seen in Figure 1. The *System User* interacts with a Web-based *User Interface*. The core of the interface is driven by the *Google Earth API*. The *User Interface* is implemented using technologies such as HTML, JavaScript, and AJAX (Asynchronous

JavaScript and XML). The use of the *Google Earth API* relieves the load of the *Http Server* as the communication with the API is done via client side scripting such as JavaScript. The *Http Server* acts as a mediator among the *User Interface*, *H*Wind*, and *YouTube*. Based on the requests from the client, the *Http Server* collects the requested data from *YouTube* and *H*Wind*, processes it, and sends it back to the user interface for display. We use Java Servlet technology to collect the needed information from *H*Wind* and *YouTube* separately, perform necessary quality control, encapsulate it to a keyhole markup language (KML) message, and send it back to the requesting *User Interface* instance. The KML messages are then communicated to the *Google Earth API* which in return tells the *User Interface* how to display the integrated data from *H*Wind* and *YouTube* to the user.

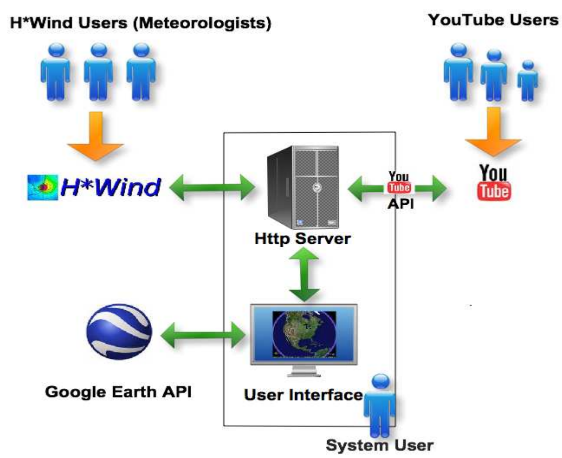

Figure 1. System architecture global view

3.1. HRD Surface Wind Analysis (H*Wind)

The Hurricane Research Division (HRD) at the Atlantic Oceanographic and Meteorological Laboratory (AOML) at Miami, Florida has developed a tool called H*Wind [7, 16] as a part of their surface wind analysis program to provide real-time objective analysis of wind speed distribution in a tropical cyclone (TC). H*Wind enables the scientists and meteorologists to view wind speed observations collected from reconnaissance aircrafts, satellites, and marine and land observation platforms, analyze it, and better understand the extent and strength of the wind-field and hurricane intensity.

H*Wind generates snapshot images which visualize the wind-field of the storm at specific location and time. We use these images in our proposed system to help the user understand the current conditions at areas affected by the storm. Such information could help the user understand the awaiting dangers and prepare better for the storm. However, in order to make such predictions and assessments, some scientific and meteorological knowledge is needed. Therefore, to help the users understand the situation better, we enhance this information by adding relevant user-generated content taken from YouTube.

3.2. YouTube

YouTube is a freely available Web 2.0 application which enables the users to upload videos they have created and share them. To easily integrate YouTube to any website, Google has made the YouTube API [25] available freely as well. The components of the API which are relevant to our work are the Data and Player APIs. The Data API is used to perform searches and retrieve videos from the YouTube database; while the player API is used to playback and control the retrieved YouTube videos. The amount of video available on YouTube is incredibly large. In order to use only the relevant videos and not overwhelm both the system and the users, we have developed a concept detection mechanism which will select and rank videos for the users' viewing.

3.3. Concept Detection System

Currently, the main two browsing features in video retrieval systems such as YouTube are search and recommendations. If a user is interested in watching videos related to "Hurricane Ike" for example, he/she will use these words as the search keywords. Once a user selects a video to view, the engine generates a list of recommended videos based on different characteristics like similar tags, similar keywords, etc. As mentioned in Section 2, the metadata related to videos in YouTube, including the keywords and tags, are generated by the users and therefore are not very reliable. For this reason, we have developed a content based concept detection component.

3.3.1 Investigated Concepts

First, we have to identify the desired concepts (semantics) we want to present to the users. In the past [10], we have developed a system to detect weather related shots from news broadcast videos. Modeling the weather shots was a difficult task due to the variability of the audio-visual content. An empirical study of videos retrieved from YouTube helped us identify three concepts that may provide us the ability of discriminating the relevant from the non-relevant videos for our proposed system. These concepts are as follows:

- Windy - videos that the existence of wind can be detected in parts or all of its audio track. We are mostly interested in videos describing the outside weather conditions during a storm;

- Music - many YouTube users add music to their videos. Added music may indicate that the user had time to structure the videos and therefore, the video may not be very recent. We are mainly interested in un-structured raw videos that will provide information considered as up-to-date as possible and will impact the viewer properly;
- Speech - many videos in YouTube are of documentary or self portrait form. In both cases, they represent videos that might be more relevant than the ones with music but less relevant than the windy videos. These videos also include news and news-like broadcasts.

We have chosen to model the above concepts using the features extracted from the audio track. This was decided mainly due to two reasons. First, audio processing is not as computationally expensive as video processing and would allow us to provide fairly close to real-time performance. Second, the visual characteristics of the investigated concepts in different YouTube videos we have observed so far are very different, which makes the discrimination process extremely difficult, if not impossible, in some cases.

3.3.2 Feature Extraction

In this paper, we have explored the frequency characteristics of our data. The Zero Crossing Rate (ZCR) is known to be a good frequency estimator and speech indicator. ZCR refers to the rate in which an audio signal crosses the zero line. According to [26], when the short time zero crossing rate is high, the speech signal is unvoiced, and when it is low, the speech signal is voiced. Therefore, the variance of the ZCR of a speech signal will be very high, as opposed to the ZCR variance of music signal which is very low. We chose to use ZCR in our work to help our classifier discriminate among clips that contain speech, 'dirty' speech (for example people talking while filming a windy area outside), and music. It can be seen from Figure 2 that the ZCR variance of a speech segment is much higher than the ones of windy and music segments.

We also investigated the power spectral density (PSD) of the investigated audio segments. The PSD provides information regarding how the power of a signal is distributed with frequencies. In this paper, the PSD for each of the investigated segments is calculated and then the average power of the following frequency sub bands is also calculated, (i) 0 - 500 Hz, (ii) 501 - 5000 Hz, (iii) 5001 - 10000 Hz, (iv) 10001 - 15000 Hz, and (V) 15000 - 20000 Hz. This was done in order to try and further characterize the frequency distribution of the investigated concepts.

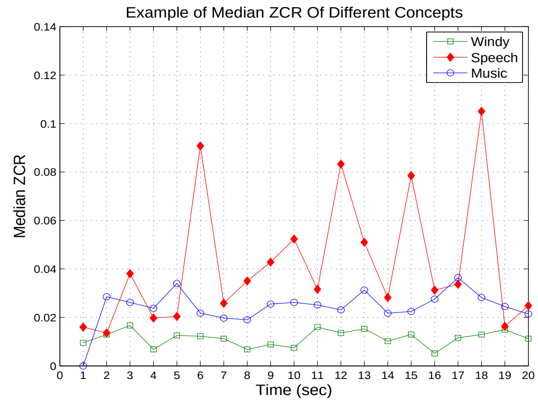

Figure 2. Median ZCR values of 3 segments representing the 3 investigated concepts

3.3.3 Concept Detection

In this paper, we have extracted 10 audio features. Four of them are related to ZCR, including median, average, dynamic range, and variance, and six of them are related to the average power of the aforementioned spectral bands. These features were computed per each 33 milliseconds of audio and then averaged per each one second interval. Under the assumption that the users would most probably not watch each YouTube in its entirety, we have chosen to extract the features for the first 2 minutes of each video and used that as a representation for the entire video. Videos that were shorter than two minutes were processed entirely. Next, the values for each feature were averaged for each video and these values were used to represent the entire video. Finally, we have used the well known decision tree classifier to assign classification labels to the investigated videos.

3.4. System Prototype

We developed an initial prototype system to demonstrate the expected performance of the proposed framework. This Web-based prototype system demonstrates the integration capability of our system via a simple User Interface. This user interface was developed using the HTML and JavaScript Technologies. This interface integrates data retrieved from YouTube and H*Wind by displaying it over the web version of Google Earth using the Google Earth API. An example of the displayed integrated data can be seen in Figure 3, where a snapshot of H*Wind analysis representing Hurricane Ike's wind intensities on September 13, 2008 at 1:30 am UTC is displayed to the user. At the same time, a YouTube icon is placed on the map signifying that a relevant video is available at that time. Once the user clicks on the YouTube icon, a balloon opens with the video in it for the user to view. The video in the balloon would be the video that the concept detection system ranked as the most relevant. If the user wants to view more videos available at that time in our system, he/she will click the provided link which will take them to a page with all the currently available videos ranked based on the concept detection system. In this specific example, it can be seen that as Hurricane Ike was approaching Galveston Island, some YouTube videos showing the storm surge created by Ike at Galveston Island was already available for viewing.

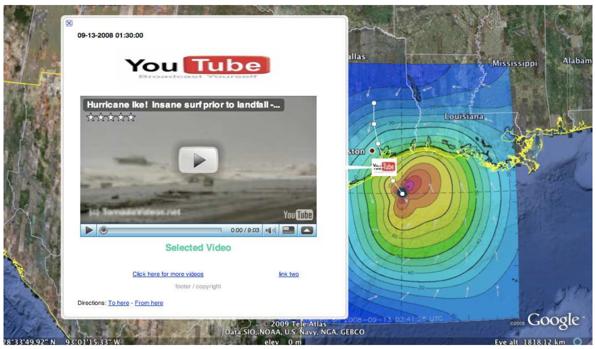

Figure 3. System Prototype

4. Experiments and Results

An experiment is designed to examine the concept detection capability of the proposed framework to provide a preliminary system evaluation. We have queried the YouTube database using the keywords “Hurricane Ike” and collected the videos returned in the first 2 result pages. Videos that had no audio information were discarded, resulting in a data set of the top 34 returned videos. Each video was manually labeled to belong to one of the 3 classes based on the presence of the investigated concepts in each video. The audio data was extracted from the videos to a mono signal sampled at 44.1 KHz and 16 bits. We kept the full sample rate (as opposed to the down-sampling we performed in the past) in order to be able to investigate a full frequency range. After extracting the 10 features, each example (video clip) had one value per each feature.

Table 1. Average precision (Pre), recall (Rec) and F1-score (F1) for the investigated concepts using WEKA unpruned DT

	Windy	Speech	Music
Pre	0.833	0.833	0.22
Rec	0.833	0.714	0.25
F1	0.833	0.769	0.235

For classification, the decision tree (DT) implementation in WEKA [22] is used. In this experiment, we have used the unpruned version of DT and performed a 3-fold cross validation using the WEKA built-in cross validation mechanism. The experiment is executed 3 times, once per each concept. The precision (truly positive examples among all examples that were classified as positive), recall (examples that are classified as positive among all examples which are truly positive), and F-measure results were documented in Table 1. As shown in Table 1, the detection performance for windy and speech videos is good. In both cases, the ZCR variance feature plays a major role in the model. In the case of the speech concept detector, the average power of the entire spectrum plays a discriminating role as well. The poor performance of the music detector has two main explanations. First, we had only 8 positive examples for this concept versus 26 negative ones, which makes it difficult on DT to model this concept. In the case of the windy and speech videos, the counts were 12 positive versus 22 negative and 14 positive versus 20 negative examples, respectively. In addition, the extracted features have a difficult time discriminating between windy and music videos. There were 5 windy videos that were classified by both the windy and music video detectors. Some of the music videos that were misclassified by the other two detectors were of different genres and included different instruments that may have similar frequency characteristics to those of the other two concepts. The addition of harmony consideration may improve the music detector and the performance of the overall system.

Despite the poor performance of the music detector, the overall system may still have performed to the user’s satisfaction. We were able to generate clusters of videos based on the following assigned class labels: (i) windy only, (ii) windy + music, (iii) speech only, (iv) music only, (v) speech + music, and (vi) no class. If we would have sorted the resulting videos to the user based on these clusters, the results could be considered satisfactory. For example, out of the 7 videos that were not classified, only 2 videos would have been considered as not belonging to this group, and the rest 5 videos would be properly ranked low. Also, if the user is interested in re-ordering the group/clusters ranking, it could be easily achieved.

5. Conclusion and Future Work

In this paper, a semantic integration system that integrates the hurricane wind analysis and multimedia content analysis is proposed for the purpose of public outreach. The preliminary performance evaluation on the concept detection component of the proposed system shows that it achieves promising precision and recall values. A prototype is developed to demonstrate the capabilities of the system by showing its potential to increase public awareness and outreach using multimedia content analysis and Web 2.0, and the impact it may have on our society. The importance of the integration of multimedia content analysis and hurricane wind analysis utilizing Web 2.0 is also presented.

Future work will include addressing some of the issues identified in this paper, such as the low performance of the music detector and the low ability to discriminate between windy and music videos. In addition, the use of associative based learning and classification will be integrated into the system as a continuous effort to improve the concept detection component of the proposed system.

6. Acknowledgment

For Mei-Ling Shyu and Guy Ravitz, this research was carried out in part under the auspices of the Cooperative Institute for Marine and Atmospheric Studies (CIMAS), a Joint Institute of the University of Miami and the National Oceanic and Atmospheric Administration, cooperative agreement #NA17RJ1226. “The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the funding agency.”

References

[1] L. Begeja and P. Van Vleck. Contextual advertising for iptv using automated metadata generation. In *6th IEEE Consumer Communications and Networking Conference*, pages 1–5, January 2009.

[2] J. L. Beven, L. A. Avila, E. S. Blake, D. P. Brown, J. L. Franklin, R. D. Knabb, R. J. Pasch, J. R. Rhome, and S. R. Stewart. Annual summary atlantic hurricane season of 2005. *AMS Monthly Weather Review*, 136:1109–1173, March 2008.

[3] S. Boll. Multitube–where web 2.0 and multimedia could meet. *IEEE Multimedia*, 14(1):9–13, January-March 2007.

[4] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. Analyzing the video popularity characteristics of large-scale user generated content systems. *Accepted for future publication, IEEE/ACM Transaction on Networking*, March 2009.

[5] E. Chang, M. Davis, P. Schmitz, and S. Boll. Panel discussion: Web 2.0 and multimedia: Challenges, hype, synergy. In *Proceedings of ACM Multimedia*, page 752, March 2006.

[6] X. Cheng, C. Dale, and J. Liu. Statistics and social network of youtube videos. In *16th International Workshop on Quality of Service*, pages 229–238, June 2008.

[7] H*Wind. *http://www.aoml.noaa.gov/hrd/data_sub/wind.html*.

[8] A. Karpenko and P. Aarabi. Tiny videos: Non-parametric content-based video retrieval and recognition. In *Tenth IEEE International Symposium on Multimedia*, pages 619–624, December 2008.

[9] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley. Face tracking and recognition with visual constraints in real-world videos. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.

[10] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen. Video semantic concept discovery using multimodal-based association classification. In *IEEE International Conference on Multimedia and Expo*, pages 859–862, Beijing, China, July 2-5, 2007.

[11] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen. Correlation-based video semantic concept detection using multiple correspondence analysis. In *IEEE International Symposium on Multimedia (ISM)*, Berkeley, CA, USA, December 15-17, 2008.

[12] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen. Effective feature space reduction with imbalanced data for semantic concept detection. In *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, pages 262–269, Taiwan, R.O.C., June 11-13, 2008.

[13] S. Liu, M. Zhu, and Q. Zheng. Mining similarities for clustering web video clips. In *International Conference on Computer Science and Software Engineering*, volume 4, pages 759–762, December 2008.

[14] S. Murugesan. Understanding web 2.0. *IT Professional*, 9(4):34–41, July-Aug. 2007.

[15] J. Paolillo. Structure and network in the youtube core. In *41st Annual Hawaii International Conference on System Sciences*, pages 156–156, January 2008.

[16] M. D. Powell, S. H. Houston, L. R. Amat, and N. Morisseau-Leroy. The hrd real-time hurricane wind analysis system. *Journal of Wind engineering and Industrial Aerodynamics*, 77(78):53–64, 1998.

[17] M. D. Powell and T. A. Reinhold. ropical cyclone destructive potential by integrated kinetic energy. *Bull. Amer. Meteo. Soc.*, 88(4):513–526, 2007.

[18] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press.

[19] M. Soraya, M. Zamani, and A. Abhari. Modeling of multimedia files on the web 2.0. In *Canadian Conference on Electrical and Computer Engineering*, pages 1387–1392, May 2008.

[20] S. Venkatesh, B. Adams, D. Phung, C. Dorai, R. Farrell, L. Agnihotri, and N. Dimitrova. you tube and i findpersonalizing multimedia content access. *Proceedings of the IEEE*, 96(4):697–711, April 2008.

[21] Wikipedia. *http://en.wikipedia.org/wiki/2008_Atlantic_hurricane_season*.

[22] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[23] X. Wu, C.-W. Ngo, A. Hauptmann, and H.-K. Tan. Real-time near-duplicate elimination for web video search with content and context. *IEEE Transactions on Multimedia*, 11(2):196–207, February 2009.

[24] YouTube. *http://www.youtube.com/*.

[25] YouTube-API. *http://code.google.com/apis/youtube/*.

[26] T. Zhang and C.-C. Jay Kuo. *Content-Based Audio Classification And Retrieval For Audiovisual Data Parsing*. Kluwer academic Publishers, 2001.