

# 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.

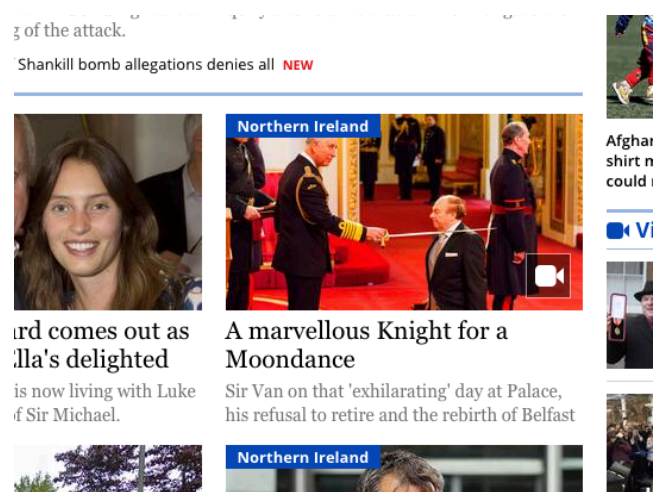


Figure 1: Screenshot of the Belfast Telegraph frontpage (<http://www.belfasttelegraph.co.uk/>) on the 5th of february, showing a thumbnail.

## 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 analysis 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 New challenges for moving thumbnails

Thumbnails are not the main focus of the page, and will not be started by a user action like a regular video would by clicking a play button. The user's attention might also not directly be focussed on the moving thumbnail, and sometimes even multiple moving thumbnails may show on a page. These issues are also present in static thumbnails, but since they have no dimension of time, this does not present itself as an issue. The workings and design of a moving thumbnail should account for this lack of initial focus.

A static thumbnail will show only a single frame of the original video, revealing little of the actual content. A moving thumbnail however, might already show the punch line of the video, eliminating the need for a user to click through to the full video. For end-users, this might not impose as a problem, however media companies gain their revenues by serving advertisements before a video, or next to an article. This highlights the need to create moving thumbnails that tease the user, instead of showing a summary.

A more technical challenge is the design of the moving thumbnail in relation to processing power, internet speeds and device limitations. Even small videos can be significantly larger in file size than static thumbnails, and may consume more processing power, especially on mobile devices. The technical details of the moving thumbnail should consider these limitations.

## 2.2 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.

The system that automatically generates moving thumbnails from a video has to take several requirements into account. Examples of these requirements are that the video should be suited for small dimensions or that the user is invited to watch the whole video. These specific requirements should be defined in a later state, based on existing previous research in the field of static thumbnail generation, video navigation and video summarisation.

# 3. PROBLEM STATEMENT

As of today, there is no system that generates moving thumbnails with the goal to tease the user to view the full video. The specifications and requirements for a moving thumbnail are unclear, a method to measure the effectiveness of the output of the system does not exist for this particular case, and since the concept is new, there is no data available to accurately train a system for this purpose.

The absence of data leads to the core of this research: The goal is to build a system that generates moving thumbnails and evaluates the results automatically to validate the output and allows for rapid improvements without user testing. The automatic evaluation can not rely on real world data (since it does not exist yet) and functions as some sort of sanity check. When the system scores acceptable on the automatic evaluation, the results are implemented in an end user test scenario to evaluate the system and gain real world data for the purpose of generating moving thumbnails.

The research question is as follows:

**How can we build a system that automatically generates moving thumbnails that invite the user to view the full video?**

The following subquestions will support the main research questions:

*How can the automatic thumbnail generator be automatically evaluated?*

*What features play a role in the automatic generation of moving thumbnails?*

The user testing that follows allows us to check whether we have actually succeeded in building such a system. This results in the following subquestion:

*How does the output of the system behave in a real world scenario?*

By answering these questions, we can describe a system that automatically generates moving thumbnails and evaluates its results, which are validated by a real world survey.

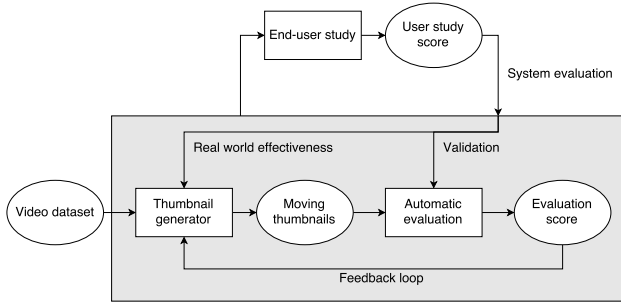
## 3.1 Context

The origin of the new concept of moving thumbnails lies at media companies, who constantly try to seek new ways to improve their website and the way users navigate their news. At Blue Billywig, we received a request to offer a way to show moving thumbnails, created manually by the editors. After investigating this new opportunity, it was decided that an automatic way of generating these moving thumbnails was required.

Blue Billywig is a company that builds and maintains a video management systems, in which its customers can upload videos, edit metadata, adjust the way these videos are displayed and publish the videos on their website. In a lot of ways, the system works the same way as the popular video site YouTube does [3], only focussed on the use by companies and without a community.

## 4. APPROACH

As mentioned in section 3, the research can be divided into three closely related subjects. The first subject is about the moving thumbnail generator itself, a system that uses various features to determine what to use from an input video to create a moving thumbnail as output. The second subject is about a validator that automatically validates the output of the system, to check the integrity of the generated moving thumbnails. The third subject is about the user test scenario that tests the output of the generator with actual users, which results can in turn be used to compare the validator results with the actual user testing, and show real world effectiveness of the thumbnail generator. This workflow is shown in figure 4. These three subjects combined form a study that covers all grounds in this new concept of automatically generating thumbnails.



**Figure 2: Schematic workflow overview of the complete system, showing feedback directions on the generator and evaluation.**

The dataset that will be used will be chosen on the availability of the data, the domain that the research aims for (news media) and the amount of metadata that is needed to gain accurate results. Blue Billywig has a number of customers that have data related to the research, properly annotated with relevant metadata and in satisfying quantity. Another advantage of using an ‘in-house’ dataset is the accessibility of the data and video files. Blue Billywig encodes every video file to a number of different formats to allow playback on any device and internet connection. This preparation of data might significantly increase the progress of the study.

## 5. LITERATURE STUDY

Since the generation of moving thumbnails is a new concept, there is no directly related literature. However, it relates close to a number of other domains: Video navigation and interfaces, static thumbnail generation and video summarisation.

### 5.1 Video navigation

There are a lot of interfaces designed to assist humans in navigating a video library. A summary by Schoeffmann et al. [4] describes over 40 different interfaces that use different techniques to allow convenient browsing through a collection of videos. These range from displaying a key frame best describing the video, or a collection of keyframes with their sizes related to the importance of the frame. Most of these interfaces have some sort of system that automatically determines what to display on the screen, taken into

consideration the different conditions in which the system is used.

A study by Hürst et al. [5] describes a user study to the recognition of video using different thumbnail sizes, numbers and various movement in the thumbnails. The study shows that users are able to handle multiple small thumbnails on mobile devices, especially when the thumbnails included motion.

Since one of the goals of the moving thumbnail is to improve the navigation of users in a news media website, video navigation literature could especially prove useful in the design and implementation of moving thumbnails in overview pages.

### 5.2 Static thumbnail generation

The issue of video navigation using thumbnails has been an active topic of research. In 2015, a study by Kim et al. [6] describes a system that automatically combines video frames to generate a thumbnail containing more information than a single frame. Another study describes thumbnail candidate selection using image quality evaluation [7]. A combination of internal and external analysis of the video content to select thumbnails is used by a study by Liu et al. [8]. The techniques and analysis methods used in these systems can possibly be of use when evaluating the generated moving thumbnails.

Many systems that generate thumbnails use a ranking of different frames to propose a suitable thumbnail [9, 10, 11]. In static thumbnail generation, this ranking can be used to select the best thumbnail. However, in moving thumbnail generation (or other video navigation interfaces) the ranking can be used to create a composition of the video. This creates opportunities to generate moving thumbnails that consist of different shots from the original video.

### 5.3 Video summarisation

Video summarisation is the domain that is closest to the concept of the moving thumbnail, since a moving thumbnail is basically a video summarisation with a few major differences. The length of a video summary would depend on the contents of the actual video, an hour long video would need a longer summary to contain all concepts in the video than a summary for a minute long video. The moving thumbnail however, should ideally always be around the same size, since it only has to attract users to view the full video, instead of showing what the video is about.

Techniques for video summarisation can be divided into three different categories: Internal analysis, external analysis and a hybrid of the two. Internal analysis techniques use information that is gained from analysing the video itself, while external analysis uses the information available outside the video (metadata like title, description or tags). The hybrid techniques combine both internal and external analysis [12].

The knowledge gained from video summarisation can be used to generate and extract relevant parts from a video, and those in turn can be used to create moving thumbnails. The difference between video summarisations and moving

thumbnails has to be highlighted in order to modify (or use techniques from) video summarisation techniques in order to use them for another purpose. Various frameworks and surveys on different techniques highlight the differences and effectiveness in domains and content [12, 13], which might be useful in the design of the moving thumbnail generator.

## 6. PLANNING

Development of the system will be planned in a scrum-like planning, implemented in the scrum planning of Blue Billywig. This means that the planning is divided into sprints of two weeks. The planning, as shown in figure ??, describes a number of topics that should be covered in those sprints. The columns ‘BB’ describes the sprint numbers as used in the Blue Billywig sprint planning.

Working on the master thesis document itself is done throughout the whole planning, and has no defined deadlines other than the draft versions in sprint 9 and 10.

### 6.1 Plan B

One of the highest priorities in the system is the user study. Without the user study, the results of the system, as well as the effectiveness of the validator cannot be measured. Thus, it is important that the user study is ready in sprint 6 to gather enough data to gain significance. Sprint 5 is mostly used to improve the overall system, and can be changed when the workflow isn’t properly set up in sprint 4. The overall planning can be extended with one sprint by removing one of the draft versions of the master thesis. However, this should be an extreme measure when the system has severe delays and the system can not be validated yet.

## 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](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] Google. *YouTube*. URL: <https://www.youtube.com/>.
- [4] Klaus Schoeffmann et al. “Video browsing interfaces and applications: a review”. In: *Journal of Photonics for Energy* (2010), DOI: 10.1117/6.0000005. URL: <http://photonicsforenergy.spiedigitallibrary.org/article.aspx?articleid=1305555>.
- [5] Wolfgang Hürst et al. “Size Matters! How Thumbnail Number, Size, and Motion Influence Mobile Video Retrieval”. English. In: *Advances in Multimedia Modeling*. Springer Berlin Heidelberg, Jan. 2011, pp. 230–240. ISBN: 978-3-642-17828-3. DOI: 10.1007/978-3-642-17829-0\_22. URL: [http://link.springer.com.proxy.uba.uva.nl:2048/chapter/10.1007/978-3-642-17829-0\\_22](http://link.springer.com.proxy.uba.uva.nl:2048/chapter/10.1007/978-3-642-17829-0_22).
- [6] Jongdae Kim et al. “Comprehensible Video Thumbnails”. English. In: *Computer Graphics Forum* 34.2 (May 2015), pp. 167–177. DOI: 10.1111/cgf.12550. URL: <http://onlinelibrary.wiley.com.proxy.uba.uva.nl:2048/doi/10.1111/cgf.12550/full>.
- [7] Weigang Zhang et al. “Web video thumbnail recommendation with content-aware analysis and query-sensitive matching”. English. In: *Multimedia Tools and Applications* 73.1 (2014), pp. 547–571. DOI: 10.1007/s11042-013-1607-5. URL: <http://link.springer.com.proxy.uba.uva.nl:2048/article/10.1007/s11042-013-1607-5/fulltext.html>.
- [8] Wu Liu et al. “Multi-Task Deep Visual-Semantic Embedding for Video Thumbnail Selection”. In: *Proceedings of the IEEE ...* (2015), pp. 3707–3715. URL: [http://www.cv-foundation.org/openaccess/content\\_cvpr\\_2015/html/Liu\\_Multi-Task\\_Deep\\_Visual-Semantic\\_2015\\_CVPR\\_paper.html](http://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Liu_Multi-Task_Deep_Visual-Semantic_2015_CVPR_paper.html).
- [9] Jiwon Choi and Changick Kim. “A framework for automatic static and dynamic video thumbnail extraction”. English. In: *Multimedia Tools and Applications* (2015), pp. 1–17. DOI: 10.1007/s11042-015-2909-6. URL: <http://link.springer.com.proxy.uba.uva.nl:2048/article/10.1007/s11042-015-2909-6/fulltext.html>.
- [10] Weigang Zhang et al. *A Novel Framework for Web Video Thumbnail Generation*. English. IEEE, 2012. ISBN: 978-1-4673-1741-2. DOI: 10.1109/IIH-MSP.2012.89. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6274250>.
- [11] Yuli Gao, Tong Zhang, and Jun Xiao. “Thematic video thumbnail selection”. English. In: *Image Processing (ICIP)* (2009), pp. 4333–4336. ISSN: 1522-4880. DOI: 10.1109/ICIP.2009.5419128. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5419128>.
- [12] Arthur G Money and Harry Agius. “Video summarisation: A conceptual framework and survey of the state of the art”. English. In: *Journal of Visual Communication and Image Representation* 19.2 (Feb. 2008), pp. 121–143. DOI: 10.1016/j.jvcir.2007.04.002. URL: <http://linkinghub.elsevier.com/retrieve/pii/S1047320307000247>.
- [13] Muhammad Ajmal et al. “Video Summarization: Techniques and Classification”. English. In: *Computer Vision and Graphics*. Springer Berlin Heidelberg, Sept. 2012, pp. 1–13. ISBN: 978-3-642-33563-1. DOI: 10.1007/978-3-642-33564-8\_1. URL: [http://link.springer.com.proxy.uba.uva.nl:2048/chapter/10.1007/978-3-642-33564-8\\_1](http://link.springer.com.proxy.uba.uva.nl:2048/chapter/10.1007/978-3-642-33564-8_1).