FAME video browser – face recognition based metadata generation for performing art videos

Kenzo Milleville[1], Alec Van den broeck[1], Rony Vissers[2], Bart Magnus[2], Nastasia Vanderperren[2], Astrid Vergauwe[2], Ellen Van Keer[2], Tom Ruette[3], Steven Verstockt[1]

[1] Ghent University – imec, IDLab, AA Tower, Technologiepark-Zwijnaarde 122, 9052 Gent, Belgium
[2] meemoo, Flemish Institute for Archives, Ham 175, 9000 Gent, Belgium
[3] Flanders Arts Institute, Ravensteingalerij 38, 1000 Bruxelles, Belgium

This demonstrator focuses on one of the end results of the FAME project [1]. The main focus of this project is the development of a generic open-source face recognition pipeline that can be applied on a broad variety of cultural heritage video/image archives, the performing arts collection of the Flanders Arts Institute [2] being one of them. Performing arts images/videos are among the most challenging types of video for face recognition given their dynamic and heterogeneous nature and the variety of domains they span. If it will work on this type of content, it will most probably also work on other types of archived content.

Due to the lack of accurate shot-level metadata, the searchability of performing art videos is still rather limited. Mostly they can only be queried based on some global metadata. Retrieving the exact shot(s) where an actor or dancer appears in the video, for example, is currently impossible. The shots can of course be manually annotated, but this is an error-prone and labour-intensive process, definitely when it needs to be performed on a large collection. Within the FAME project, we address this issue and proposes a video content annotation tool for the automatic annotation of faces in performing art videos. The resulting JSON metadata (which contains the Wikidata IDs of the recognized people and the timestamps when they appear in the video) is used in a video browser web application which allows to query the video based on the recognized performers and find the shots they appear in. A screenshot of the timeline based FAME video browser – which is built in React (A JavaScript library for building user interfaces) - is shown in Figure 1.

The automatic generation of fine-grained facial metadata for each video sequence is based on shot detection, face detection/recognition and context based filtering. For the shot detection, the Adaptive Detector [3] of PySceneDetect (version 0.5.6.1) generates the best results over all types of investigated content. Further optimization is possible by grouping shots with high semantic similarity, as is done in the logical story unit detection work of Kumar et al. [4]. Once all shots have been detected, we randomly select 3 frames within the shot. Those will be used to do the face detection/recognition. A subset of the selected frames for the Ivanov play by theatre collective Stan (1992) are shown in Figure 2. As can be seen, similar shots are nicely grouped together.

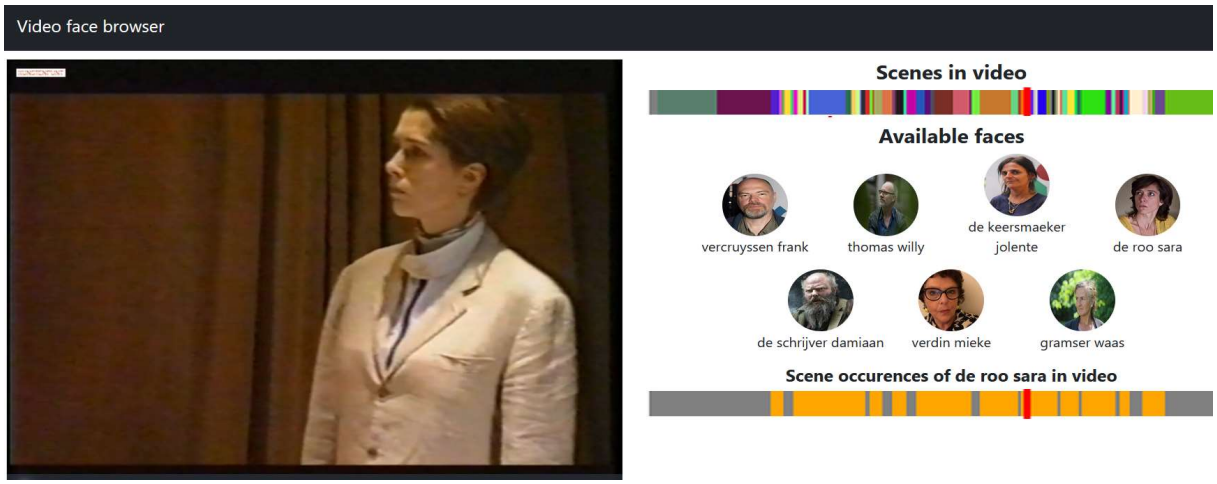


Figure 1 - Timeline based FAME video browser. For each actor, all shots he/she appears in are highlighted on the timeline. By clicking on the timeline, the video browser starts playing the video at that particular location.

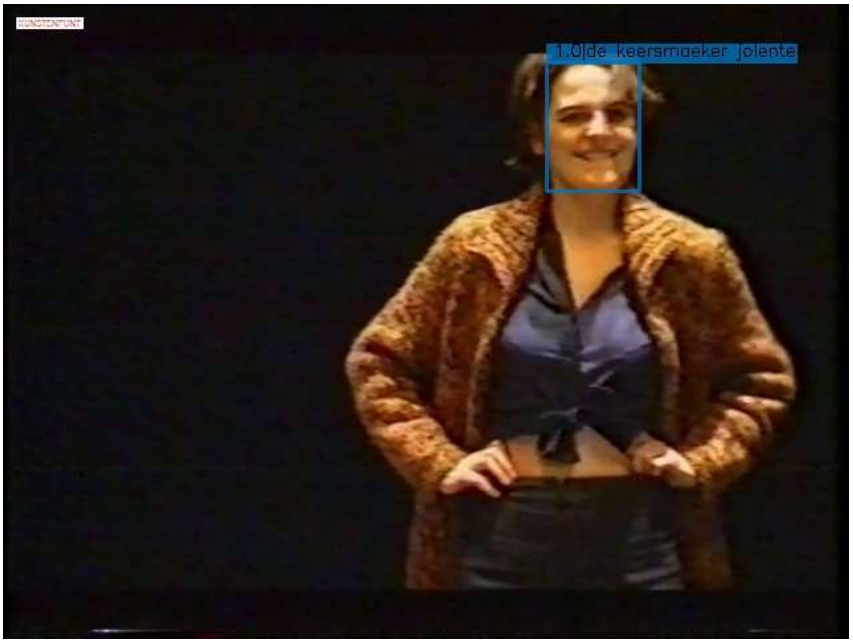
Figure 2 – Shot detection results for Ivanov (1992 – tg Stan). For each detected shot, 3 random frames are chosen on which the face detection/recognition will be performed.

The face detection (of which some examples are given in Figure 3) is based on Insightface [5], an open-source, all-in-one model with state-of-the-art performance and low computational cost that outperformed the other models in our evaluation. For each detected face, it generates a 512D feature vector, i.e., the face profile. Furthermore, it also predicts age and gender – both can be useful in the context filtering step. After extracting the profiles for each face, these can now be matched to a reference (ground truth) dataset of known performers (2252 persons, 28 397 validated images) using a normalized pairwise cosine similarity score. Some examples of faces that were recognized in the Ivanov play are shown in Figure 3. A higher recognition score denotes a more similar match to the reference images of that actor. Side angles, worn items (masks, sunglasses, etc.), beautification [6] and time differences between the video images and reference images of course can have impact on the score. How to optimally cope with these disruptors is currently been further investigated. Face pose estimation [7] and facial accessories recognition [8], for example, can probably be used to cope with the first two problems.

Based on the global metadata of the production (e.g. year, actor list, location) the list with candidate persons that is retrieved in the face recognition step can be further filtered. Furthermore, semantic relations (i.e., co-occurences of actors that appear in a particular setting) can also help to remove outliers/misdetections. Both examples of context based filtering further increase the overall accuracy of the proposed methodology.

Acknowledgements

The research activities as described in this paper were funded by Ghent University, imec, meemoo, the Belgian Science Policy Office (BELSPO), and the Flanders Department of Culture, Youth and Media.



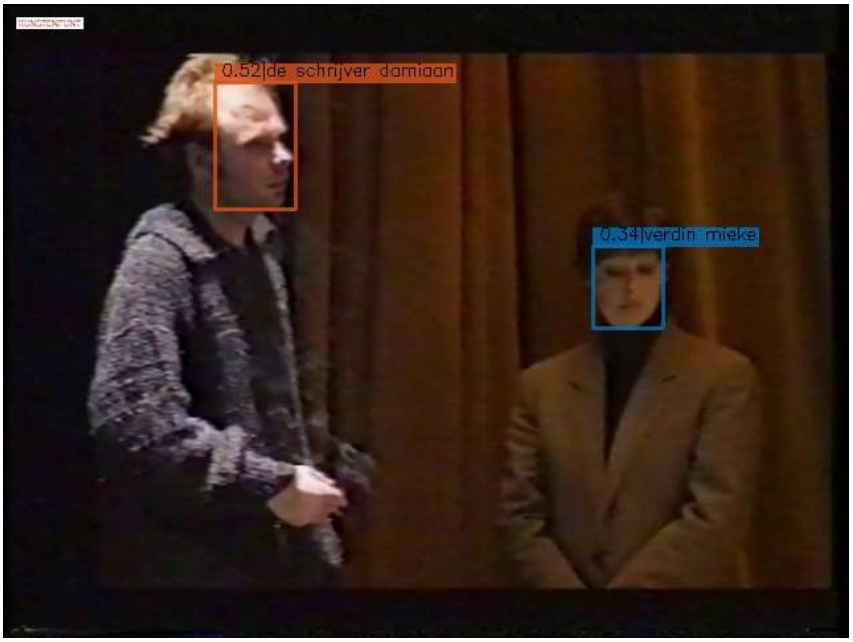


Figure 3 – InsightFace face detection/recognition results for Ivanov (1992 – tg Stan)

References

[1] meemoo (2022, 17 January), FAME: facial recognition as a tool for metadata creation, https://meemoo.be/en/projects/fame-facial-recognition-as-a-tool-for-metadata-creation

[2] Flanders Arts Institute (2022, 17 January), Expertise centre for performing arts, music and visual arts, https://www.kunsten.be/en/performing-arts/

[3] PySceneDetect (2021, 11 October), Intelligent scene cut detection and video splitting tool, https://pyscenedetect.readthedocs.io/en/latest/reference/detection-methods/

[4] Thirukokaranam Chandrasekar, Krishna Kumar, and Steven Verstockt. “Context-Based Structure Mining Methodology for Static Object Re-Identification in Broadcast Content.” APPLIED SCIENCES-BASEL, vol. 11, no. 16, 2021, doi:10.3390/app11167266.

[5] Insightface (2021, 30 November), State of the art deep face analysis library, https://insightface.ai/, https://github.com/deepinsight/insightface

[6] Impact and Detection of Facial Beautification in Face Recognition: An Overview, https://hal.inria.fr/hal-02378939/document

[7] Baltanas, Samuel-Felipe, Jose-Raul Ruiz-Sarmiento, and Javier Gonzalez-Jimenez. "Improving the Head Pose Variation Problem in Face Recognition for Mobile Robots" *Sensors* 21, no. 2: 659, 2021, doi:10.3390/s21020659

[8] Zeng, Dan & Veldhuis, Raymond & Spreeuwers, Luuk. (2021). A survey of face recognition techniques under occlusion. IET Biometrics. 10. doi:10.1049/bme2.12029