Sketch-based Similarity Search for Collaborative Feature Maps

Andreas Leibetseder, Sabrina Kletz, and Klaus Schoeffmann

Klagenfurt University (AAU)
Institute of Information Technology
9020 Klagenfurt, Austria
{aleibets,sabrina,ks}@itec.aau.at

Abstract. Past editions of the annual Video Browser Showdown (VBS) event have brought forward many tools targeting a diverse amount of techniques for interactive video search, among which sketch-based search showed promising results. Aiming at exploring this direction further, we present a custom approach for tackling the problem of finding similarities in the TRECVID IACC.3 dataset via hand-drawn pictures using color compositions together with contour matching. The proposed methodology is integrated into the established Collaborative Feature Maps (CFM) system, which has first been utilized in the VBS 2017 challenge.

Keywords: interactive video search, collaboration, sketch-based search

1 Introduction

Collocated at the International Conference on Multimedia Modeling (MMM), VBS[1] is challenging participating teams to compete in solving former TRECVID tasks of known-item search (KIS) [10] and a currently also include ad-hoc video search (AVS) [1] in a large video dataset[2]. Past VBS events showed that the concept of drawing simple sketches for relevant shot retrieval can be leveraged to achieve excellent results [5], even in cases surpassing all other approaches and coming out victorious, as has been the case in 2014 [6] as well as 2015 [2]. CFM [9], first introduced in VBS 2017, offer a great variety of collaborative search modalities, yet, other than a simple dominant color filter, it does not offer the possibility of sketch-based similarity search. Further improving the system, we thus incorporate custom sketch-based similarity search as an extension, as illustrated in Figure 1. Among several already built-in extensions, the newly integrated feature takes a hand-drawn image as an input and in a first filtering step retrieves TRECVID keyframes exhibiting a similar color composition. In a second step these results are further refined by matching selected contours from the sketch. Following sections describe the proposed addition in more detail: Section 2 gives an overview over the proposed approach's architecture, starting

[1] VBS has first been organized in 2012 [8].
[2] The current IACC.3 dataset contains 600 hours of video.

out with Section 2.1 describing the user interface and ending in Section 2.2 detailing the underlying methodologies. Finally, Section 3 concludes this work.

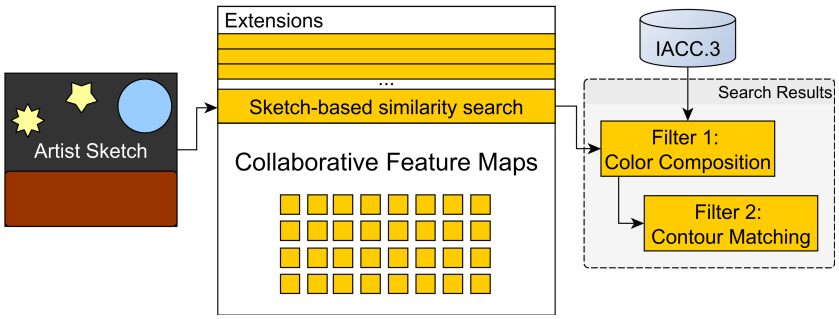

Fig. 1: Collaborative Feature Maps extension: sketch-based workflow.

2 Sketch-based Similarity Search

As indicated above, the extension's main workflow can be split up into two separate tasks: drawing and reviewing results. Although created sketches should be fairly rudimentary, as is shown in Figure 2 portraying the VBS 2017 query topics "palm tree" and "playing guitar outdoor", the activity can be much more time-consuming than for example text-based search. Hence, it is best for the collaborating team members to dynamically assume different roles as needed. Accordingly, a plausible scenario for handling a given query could well be all team members starting out drawing and gravitating towards either continuing this task, i.e. illustrating an alternate perspective or adding more details, or processing result lists of already created sketches.

Fig. 2: Sample sketches drawn by an artist.

Since tablets, smart phones or any other pen input device are growing in illustration usage, these devices provide simple sketch-based drawing tools, which

are sufficient to create rudimentary sketches. Therefore, the system does not provide a supplementary interface to create sketches. The artists simply use tools of their choice to create digital images, which they transfer to a location visible for the system. Upon receiving completed sketches the system conducts implemented similarity searches (see Section 2.2), to retrieve result lists for each of the provided paintings, displayed in the provided result processing interface.

2.1 Interface

We extend the traditional interface of the system by an additional view, shown in Figure 3. After choosing a particular drawing from any available input device, a user is provided with a simple sketch visualization and ranked result lists according to several partially user-defined criteria.

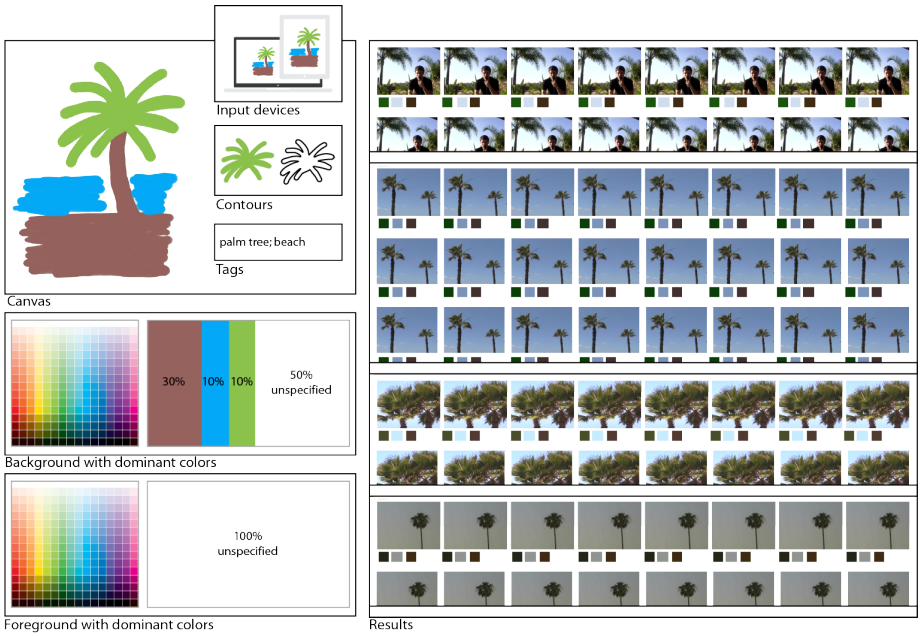

Fig. 3: Screenshot of the extended color sketch view - prototype UI

The main focus of the sketch view is the refinement of paintings to narrow down the result set. A reduced set of dominant colors, i.e. color composition, are extracted for the sketch, which initially is treated entirely as belonging to an image background [4, 7]. Optionally, a user can define parts of the drawing as a foreground region, which is useful for bringing focus to moving objects. All detected dominant colors are displayed according to their visibility in the sketched scene. The presence of colors is indicated by the percentage of their

occurrence including the unpainted or transparent area, intentionally left blank by the artist. Collaborators are able to change, add or remove colors using a preset reduced color palette and even re-weight color dominance.

Beside the refinement of the color compositions, contours are an essential feature outlining shapes. In the majority of cases a rough color sketch consists of color regions, which do not necessarily contain meaningful contours. Therefore coherent areas, such as the crown of the palm tree can be selected and used as expanded matching criteria. Since the sketch view is integrated into the CFM system and in the case that shapes do not result with the desired content, the underlying query can as well be searched via textual filtering (see 'Tags' filtering option in figure), which utilizes the indexed concepts already provided by the system.

Finally, when a new sketch is processed, i.e. all filtering criteria have been applied, the system interface displays the results in a storyboard arrangement. The result set is split up into shots with one storyboard each. The storyboards are arranged in sub views line by line and consist of frames corresponding to the individual shots. In each sub view, the frames are aligned from left to right and from top to bottom. The visible part of the sub view starts with the keyframe containing the sketched scene or object and the following frames are arranged by time stamp. To get an overview of the whole shot, the corresponding sub view can be expanded and still remaining frames will be loaded. The sub views are ranked according to their similarity to the input sketch and can be re-ranked by various matching criteria, as described previously.

2.2 Details

While traditional sketch-based retrieval systems provide line-based sketches, either with colored lines or simple black lines, color sketches that outline a scene with color regions lack in detail to use shapes as matching criteria [3, 11]. Motivated by this drawback, the system extension exploits a two-fold approach, where for any given sketch first color compositions are matched against precalculations on the TRECVID dataset and then selected coherent contours are used for shape matching. Whereas drawing objects with contours is often problematic due to the vast amount of different potential perspectives, dominant colors are considered the best choice to describe a scene independent of its position in an image and contained movement. Even black and white scenes can be outlined by means of different grayscale regions. Since shots contain several dominant color compositions changing in time and space, the dominant colors of each contained frame are separately addressed and used as a matching criteria.

The implementation of color compositions is based on image segmentation approaches [4, 7]. This has the advantage that due to the color reduction, the most dominant colors and as well roughly outlined shapes can be identified. In case the scene contains moving objects, they are indicated as belonging to the foreground when applying background subtraction. Therefore, user-defined foreground objects in drawings can merely match TRECVID scenes containing moving objects, i.e. with an existing foreground. Regardless of background or

foreground objects, edge detection approaches provide at least some contours that can be used to align outlines using distance transformations [11]. However, drawing color regions and objects and matching them to their most similar occurrence in shots requires several computationally intensive steps. Therefore, dominant colors and objects in the dataset are pre-indexed, while merely a sketch is processed online, i.e dominant colors and contours are extracted for applying two-fold similarity measures. Thus, an initial result list can be produced according to color composition and then narrowed down through matching contours.

3 Conclusion

We present an extension for the Collaborative Feature Maps system used in the Video Browser Showdown 2017 challenge based on hand-drawn sketches of collaborating artists. The extension employs a two-fold strategy for finding similar shots to corresponding drawings: matching dominant colors on the pre-indexed TRECVID dataset and allowing for further result refinement via determining similar object contours. Segmentation techniques further allow for a more fine-grained sketch partitioning and retrieval improvement.

References

1. Awad, G., Fiscus, J., Michel, M., Joy, D., Kraaij, W., Smeaton, A.F., Quénot, G., Eskevich, M., Aly, R., Ordelman, R.: TRECVID 2016: Evaluating Video Search, Video Event Detection, Localization, and Hyperlinking. In: Proceedings of TRECVID. vol. 2016 (2016)
2. Blažek, A., Lokoč, J., Matzner, F., Skopal, T.: Enhanced Signature-Based Video Browser, pp. 243–248. Springer International Publishing, Cham (2015)
3. Bui, T., Collomosse, J.: Scalable Sketch-Based Image Retrieval Using Color Gradient Features. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1012–1019 (2016)
4. Chen, J., Pappas, T.N., Mojsilović, A., Rogowitz, B.E.: Adaptive perceptual color-texture image segmentation. IEEE Transactions on Image Processing 14(10), 1524–1536 (2005)
5. Cobârzan, C., Schoeffmann, K., Bailer, W., Hürst, W., Blažek, A., Lokoč, J., Vrochidis, S., Barthel, K.U., Rossetto, L.: Interactive video search tools: a detailed analysis of the video browser showdown 2015. Multimedia Tools and Applications 76(4), 5539–5571 (Feb 2017)
6. Lokoč, J., Blažek, A., Skopal, T.: Signature-Based Video Browser, pp. 415–418. Springer International Publishing, Cham (2014)
7. Mojsilović, A., Hu, J., Soljanin, E.: Extraction of perceptually important colors and similarity measurement for image matching, retrieval, and analysis. In: IEEE Transactions on Image Processing. vol. 11, pp. 1238–1248 (nov 2002)
8. Schoeffmann, K.: A user-centric media retrieval competition: The video browser showdown 2012-2014. IEEE MultiMedia 21(4), 8–13 (Oct 2014)
9. Schoeffmann, K., Primus, M.J., Muenzer, B., Petscharnig, S., Karisch, C., Xu, Q., Huerst, W.: Collaborative Feature Maps for Interactive Video Search. In: Proceedings of the International Conference on Multimedia Modeling, MMM 2017 (2017)

10. Smeaton, A.F., Over, P., Kraaij, W.: Evaluation campaigns and trecvid. In: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval. pp. 321–330. MIR '06, ACM, New York, NY, USA (2006)
11. Sun, X., Wang, C., Xu, C., Zhang, L.: Indexing billions of images for sketch-based retrieval. In: Proceedings of the 21st ACM international conference on Multimedia - MM 2013. pp. 233–242 (2013)