Storyboard-Based Video Browsing Using Color and Concept Indices

Wolfgang Hürst[1(✉)], Algernon Ip Vai Ching[1], Klaus Schoeffmann[2], and Manfred J. Primus[2]

[1] Utrecht University, Utrecht, Netherlands
huerst@uu.nl, A.A.K.IpVaiChing@students.uu.nl
[2] Klagenfurt University, Klagenfurt, Austria
{ks,mprimus}@itec.aau.at

Abstract. We present an interface for interactive video browsing where users visually skim storyboard representations of the files in search for known items (known-item search tasks) and textually described subjects, objects, or events (ad-hoc search tasks). Individual segments of the video are represented as a color-sorted storyboard that can be addressed via a color-index. Our storyboard representation is optimized for quick visual inspections considering results from our ongoing research. In addition, a concept based-search is used to filter out parts of the storyboard containing the related concept(s), thus complementing the human-based visual inspection with a semantic, content-based annotation.

Keywords: Video browsing · Visual inspection · Concept-based indexing

1 Motivation

Previous years of the Video Browser Showdown (VBS) competition have shown that optimized storyboard representations are an effective way for quickly skimming the content of reasonably sized video archives. For example, in 2015, we participated in the VBS with a storyboard-based visualization of a 100-hour video database [2]. Despite no content analysis, the optimized storyboard layout enabled users to quickly and efficiently search for known items, resulting in the 3rd place of the overall competition. Likewise, concept-based search, where a set of predefined and trained semantic concepts is used to search in the videos, has shown to work particularly well with larger databases. For example, in 2016, when the size of the database was increased to 200 h, the concept-based search in our system performed particularly well in the part of the VBS competition where targets have been presented by textual descriptions [1], resulting in the 2nd place of the overall competition. Our contribution for this year, where the data size has been significantly increased to 600 h, therefore aims at combining the advantages of both approaches into one system. In the following, we summarize the preprocessing of the data and indexing – including storyboard layout, color index, and concept detection. Then, we describe a common search process, also illustrating the basic idea behind our approach, and conclude with a short discussion.

L. Amsaleg et al. (Eds.): MMM 2017, Part II, LNCS 10133, pp. 480–485, 2017.
DOI: 10.1007/978-3-319-51814-5_45

2 Indexing

To visually inspect and find a target scene, we represent video content via a storyboard, i.e., a temporarily sorted set of thumbnails of frames extracted from the videos. The layout of this temporal arrangement is optimized based on our research about visual perception [3, 4] and experience from previous years' VBS participations [1, 2]. Figure 1 illustrates the basic design.

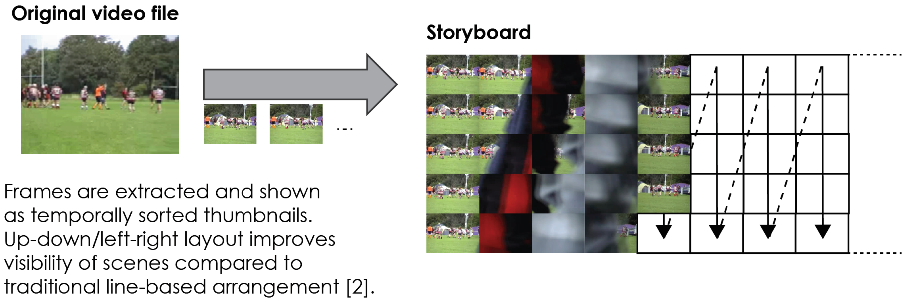

Fig. 1. Storyboard layout. (Color figure online)

In the original storyboard version from the 2015 competition [2], we used a large storyboard containing all video files sorted by their duration. Given the significantly increased dataset of 600 h for this year's event, single video files are now no longer represented as a whole but instead separated into segments created during indexing. Figure 2 summarizes this indexing process. First, a storyboard for each video is created by extracting frames and aligning them in the layout introduced in Fig. 1. These individual storyboards for each video in the database are later also used in the search process (cf. Sect. 3).

Then, each video is split into segments of 25 frames. This can be done in a brute force manner by just combining 25 consecutive frames. Alternatively, content analysis, such as automatic shot detection can be used. In the latter case, segments are formed by taking 25 equally distributed frames from a shot to form the scene. Experiments comparing the impact of these two and other content-related segmentations on search performance are part of our ongoing research.

In the next step, an index is created for each segment. First, an identifying color is assigned by analyzing the histograms of the 25 frames. A related score expresses how much this color represents the whole scene depicted by these frames. Then, semantic concepts are assigned to the center frame using comparable techniques like in our system from the 2016 VBS competition [1]. Similar to the dedicated color, each assigned concept has an associated confidence score indicating the likelihood of that concept appearing in that particular video frame.

Based on this index, an initial storyboard containing the whole database is created. To do this, segments are sorted by color, and then by their related color confidence score. This color-sorted storyboard is then used as starting point for the search, as we will describe in the next section.

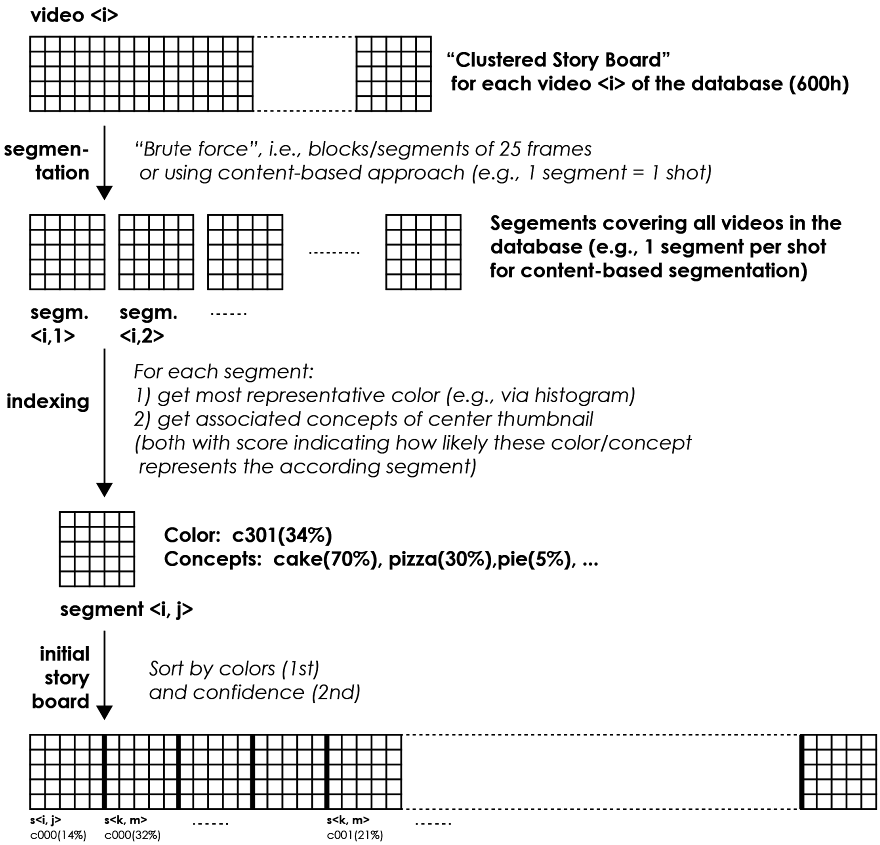

Fig. 2. Indexing (overview).

3 Searching

The initial storyboard of color-sorted segments created in the indexing process is shown on the screen at the beginning of the search. Users can skim it by scrolling up and down. Direct access to particular parts is possible via a color index on the left side. Concept-based browsing is done via text entry or selection of concepts from a list via a popup menu. In the following we describe a typical search process as it appears in our system for known item search (KIS) and ad-hoc video search (AVS) – the two tasks evaluated at the VBS. Figure 3 gives an overview of the procedure for color-based search.

For KIS tasks, a 20 s clip is played, which represents the target, i.e., the "known item" that the user has to find. In our approach, the user identifies a scene or shot that has a dominating color – this might be the most frequent one (e.g., green from the grass of a football field in a soccer video) or a "color tone" reflected in the whole image (e.g., a landscape colored in orange due to the setting sun). Clicking on the corresponding

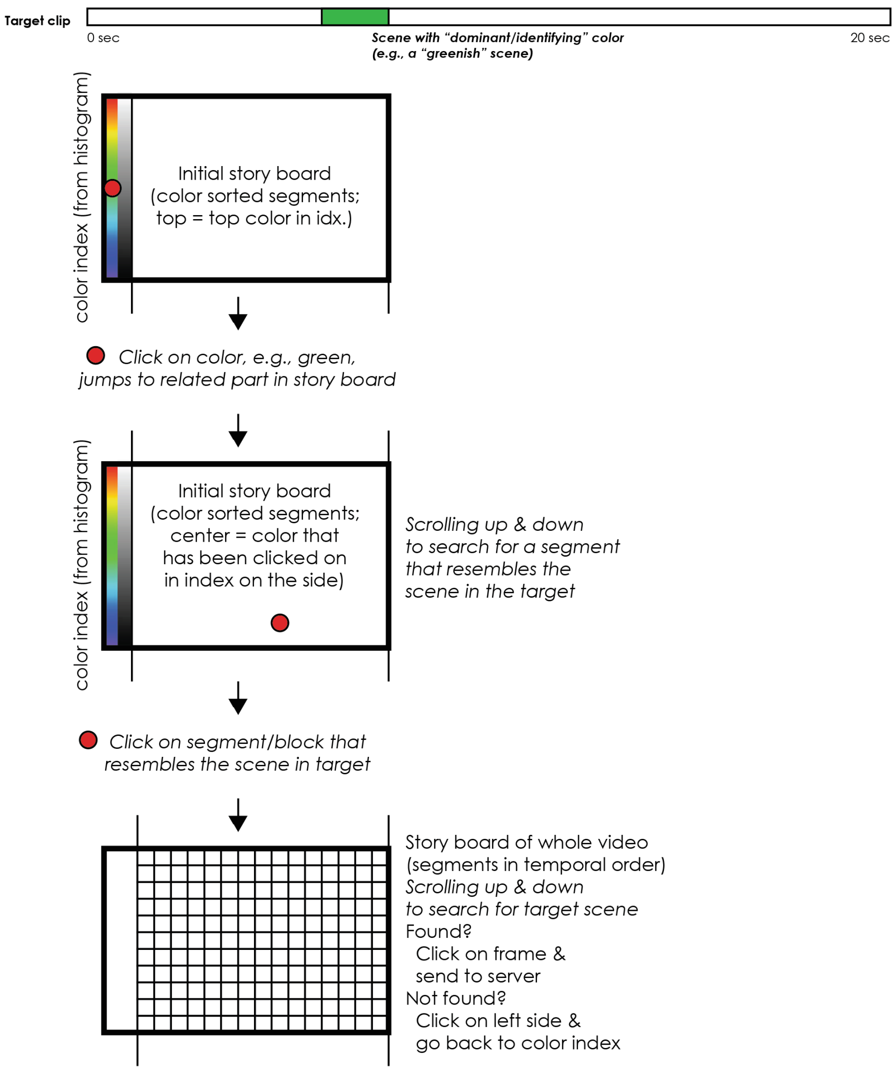

Fig. 3. Color-based search (exemplary overview). (Color figure online)

color in the color bar on the left repositions the storyboard around the segments that have this color assigned in their index. Users can then visually inspect the data by scrolling the storyboard up and down. Of course, we cannot assume that the segment is directly visible. First, identifying the dominating color is not an easy task for humans, and second, there might be many segments with the same or very similar colors. Yet, the optimized visualization and scrolling mechanisms also used in our system from 2015 [2] should enable users to quickly find segments similar to the searched target. Once such a target candidate has been identified, users can click on it to get a view of the whole file's storyboard that was created for each video at the begin of the indexing process (cf. Fig. 1).

AVS tasks rely on a textual description of a subject, object, or event, and thus, color-based search is only useful in some situations. Instead, concept-based search is generally a more promising approach. In our system, users can select a concept from the index via text entry or pop-up menu. This removes all segments from the original, color-sorted storyboard that do not have this concept in their index. If the concept is very sophisticated or if there are rather few segments with that concept in the database, users can then just search for targets by scrolling through the reduced storyboard. In other cases, users can select a second concept, further reducing the size of the storyboard, and so on. Likewise, users can jump to a particular color in the reduced storyboard – which is still sorted by colors. For example, a user can select the concept "soccer", which will likely show many segments with green as dominating color, but then jump to the color white if the target scene happens to be a game that was played in winter with snow covering the green grass of the playing field. Similarly, color-based search can be followed by concept-based search. For example, a scene with people partying on the deck of a boat on the ocean can be found searching for blue segments, if the dominating color is blue from the water surrounding the boat, followed by applying a filter for the concepts "ship/boat" and "people".

4 Discussion

Our system uses a rather simplified approach for search, and relies heavily on human browsing and human search performance. For the color-based search, users need to be able to identify dominating colors in scenes and visually match the small thumbnail representations of segments to target scenes. For concept-based search, users have to identify the concepts in the system and then again preform some challenging visual inspections of the presented target candidates. Yet, our previous work and participations from the last two years have shown that humans perform extremely well for such tasks, and thus the combination with simplistic, but powerful search functionality is a promising approach that we now further optimize with this year's system design.

References

1. Hudelist, M.A., Cobârzan, C., Beecks, C., Werken, R., Kletz, S., Hürst, W., Schoeffmann, K.: Collaborative video search combining video retrieval with human-based visual inspection. In: Tian, Q., Sebe, N., Qi, G.-J., Huet, B., Hong, R., Liu, X. (eds.) MMM 2016. LNCS, vol. 9517, pp. 400–405. Springer, Heidelberg (2016). doi:10.1007/978-3-319-27674-8_40
2. Hürst, W., van de Werken, R.: Human-based video browsing-investigating interface design for fast video browsing. In: 2015 IEEE International Symposium on Multimedia (ISM), pp. 363–368. IEEE, December 2015

3. Hürst, W., Darzentas, D.: Quantity versus quality: the role of layout and interaction complexity in thumbnail-based video retrieval interfaces. In: Proceedings of the 2nd ACM International Conference on Multimedia Retrieval (ICMR 2012), Article 45, p. 8. ACM, New York (2012)
4. Hürst, W., Snoek, C.G.M., Spoel, W.-J., Tomin, M.: Keep moving! Revisiting thumbnails for mobile video retrieval. In: Proceedings of the International Conference on Multimedia (MM 2010), pp. 963–966. ACM, New York (2010)