1-2017

Concept-based interactive search system

Yi-Jie LU
City University of Hong Kong

Phuong Anh NGUYEN
City University of Hong Kong

Hao ZHANG
City University of Hong Kong

Chong-wah NGO
Singapore Management University, cwngo@smu.edu.sg

Concept-Based Interactive Search System

Yi-Jie Lu(✉), Phuong Anh Nguyen, Hao Zhang, and Chong-Wah Ngo

Department of Computer Science,
City University of Hong Kong, Kowloon, Hong Kong
{yijie.lu,panguyen2-c,hzhang57-c}@my.cityu.edu.hk, cscwngo@cityu.edu.hk

Abstract. Our successful multimedia event detection system at TRECVID 2015 showed its strength on handling complex concepts in a query. The system was based on a large number of pre-trained concept detectors for textual-to-visual relation. In this paper, we enhance the system by enabling human-in-the-loop. In order to facilitate a user to quickly find an information need, we incorporate concept screening, video reranking by highlighted concepts, relevance feedback and color sketch to refine a coarse retrieval result. The aim is to eventually come up with a system suitable for both Ad-hoc Video Search and Known-Item Search. In addition, as the increasing awareness of difficulty in distinguishing shots of very similar scenes, we also explore the automatic story annotation along the timeline of a video, so that a user can quickly grasp the story happened in the context of a target shot and reject shots with incorrect context. With the story annotation, a user can refine the search result as well by simply adding a few keywords in a special "context field" of a query.

Keywords: Video search · Known-Item Search · Concept bank · Semantic query · Video reranking · Story annotation

1 Introduction

In TRECVID 2015, we developed a multimedia event detection system for zero-example event detection that achieved the best performance [4]. The core of the system is a large concept bank that contains about 2,800 pre-trained concept detectors covering common objects, actions, scenes and everyday activities. To perform a text query search in an unannotated video corpus, the crux of the system is to solve the textual-to-visual relation using the concept bank as a knowledge base.

We have studied several facts which significantly impact retrieval performance. Such facts include the number of concepts, concept specificity, and concept discriminativeness regarding the query. However, the performance of an automatic video retrieval system is still far from perfection, especially when no precisely matched concepts can be found in the concept bank. In this case, the system would propose concepts with the smallest word distance towards the query. This metric, however, often suggests off-topic concepts due to a lack of

L. Amsaleg et al. (Eds.): MMM 2017, Part II, LNCS 10133, pp. 463–468, 2017.
DOI: 10.1007/978-3-319-51814-5_42

common sense that can distinguish a concept from the context of the query topic. A feasible solution is to employ a human evaluator to quickly adjust the result by screening the machine-proposed concepts. On the other hand, although our existing system can be adapted to Ad-hoc Video Search, it is inefficient for Known-Item Search. This is because a text query is insufficient to describe the fine details which are required to mine the exact query clip from a number of clips sharing the same semantic content. Hence, a human needs to painstakingly dig into hundreds of results to find the correct match even if the top results are all relevant. We, therefore, seek help from an interactive search where a user can refine a first-time search result with different methods so that the correct match has a higher chance to show up.

Video Browser Showcases [8] in previous years suggest using high-level visual concepts [5–7] and low-level visual descriptors [1,2] as two lines of approach. For Known-Item Search, the systems using low-level features generally have an advantage over those using high-level concepts. It is worth to mention that a color sketch search method was shown to be very effective in 2014 and 2015 [1,3]. But as low-level features do not contain semantic information, the systems with high-level concepts have their inherent benefit on Ad-hoc Video Search where queries are only formed by text. In this paper, we tend to integrate both methods into an interactive search system. The concept-based search system is mainly used for generating the first-time search result. Then, we implement different reranking techniques to incorporate the strength of both high-level concepts and low-level features. Specifically, highlighted concept reranking is a simple and quick method for a user to emphasize a particular characteristic in the query. When a user finds one or more visually relevant clips in the search result, either relevance feedback or color sketch can be further exploited to refine the result so that the user has a better chance to hit the correct answer. Furthermore, there is an increasing awareness of difficulty in distinguishing shots sharing very similar scenes in the search result [1]. We recount the dominant concepts along the timeline of a video to facilitate video browsing so that a user can quickly grasp the context of a target shot even though the shot itself is not distinctive. We also implement a *context field* in the query to quickly refine the result in this scenario. The following sections detail each component of our system.

2 Concept-Based Video Search System

We adapt our zero-example event search system to general-purpose video search. The search system is backed by a large concept bank which contains thousands of concept detectors for textual-to-visual relation. The most important module in our system is called *semantic query generation* which generates the internal query representation by calculating the distance from a query to each concept. The internal query, a.k.a. the *semantic query* is formed by a number of selected concepts with their weights. The weight calculation and concept selection are discussed in our paper [4].

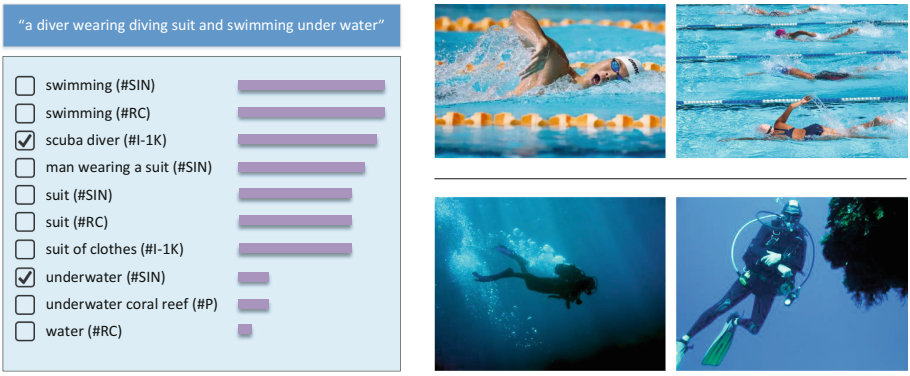

Fig. 1. (a) An example of semantic query editing. (b) Training examples for the concept "swimming" vs. the meaning of the term *swimming* under the context of the query.

As the queries in Ad-hoc Video Search are generally more specific and shorter, the term weights based only on the query are unreliable compared to the prolonged event query used in multimedia event detection. We, therefore, shed more light on semantic query editing by involving human-in-the-loop. As illustrated in Fig. 1a, given a text query *"a diver wearing a diving suit and swimming under water"* without any further editing, the system first generates a list of candidate concepts[1] loosely relevant to the query. The weight of each concept is indicated by a weight bar. A user then can quickly refine the concept list by removing wrong and non-discriminative [4] concepts, and watch the search result change at the same time. Figure 1b shows a typical wrong concept "swimming" which is easily identified by a human but difficult by a machine algorithm. As in a human's sense, the term *swimming* under the context of the query means *underwater diving* which is visually different from the sport *swimming* the concept automatically proposed. Furthermore, we also allow a user to adjust a concept's weight in order to strengthen or weaken the concept. For example, the concept "person" is not important in most query examples because the term is too common. While in some rare cases, such as *"a person sitting beside a laptop,"* the concept "person" should not be depreciated. A user thus can manually increase the weight for "person."

3 Video Reranking

In order to facilitate Known-Item Search, we implement three methods for video reranking. A user can adjust the scope of a reranking method. By default, a reranking is only performed within the top videos recommended by the concept-based video search. For example, *highlighted concept reranking* is most effective

[1] The tag in the brackets of Fig. 1a denotes the dataset from which the concept comes.

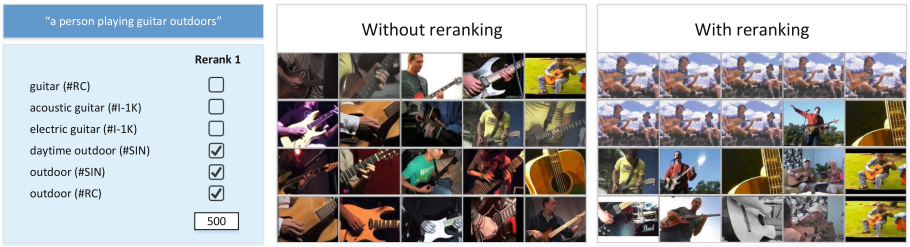

Fig. 2. (a) Concept reranking by emphasizing concepts about "outdoor". (b) The top results change accordingly.

in the scope of the top 300–500 videos. This limitation ensures that the algorithm is not applied to the semantically irrelevant videos at the bottom of the rank list.

Highlighted concept reranking is a simple and quick reranking approach used to highlight particular characteristics in a semantic query. Figure 2a emphasizes the concept "outdoor" in the query *"a person playing guitar outdoors."* As shown in Fig. 2b, the top retrieval result of the original semantic query mixes guitar playing both indoors and outdoors. It is reasonable to highlight the concept "outdoor." But, if we simply increase the weight of "outdoor" in the semantic query, it would pull up noisy outdoor activities which do not contain guitar playing at all. A feasible way is thus to rerank only within the clips about guitar playing. Figure 2b shows the reranking result in the scope of the top-500 clips.

Relevance feedback is used when a user identifies one or more visually relevant clips. Even with highlighted concept reranking, the retrieval result is still diverse if the search system is only based on the high-level semantic concepts. Once a user has identified some relevant clips, these clips can be served as training examples having fine-grained visual details. We intuitively want to refine the result using these visual details. We train SVM classifiers for the user picked clips and rerank the result according to this feedback. The new result is expected to be much more specific and focused on the visually similar clips according to the user's choice.

Color sketch was a very successful approach in Video Browser Showcase 2014 and 2015 [1,3]. Basically, color sketch uses position-color features. These low-level features characterize the colors with their positions on a keyframe. A user can perform the search by simply drawing a few color circles on the empty canvas. We incorporate this search approach to be a reranking alternative mainly for its accuracy on Known-Item Search. In our system, the user can not only draw a new sketch but also use the color sketch automatically extracted from several marked clips in the search result for reranking.

4 Context Annotation

Video Browser Showcase 2014 raised a critical problem in Known-Item Search that it was difficult to distinguish the shots with very similar scenes in a search

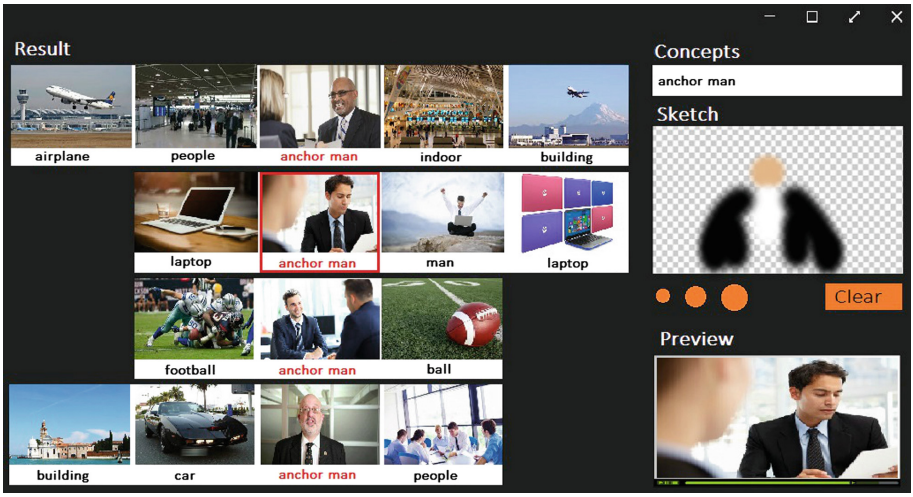

Fig. 3. The context view for a search result. (Color figure online)

result. The problem was noticeable when a large portion of the query clip was about a TV studio scene [1]. To tackle this problem, we automatically annotate the master shots along the whole timeline of a video by its dominant concepts, then fold the adjacent shots sharing the same dominant concept. This process is called *story annotation*. With this annotation, we can enhance the result presentation by showing a *context view* of a target shot. Figure 3 is an example. When the query is a common concept/scene, such as an anchor man, we expect multiple relevant shots of the similar scene to appear in the search result (red centered in Fig. 3). Although hardly any decision can be made by the shots themselves, by expanding the result to a *context view* (images with black text underneath), we can easily grasp the story around each shot and thus distinguish these shots. The benefit of story annotation is not limited to the result presentation. In addition, we implement a special *context field* in the query for quickly screening the search result. A user may simply add a few keywords in the context field of the query to refine the search results, eventually coming up with the shots having matched context only. For instance, when querying a report of a flooded village but the query clip is mostly an anchor person in a news studio, other than describing the exact query clip in the system query, we tend to add keywords like *flood*, *rooftop*, *rescue man*, and even *river* (which is visually similar to flood) in the context field.

Acknowledgments. The work described in this paper was supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (CityU 11210514 and CityU 11250716).

References

1. Blažek, A., Lokoč, J., Matzner, F., Skopal, T.: Enhanced signature-based video browser. In: He, X., Luo, S., Tao, D., Xu, C., Yang, J., Hasan, M.A. (eds.) MMM 2015. LNCS, vol. 8936, pp. 243–248. Springer, Heidelberg (2015). doi:10.1007/978-3-319-14442-9_22
2. Cobârzan, C., Hudelist, M.A., Fabro, M.: Content-based video browsing with collaborating mobile clients. In: Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., O'Connor, N. (eds.) MMM 2014. LNCS, vol. 8326, pp. 402–406. Springer, Heidelberg (2014). doi:10.1007/978-3-319-04117-9_46
3. Lokoč, J., Blažek, A., Skopal, T.: Signature-based video browser. In: Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., O'Connor, N. (eds.) MMM 2014. LNCS, vol. 8326, pp. 415–418. Springer, Heidelberg (2014). doi:10.1007/978-3-319-04117-9_49
4. Lu, Y.J., Zhang, H., de Boer, M., Ngo, C.W.: Event detection with zero example: select the right and suppress the wrong concepts. In: ACM ICMR (2016)
5. Moumtzidou, A., Mironidis, T., Apostolidis, E., Markatopoulou, F., Ioannidou, A., Gialampoukidis, I., Avgerinakis, K., Vrochidis, S., Mezaris, V., Kompatsiaris, I., Patras, I.: VERGE: a multimodal interactive search engine for video browsing and retrieval. In: Tian, Q., Sebe, N., Qi, G.-J., Huet, B., Hong, R., Liu, X. (eds.) MMM 2016. LNCS, vol. 9517, pp. 394–399. Springer, Heidelberg (2016). doi:10.1007/978-3-319-27674-8_39
6. Ngo, T.D., Nguyen, V.H., Lam, V., Phan, S., Le, D.-D., Duong, D.A., Satoh, S.: NII-UIT: a tool for known item search by sequential pattern filtering. In: Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., O'Connor, N. (eds.) MMM 2014. LNCS, vol. 8326, pp. 419–422. Springer, Heidelberg (2014). doi:10.1007/978-3-319-04117-9_50
7. Rossetto, L., Giangreco, I., Heller, S., Tănase, C., Schuldt, H., Dupont, S., Seddati, O., Sezgin, M., Altıok, O.C., Sahillioğlu, Y.: IMOTION – searching for video sequences using multi-shot sketch queries. In: Tian, Q., Sebe, N., Qi, G.-J., Huet, B., Hong, R., Liu, X. (eds.) MMM 2016. LNCS, vol. 9517, pp. 377–382. Springer, Heidelberg (2016). doi:10.1007/978-3-319-27674-8_36
8. Schoeffmann, K.: A user-centric media retrieval competition: the video browser showdown 2012–2014. IEEE MultiMed. **21**(4), 8–13 (2014)

Enhanced Retrieval and Browsing in the IMOTION System

Luca Rossetto[1(✉)], Ivan Giangreco[1], Claudiu Tănase[1], Heiko Schuldt[1], Stéphane Dupont[2], and Omar Seddati[2]

[1] Databases and Information Systems Research Group, Department of Mathematics and Computer Science, University of Basel, Basel, Switzerland
{luca.rossetto,ivan.giangreco,c.tanase,heiko.schuldt}@unibas.ch
[2] Research Center in Information Technologies, Université de Mons, Mons, Belgium
{stephane.dupont,omar.seddati}@umons.ac.be

Abstract. This paper presents the IMOTION system in its third version. While still focusing on sketch-based retrieval, we improved upon the semantic retrieval capabilities introduced in the previous version by adding more detectors and improving the interface for semantic query specification. In addition to previous year's system, we increase the role of features obtained from Deep Neural Networks in three areas: semantic class labels for more entry-level concepts, hidden layer activation vectors for query-by-example and 2D semantic similarity results display. The new graph-based result navigation interface further enriches the system's browsing capabilities. The updated database storage system ADAM$_{pro}$ designed from the ground up for large scale multimedia applications ensures the scalability to steadily growing collections.

1 Introduction

In this paper we introduce the 2017 version of the IMOTION system which is the third iteration (after [11,13]) of the system participating in the Video Browser Showdown [2].

We provide a brief overview of the overall architecture of the system in Sect. 2, and elaborate in greater detail on the improvements made since the previous version in Sect. 3. Section 4 concludes.

2 The IMOTION System

2.1 Overview

The IMOTION system is a sketch-based video retrieval system which supports a large variety of query paradigms, including query-by-sketch, query-by-example, query-by-motion and querying using semantic concepts. It allows to search using multiple query containers, e.g., a still image, a user-provided sketch, the specification of motion via flow fields or by denoting a semantic concept. The IMOTION system is built in a flexible and modular way and can easily be extended to support further query modes or feature extractors.

L. Amsaleg et al. (Eds.): MMM 2017, Part II, LNCS 10133, pp. 469–474, 2017.
DOI: 10.1007/978-3-319-51814-5_43

2.2 Architecture

The 2017 IMOTION system is based on the ADAM$_{pro}$ database [3] and the Cineast retrieval engine [12] which are both part of the vitrivr[1] open-source content-based multimedia retrieval stack [14]. The IMOTION system has a custom browser-based front end which communicates with the storage and retrieval back-end via a web server which also serves the static content such as videos and preview images. Figure 1 shows an overview of the architecture of the IMOTION system.

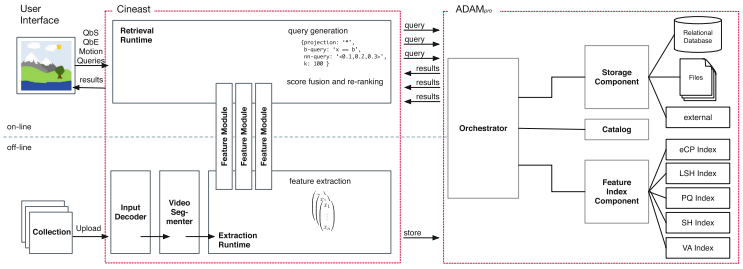

Fig. 1. Architectural overview of the IMOTION system.

3 New Functionality

3.1 Concept Detection

Since the last edition, we have expanded the set of semantic features supported by IMOTION. All these features are based on Deep Neural Network classifiers:

- We have extracted semantic categories representing entry-level labels of environments from the Places2 dataset. Classification was performed using the pre-trained VGG16-places365 network [18].
- We have trained image-level classifiers for the 80 classes of the MS COCO Detection challenge [9]. The feature data is obtained from the last fully connected layer ("fc7") of a VGG convolutional network. The model is trained on the MS COCO train2014 data and it learns the 80 labels independently using multinominal logistic regression.
- We kept the 325 semantic entry-level categories obtained from n-grams from last year [11].

Given the participation in this year's TRECVID Ad Hoc Search task[2], which also operated on the IACC.3 data, we integrated the result scores for our estimated best run into the search engine. We have extended the list of 30 AVS

[1] https://www.vitrivr.org/.
[2] http://www-nlpir.nist.gov/projects/tv2016/tv2016.html#avs.

textual queries with several queries we consider useful for browsing e.g., "shots with two people", "shots showing cartoons", etc.

As in our previous system, we use multiple ConvNets for feature extraction and object/action recognition. We replaced the temporal ConvNet trained on dense optical flow maps with ConvNets that are able to recognize visual actions that may be detected from single images. In order to train these ConvNets, we used the two databases Stanford 40 [17] with 40 categories of actions and COCO-a with 140 categories [10].

We also use a modified version of the DenseCap [7] language model (LM). We use a beam search approach in order to keep multiple results at each generated word. We hence end up with a number of alternatives sentences for each region of interest. From these sentences, we recover a set of words corresponding to objects and attributes. We also use downsampled (bilinear sampling) features extracted with DenseCap ConvNet. This ConvNet was trained on the Visual Genome [8] dataset.

3.2 Semantic Class Selection

As with the previous version of the system, one supported query mode is to search for instances of detected semantic concepts. In the 2016 IMOTION system [11] we implemented the interface for the selection of these concepts as a list of icons which could be added to a canvas via drag and drop. Figure 2 shows an example of this UI element. The new selection interface for VBS 2017, depicted in Fig. 3, uses a text box with an auto-complete feature to select semantic classes. Every class adds a weight slider by which the importance of this class with respect to the query can be specified.

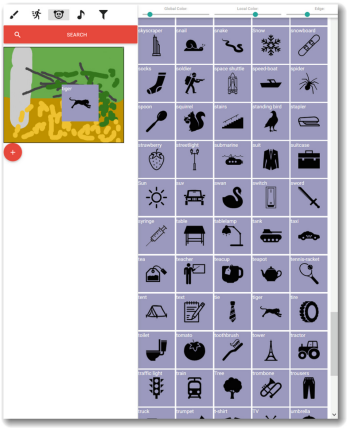

Fig. 2. Semantic class selection in the 2016 IMOTION system.

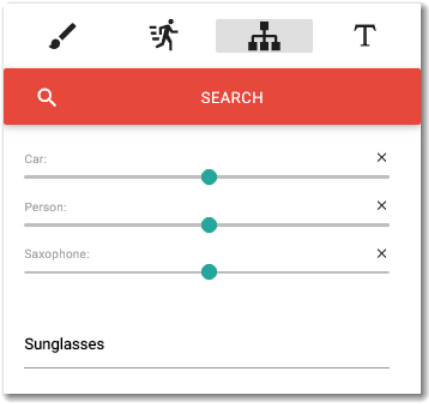

Fig. 3. New semantic class selection in the 2017 IMOTION system.

3.3 Result Presentation and Browsing

In addition to the existing querying capabilities, for the 2017 version of the system we put additional emphasis on exploratory search and browsing capabilities. In a manner similar to several of the 2016 VBS systems (e.g., [1]), we have implemented a similarity-based navigation interface. The new interface allows to navigate through the resulting grid by panning and zooming as it places visually and semantically similar results close to each other.

3.4 Text-Based Retrieval

At VBS 2017, we use traditional text retrieval based on Lucene to search in the text extracted from the ASR (as provided with the video data), and captions extracted from the keyframes using DenseCap [7].

3.5 ADAM$_{\mathbf{pro}}$

In the most current version, the IMOTION system uses the new ADAM$_{pro}$ database. The ADAM$_{pro}$ database [3] is geared towards offering storage and retrieval capabilities for multimedia objects and the corresponding metadata. To this end, it supports both Boolean retrieval and k nearest neighbour similarity searches in the vector space retrieval model and is particularly tailored to support large multimedia collections. ADAM$_{pro}$ comes with various index structures that are very different in their nature: Locality-Sensitive Hashing [5] and Spectral Hashing [16] are hash-based methods and form together with Product Quantization [6] and extended Cluster Pruning (eCP) [4] a group of indexes which support a rather coarse retrieval which can be executed very quickly, however suffers from the

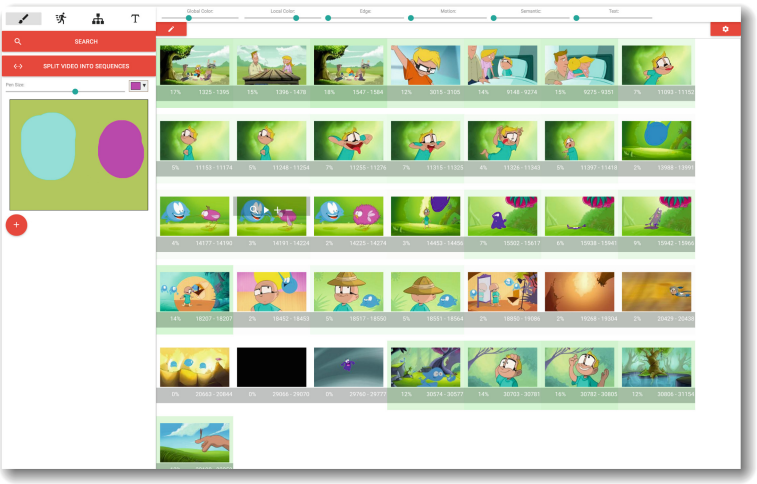

Fig. 4. Screenshot of the 2017 IMOTION system UI.

fact that it may miss result candidates as they are pruned by mistake from the candidate list. The Vector Approximation-File (VA-File) index [15], on the other hand, may degenerate to a sequential scan in worst case; however, it will not prune by mistake a true result candidate. Finally, ADAM_{pro} supports sharding a collection to multiple nodes to increase the retrieval efficiency.

4 Conclusions

The 2017 version of the IMOTION system has received significant upgrades over previous versions in both indexing and browsing (Fig. 4). Compared to last year's version, we have tripled the number of semantic classes and improved the class selection mechanism. In agreement with video browsing state of the art, the results browsing interface features semantic-based arrangement, which is supposed to significantly reduce the interaction overhead for browsing and near-hit search. Finally, the new version of IMOTION is backed up by the new ADAM_{pro} storage system, which comes with a large variety of indexing structures to decrease query latency.

Acknowledgements. This work was partly supported by the Chist-Era project IMOTION with contributions from the Belgian Fonds de la Recherche Scientifique (FNRS, contract no. R.50.02.14.F) and the Swiss National Science Foundation (SNSF, contract no. 20CH21_151571).

References

1. Barthel, K.U., Hezel, N., Mackowiak, R.: Graph-based browsing for large video collections. In: He, X., Luo, S., Tao, D., Xu, C., Yang, J., Hasan, M.A. (eds.) MMM 2015. LNCS, vol. 8936, pp. 237–242. Springer, Heidelberg (2015). doi:10.1007/978-3-319-14442-9_21
2. 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 Appl., 1–33 (2016). doi:10.1007/s11042-016-3661-2
3. Giangreco, I., Schuldt, H.: ADAMpro: database support for big multimedia retrieval. Datenbank-Spektrum **16**(1), 17–26 (2016)
4. Gudmundsson, G., Jónsson, B., Amsaleg, L.: A large-scale performance study of cluster-based high-dimensional indexing. In: Proceedings of the International Workshop on Very-Large-Scale Multimedia Corpus, Mining and Retrieval (VLS-MCMR 2010), Firenze, Italy, pp. 31–36. ACM (2010)
5. Indyk, P., Motwani, R.: Approximate nearest neighbors: towards removing the curse of dimensionality. In: Proceedings of the Symposium on the Theory of Computing, Dallas, Texas, USA, pp. 604–613. ACM (1998)
6. Jegou, H., Douze, M., Schmid, C.: Product quantization for nearest neighbor search. IEEE Trans. Pattern Anal. Mach. Intell. **33**(1), 117–128 (2011)
7. Johnson, J., Karpathy, A., Fei-Fei, L.: Densecap: fully convolutional localization networks for dense captioning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)

8. Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.-J., Shamma, D.A., et al.: Visual genome: connecting language and vision using crowdsourced dense image annotations. arXiv preprint arXiv:1602.07332 (2016)
9. Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P.: Microsoft COCO: common objects in context. ArXiv e-prints, May 2014
10. Ronchi, M.R., Perona, P.: Describing common human visual actions in images. In: Jones, M.W., Xie, X., Tam, G.K.L. (eds.) Proceedings of the British Machine Vision Conference (BMVC 2015), pp. 1–12. BMVA Press, Norwich (2015)
11. Rossetto, L., et al.: IMOTION – searching for video sequences using multi-shot sketch queries. In: Tian, Q., Sebe, N., Qi, G.-J., Huet, B., Hong, R., Liu, X. (eds.) MMM 2016. LNCS, vol. 9517, pp. 377–382. Springer, Heidelberg (2016). doi:10.1007/978-3-319-27674-8_36
12. Rossetto, L., Giangreco, I., Schuldt, H.: Cineast: a multi-feature sketch-based video retrieval engine. In: 2014 IEEE International Symposium on Multimedia (ISM), pp. 18–23. IEEE (2014)
13. Rossetto, L., Giangreco, I., Schuldt, H., Dupont, S., Seddati, O., Sezgin, M., Sahillioğlu, Y.: IMOTION — a content-based video retrieval engine. In: He, X., Luo, S., Tao, D., Xu, C., Yang, J., Hasan, M.A. (eds.) MMM 2015. LNCS, vol. 8936, pp. 255–260. Springer, Heidelberg (2015). doi:10.1007/978-3-319-14442-9_24
14. Rossetto, L., Giangreco, I., Tanase, C., Schuldt, H.: vitrivr: a flexible retrieval stack supporting multiple query modes for searching in multimedia collections. In: Proceedings of the 2016 ACM on Multimedia Conference, pp. 1183–1186. ACM (2016)
15. Weber, R., Schek, H.-J., Blott, S.: A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In: Proceedings of the International Conference on Very Large Data Bases (VLDB 1998), New York, USA, pp. 194–205 (1998)
16. Weiss, Y., Torralba, A., Fergus, R.: Spectral hashing. In: Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS 2008), Vancouver, Canada, pp. 1753–1760 (2008)
17. Yao, B., Jiang, X., Khosla, A., Lin, A.L., Guibas, L., Fei-Fei, L.: Human action recognition by learning bases of action attributes and parts. In: 2011 International Conference on Computer Vision, pp. 1331–1338. IEEE (2011)
18. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: Advances in Neural Information Processing Systems, pp. 487–495 (2014)