Thesis Design

Master Thesis Information Studies

Jorick van Hees

Blue Billywig[*]
jorick.vanhees@student.uva.nl
UvA Student Nr.: 10894020
VU Student Nr.: 2567527

Categories and Subject Descriptors

Computing methodologies [**Computer vision**]: Video summarisation

1. INTRODUCTION

With the introduction of high speed internet and the introduction of video sharing platforms, the use of online video increased dramatically and is one of the main media content formats used on the web today, with increasing popularity in news media, business matters, education and the social web. According to Cisco [1], 64 percent of the consumer internet traffic in 2014 was online video.

News media website navigation often exists of article pages and overview pages like the front page, designed to allow easy navigation to the article pages. The overview pages link to the article using a title, a small description and a visual preview. The visual preview is a small visual representation, or *thumbnail*, of the content, that is only a few centimeters in size. The thumbnail should engage the user to follow the link to view the full article and is often the only visual reference.

Thumbnails have always been a strong visual cue for the user to find relevant content. A study from Microsoft, Dziadosz, and Chandrasekar [2] in 2002 regarding thumbnails in their search engine results, showed that users made more accurate decisions when a thumbnail of the result was presented next to the textual result. In general, images tend to pull the audience to the content; it's what newspapers are doing for years with the big frontpage images. The contents of these thumbnails exist of stock photos, actual photos from the event, or other designed images containing company logos, previews of graphs and other relevant visuals. Overview pages linking to multiple news articles often use these thumbnails to improve the visual appearance of the website, but also try to tempt the user into viewing the article. These images have always been static, but with the increase of broadband internet speeds and reliability (even on mobile devices), opportunities arise to create moving thumbnails. These thumbnails may improve the user experience and could increase the chances of users watching the full article.
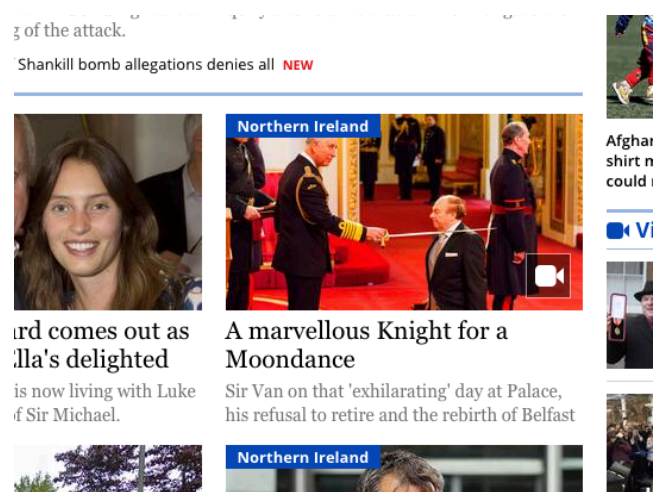

Figure 1: Screencapture of the Belfast Telegraph frontpage (http://www.belfasttelegraph.co.uk/) on the 5th of february, showing a thumbnail.

2. THE MOVING THUMBNAIL

With the rise of video as a way of communication and distributing news content, new opportunities arise to attract users to specific content. Instead of using a static frame from the video, it is possible to extract a small portion of the video and use that as a *moving thumbnail*. The moving thumbnail is a small video in both dimensions and length, that acts as a substitution for a regular thumbnail. In every way, the moving thumbnail is the same as a regular static thumbnail with the exception that it contains 'moving images'.

Static thumbnails have different purposes in news media websites in relationship to the content: it can try to accurately describe the content, or it can attempt to seduce a user to view the full content (which is what we can call a teaser). The chosen role of the thumbnail depends heavily on the type of content, the strategic goals of the media company and the location of where the thumbnail is used.

Since the moving thumbnail replaces the static thumbnail, it should serve the same purpose.

To pick the best (static) thumbnail for a desired goal, a number of systems exist to aid an editor into selecting a thumbnail. The most simple system features a time selector and a preview window in which the editor is able to select a frame on a certain time to use as thumbnail. Other systems use different video analysation techniques to propose a set of suitable options to the editor. In order to use a moving thumbnail as a replacement for the static thumbnail, the creation of these moving thumbnails should be fast and convenient.

2.1 Automatic generation of moving thumbnails

Creating moving thumbnails manually for every video can be a time consuming task, even when presented with various editing tools that allow quick trimming. In order to solve this problem, automatic generation of moving thumbnails is essential for the use and implementation of this new concept.

3. LITERATURE STUDY

3.1 Static thumbnail generation

3.2 Video summarisation

It is important to note that thumbnails are not the main content of a page. They are used to assist the user in navigating a page and scanning content, and should be available to the user at all times, wether he just landed on the page, or is still viewing the page after a few minutes.

The moving thumbnail has a start- and an endpoint. moving thumbnail, the video will loop infinitely to make sure the user will always see a moving thumbnail. In addition, the video will automatically start playing when the webpage is loaded.

Online content management systems (or video management systems) offer functionality to specify what frame is extracted from the video, or allow the content editor to upload an entirely new image to use as a thumbnail. This allows the editor to select a thumbnail that represents the video in the best way, or may attract most viewers.

Unlike static image thumbnails, the moving thumbnail videos have a set length and will stop when the end is reached. In order to create a continues moving thumbnail, the video will loop infinitely to make sure the user will always see a moving thumbnail. In addition, the video will automatically start playing when the webpage is loaded and will contain no audio. The latter is done to avoid sudden unexpected sounds that may annoy the user when they visit the page, and to allow multiple moving thumbnails playing at the same time.

4. PROBLEM STATEMENT

With the introduction of the moving thumbnail, a number of new problems arise, one of them being the creation of these moving thumbnails.

Static thumbnails for videos are often manually picked from a list of proposed frames from the video. These proposed frames are generated using techniques like key frame extraction [3] and require little extra manual labour to select an attractive thumbnail.

In order to generate a moving thumbnail, a portion of the original video has to be selected. This can be done by hand, using a video editor to make things easier. However, this introduces a lot of extra manual labour, especially when a newspaper publishes a lot of videos. In order to effectively use moving thumbnails in practise, we have to automatically generate the moving thumbnail.

4.1 Requirements of automatic generation

The goal of the moving thumbnail is to attract users to follow the underlying link to the full article. In order to do this, the moving thumbnail has a few requirements that should be met when generating it.

1. It should invite the user to watch the full video, using interesting shots that may rise the interest of the viewer.

2. The shot length and general timing of the moving thumbnail should be suited for viewing by a human being.

3. The shots used should be suited for the small dimensions of the moving thumbnail.

4. The moving thumbnail doesn't contain any audio, thus the contents should be interesting without audio.

4.2 Specification of the moving thumbnail

The moving thumbnail is not yet being used in practice and thus has no standard in terms of length, size or any other specification. In order to generate usable thumbnails, there are certain specifications that should be met. These specifications are part of the research.

1. The length of the thumbnail.

2. The dimensions of the thumbnail. It should be noted that the actual display of the thumbnail in a production website might differ from these dimensions. The dimensions should provide a standard in the same manner as '720p' or '1080p' do in current online video usage.

3. A target file size of the generated thumbnail. This is important since video quality might degrade when using heavy compression in order to save bandwidth, which in turn might decrease the clarity of the moving thumbnail.

4.3 Research question

The knowledge gained from the study should be the specified features that are used in the generation of these video thumbnails, that are known to affect the effectiveness of the moving thumbnail. These effects can range from shot and overal length, to color saturation or facial recognition.

The research question can be formulated as follows:

What features used in the generation of moving thumbnails affect the curiosity of the viewer, increasing the need for them to see the full video?

5. GENERATING INTERESTING MOVING THUMBNAILS

To fulfil the requirements stated in section 4.1, the main body of the thesis will consist of research to the appropriate features in the generation of moving thumbnails. These features are used to create a moving thumbnail that is interesting for the audience.

5.1 The dataset

The first step in this process is gathering a test and training dataset on which we can test the effect of the features. This dataset should contain both good examples (as in: interesting moving thumbnails) and bad examples (meaning non interesting moving thumbnails). Since we want to test multiple features, the number of videos in this dataset should be among the 100s to 1000s.

Since the moving thumbnail is a new concept that is not actually implemented in a real-world scenario, we cannot get a dataset specifically for this goal. This means that we have to create a new dataset, which is possible using a number of methods.

A specifically designed survey, which gathers data in a controlled environment using specifically selected videos. In order to gain accurate results, the number of people taking this survey would be too big to be realistic.

A real world implementation on an existing website that gathers data using user analytics would be another way of creating a dataset. However, the results could be affected by factors outside the test environment, thus wouldn't be reliable. Finding a website that meets the requirements would be a challenge on its own.

It might be possible to use an existing dataset that may not directly contain the metrics that are required, but can display significant changes when features are changed in other metrics that the dataset contains. An example of such a dataset is the 'Vine' dataset, containing 'Vine' videos that have a lot of views, or no views. The amount of views might specify wether a video is interesting or not. In this case, it would be very important to support the dataset with proper research and logic.

5.2 Extracting features

5.3 Assigning a score

5.4 Measuring the effect of the features

5.5 Research analysis

In order to measure the effectiveness of the system, the automatically generated moving thumbnails will be compared to static thumbnails. This will be done using a custom test setup (perhaps with A/B testing scenarios) in an isolated environment. If possible within the time constraints of the master thesis, there will be a real-world test with cooperation of a customer of Blue Billywig that is interested in using the technology when its ready for production.

An analysis of the test results will determine of the moving thumbnail generation system has added value in terms of conversion and if the moving thumbnail is successful in attracting more audience to the article.

6. DATASET

Since the system has to work with existing data and results with which it can improve itself, the dataset is a very important part of the research and has a number of requirements that have to be met.

6.1 Size

The dataset should contain around 500 to 1000 videos. The system that will evaluate the videos on their ability to function as a teaser, requires learning data. The amount of features that help the system decide this will probably more than a few, thus creating a high-dimensional space in which the system has to find a pattern, increasing the need for a lot of data.

6.2 Training data

Since the system has to learn wether certain videos (or video fragments) are fit to function as a teaser, it requires known training data. The training data is a very key component in the dataset, since it will eventually decide what features in the video affect its popularity. First, this means that training data should be available at all, for example in the form of number of views on the video.

Secondly, this data has to be reliable. Videos on YouTube are often getting watched since the uploader has a lot of 'subscribers', the custom video thumbnail might be very inviting or because the title contains effective keywords. All these factors have nothing to do with the actual quality of the video, and might result in false training data. It is therefor important that the training data is justified and that the use of it is supported with solid arguments.

7. LITERATURE STUDY

7.1 Video summarisation techniques

In order to generate a thumbnail that is actually useable (and not just a random selection of frames), we have to look at various existing techniques that are used in video summarisation in order to generate a logical video that is pleasant to watch for humans.

7.2 Feature detection techniques

The features that affect the attractiveness of the video thumbnail have to be detected in the video using the video itself, and metadata that is attached to the video. These might be high level features like concept detection, or low level features like color variance and shot tempo.

8. PLANNING

References

[1] Cisco. *Cisco Visual Networking Index*. May 2015. URL: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf.

[2] Microsoft, Susan Dziadosz, and Raman Chandrasekar. *Do thumbnail previews help users make better relevance decisions about web search results?* ACM, Aug. 2002. ISBN: 1-58113-561-0. DOI: 10.1145/564376.564446. URL: http://dl.acm.org.proxy.uba.uva.nl:2048/citation.cfm?id=564376.564446.

[3] F Dirfaux. "Key frame selection to represent a video". English. In: *Proceedings of 7th IEEE International Conference on Image Processing* 2 (2000), 275–278 vol.2. ISSN: 1522-4880. DOI: 10.1109/ICIP.2000.899354. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=899354.