Smarter Video with ProCAMS Blepo Marketplace

Johanna Björklund, Marina Kolesnik, Sigrid Lindholm, Emil Lundh, Patrik Löfgren, Markus Risberg, Jonas Sandberg, Eniko Szakasc, Urban Söderberg

Abstract

Automatic metadata annotation is in great demand in the digital market and is currently transforming the publishing and broadcasting industries. We present Blepo, a market place for Intelligent Video Analytics, and describe two usecases, one related to machine-supported face clustering and one to mood recognition.

Index Terms: video analysis, digital marketplaces, face clustering, mood recognition

1. Introduction

Video continues to gain in importance as a marketing and customer-relations channel. By 2018, approximately 80 per cent of global IP traffic is predicted to be video. To manage these kinds of volumes, accurate and automatically obtainable metadata is needed. ProCAMS, which is short for Promoting Creativeness in Augmented Video Services, is funded within the Horizon 2020 programme under agreement number 644460. The project serves to realize the online repository blepo.net for the distribution and procurement of intelligent video analytic (IVA) modules. A guiding principle is that users should be able to select, test and solicit IVAs based on their own requirements and type of content. To identify core IVAs and demonstrate the added value of the repository, a number of SMEs have been invited to contribute real-world usecase. Codemill participates with two software tools, one for machine-supported indexing of content with faces, and one for collecting customer feedback through the combination of video input and mood recognition.

2. The Bleopo Marketplace

ProCAMS uses the Blepo Marketplace to promote the exchange of specialized video analytics applications, so as to enrich the creative video content production. With eight international partner companies contributing to its operation, Blepo has been an international project from the start. The IVA market is fragmented – there are developers creating excellent software, and video content producers looking for content enhancement and analytical solutions, and yet there is no central place that brings these two together. Blepo is intended to be that place – Blepo marketplace aims to expand the reach of the IVA market in a transparent way, enabling developers and producers in the creation of world class video content.

Content providers as end users of Blepo services deal with a wide variety of IVA applications. Blepo facilitates the work of content creators by i) testing different IVA solutions against own videos, and ii) finding an ultimate IVA solution tailored best to their technical needs. From system level point of view, transparent operation, searchability, data storage are very important, and might be facilitated by IVAs. Business opportunities including marketing potential, interactive content and other commercial purposes are clearly engaging, as well. Many of the content providers are coming from the creative and media industry, such as television, advertising or visual arts. The size of only the television market in the EU compared to other creative industries is impressive; the turnover is around 90 billion euros. Presence on Blepo allows developers to test their creativeness and their innovative skills among other industry players. They get an insight into what international companies work on or what solutions of certain challenges are interesting for them. They can take it as a benchmark or some kind of quality check.

In summary, Blepo meets the market's needs by

1. creating a base for professional networks for content providers and developers; opportunity to get involved in the life of a diverse professional community
2. providing wider business development opportunities than before
   (a) for video content providers: to find innovative solutions available on the market or solicit customized solutions
   (b) for developers: to find new customers,to offer their IVA or simply to get industry benchmarks or quality check of their work.

3. Face clustering

Codemill has the privilege to work with some of the world's best known publishers and broadcasters, and we see a broad interest in content-based video search and recommendation. Highest priority is typically given to automatic recognition of speech and faces. Within ProCAMS, we developed a tool for face-based content indexing as part of a greater media asset management system. Its purpose is to mark up a video archive with respect to a large number of faces. Traditional face recognition requires the user to provide an image gallery of target persons. This makes a closed-world assumption that is reasonable for, e.g., produced tv series where the cast is known, but not

for news casts and user-generated content. For this reason, our tool detects and clusters unknown faces, and then lets the user identify them as needed in a post-processing step.

The central task is as follows: From a video containing one or more different individuals, detect and group the faces that appear. The detected faces should not be matched against an input gallery, but instead, the algorithm must decide what occurrences map to the same individual. Each detected occurrence of a face has an associated time span. The time span ends when the face disappears from view, and a new time span begins if the face reappears. All time spans belonging to the same individual are grouped together in the output data. Two or more time spans, associated with different individual faces, can thus be fully or partially overlapping. The output also contains time codes of representative frames from which the different faces can be extracted.

Face detection and recognition is by now a mature research field and there is no shortage of algorithms. For the current application, one should distinguish between the two tasks of detecting and comparing faces. First, for detecting faces, a wealth of techniques have been developed, the most popular being the Viola-Jones classification using Haar cascades [1]; implementations of this type of algorithm are available in open-source libraries such as OpenCV (opencv.org). Recent developments in deep neural networks, however, are rapidly changing the game: Neural networks are now beating the state of the art in most computer-vision subfields, and for example, form the foundation of Googles online service Cloud Vision (cloud.google.com/vision). Comparing faces is different from detection, and requires its own set of tools; again, an extensive literature has been developed and several algorithms are in use today. One class of matching tools build on plain image-matching techniques such as Speeded Up Robust Features (SURF) [2] and Scale-Invariant Features (SIFT) [3], while others are tailor-made for comparing faces, making use of facial features or building on facial-recognition techniques.

In this project, the IVA is realised with the Luxand Face SDK. The face clustering project and IVA is described in greater detail at blepo.net/faceclustering. After metadata about faces has been extracted, it is available for review through a web-based frontend. The IVA may for instance hava failed to understand that two clusters of faces belong to the same individual, in which case there is an option to merge the clusters. Similarly, clusters can be created, deleted, split, and named. After the review, the metadata can be used for searching, querying, and recommending content.

4. Mood detection

The mood surveyor is part of Smart Video, a cloud-based adtech software that increases online sales. It does so by increasing traffic, conversion, and by bridging the gap between physical and virtual store. The mood surveyor is useful in the last respect, as it gauges and collects customer feedback through the analysis of facial expressions. This technology can be used in both virtual and physical stores, but it is also applicable on existing video archives to make video searchable by emotion.

The mood surveyor is divided into a frontend, a backend, and a database layer. The frontend layer integrates with a camera and continuously sends captured frames to the backend layer for analysis. Once this is complete, the resulting metadata is stored in the database and a set of mood triggers is returned to the frontend. The triggers decide what feedback the system will give to the user, for example, an animation of a happy face or a little tune.

The mood surveyor can be used in stores and public places to collect customer feedback, for example on the service in a restaurant, or how well a public bathroom is cleaned. One of the compelling ideas behind the mood surveyor is to give feedback without having to touch anything, for example a written form, and this can lead to a higher user response count especially in public places such as bathrooms. The feedback provided could even be connected on social medias to give a customer rating based on the number of smiles or frowns.

The video analysis service is based on the well-known SHORE algorithm (short for Sophisticated High-speed Object Recognition Engine) [4], here implemented by Fraunhofer and acquired through the Blepo Marketplace. SHORE combines structure-based features and learning algorithms and is capable of recognising 4 different moods. Input is given in the form of video in the WebM format, and the output is metadata describing the mood detected and the frames where it was recognized.

5. Conclusion

Blepo allows its users to evaluate all uploaded IVAs on their own content to see what professional challenges can be solved through them. It is also possible to socilicty new research and developement work on particular tasks, to obtain custom solutions. In summary, we believe that the Blepo repository fills a current gap in the digital ecosystem, and we therefore encourage other SMEs in video technology to join the effort.

6. References

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. 511–518.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[3] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 1999, p. 1150.

[4] T. Ruf, A. Ernst, and C. Küblbeck, *Face Detection with the Sophisticated High-speed Object Recognition Engine (SHORE)*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 243–252.