2022 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-9620-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIP46576.2022.9897300

HISTORIAN: A LARGE-SCALE HISTORICAL FILM DATASET WITH CINEMATOGRAPHIC ANNOTATION

Daniel Helm, Fabian Jogl, Martin Kampel

TU Wien, Institute of Visual Computing & Human-Centered Technology,
Computer Vision Lab Austria, Vienna

ABSTRACT

Developing automated tools for sustainable film preservation of extensive historical film collections assumes an understanding of fundamental cinematographic settings. In order to be able to investigate new approaches to detect and classify cinematographic settings, this paper proposes a novel large-scale historical film dataset with cinematographic annotations (HISTORIAN), i.e., shot boundaries, shot types, camera movements. The dataset consists of 98 digitized original analog film reels related to the Second World War and 10593 film shots manually annotated with human film experts. Moreover, annotations for overscan areas such as sprocket holes are included. A baseline film analysis pipeline is introduced and evaluated. To the best of our knowledge, HISTORIAN is the first dataset that covers the challenges and characteristics of historical film documentaries and provides novel possibilities for exploring automatic film analysis tools.

***Index Terms*—** Historical Film Dataset, Film Archives, Deep Learning, Automated Film Analysis, Cinematographic Data

1. INTRODUCTION

Automatic video analysis is an ongoing research field including different subdomains such as scene understanding [1, 2] or cinematographic style analysis [3]. In order to deal with those topics, a number of datasets [4–6] have been published. They include mainly short sequences or single frames extracted from short video clips. In comparison, *professionally produced films* such as modern Hollywood productions or documentaries are not created by just recording one specific situation [7,8]. In fact, they consist of a complex film hierarchy [9] and are produced by following a lengthy editing and recording process. The smallest unit of a professional film is the *Shot*. One shot is a continuously recorded situation characterized by specific cinematographic settings. One fundamental cinematographic setting is the *Shot Type*, which is a representation of the distance between the subject of interest and the camera lens [4]. Camera movements like pans, tilts, zooms or tracks [10] are another example of cinematographic properties used in professional filmmaking. Recent research [2, 5] is mainly focused on modern feature films (e.g., Forest Gump, Titanic) while historical documentaries have been a less considered area of research [11, 12]. Feature films show highly synthetically generated recordings, while historical documentaries illustrate real-world situations. Historical films comprise different challenges due to the quality and general frame behavior [7, 11, 12]. However, knowledge of cinematographic settings is crucial for sustainable film preservation and any expert working on film archival tools. Detecting cinematographic settings allows film experts to find specific shots in large film collections needed to create new visual representations. Meaningful datasets, including authentic data sources and their annotations, are crucial to developing automated film analysis tools for large historical film collections.

This paper proposes a novel dataset called HISTORIcal Film Dataset with Cinematographic ANnotation (HISTORIAN). This dataset includes historical film documentaries recorded during the Second World War. Overall, *98 film documentaries* have been used to provide 10593 annotated shots, 838 camera movements, and 192 frames with annotated overscan artifacts. HISTORIAN captures novel challenges and characteristics of historical footage, including complex real-world situations compared to well-known benchmark datasets presented in Section 2. The annotation process collaborates with human annotators (film and computer vision experts) to provide a meaningful dataset to the research community. Additionally, an automated film analysis pipeline is presented to provide a baseline for the automated detection and classification of cinematographic settings. Finally, the entire dataset, including annotations as well as film sources and the reference implementation of the baseline film analysis pipeline, is published on Zenodo [13] and Github [14].

This paper is structured as follows: Section 2 gives an overview of state-of-the-art datasets. The properties of HISTORIAN and details about the annotation process are presented in Section 3. In Section 4 and 5, the baseline film analysis pipeline and results are demonstrated and discussed. Finally, Section 6 concludes our investigation with a summary and outlook for future research.

2. RELATED WORK

Hassanien et al. [15] and Tang et al. [16] work on automatic shot boundary detection approaches to detect abrupt and gradual transitions. Based on video sequences gathered from TRECVID [6], RAI [17], Youtube or Weibo, they have published the datasets **DeepSBD** [15] and **Clipshots** [16]. A further dataset, called **Cinescale** [4] provides a fundamental base for exploring cinematographic shot types [3] in professional film recordings. It contains a massive number of frame-based labels (792k) that have been extracted from 124 art movies (made between 1949 and 2013). The recently published dataset, **Movienet** [5], focuses on modern film productions. It contains massive annotations of cinematographic settings such as shot boundaries, shot types, and camera movements [10, 18] for entire movies and trailers. The dataset is usable in different research areas such as automatic detection of actor identification [19], movie synopsis analysis [1] or scene segmentation [20]. However, accessibility to the original movies and trailers is currently restricted, and

Visual History of the Holocaust: Rethinking Curation in the Digital Age. This project has received funding from the EU's Horizon 2020 research and innovation program under the Grant Agreement 822670. Special thanks to Jakob Zenzmaier (Key Researcher - Ludwig Boltzmann Institute for Digital History, Vienna).

only the corresponding film annotations are published. All of the datasets mentioned above are published to promote research on automatic film analysis. However, copyright constraints may prevent researchers from obtaining full access or using the videos to analyze proposed methodologies. Public datasets such as **MovieGraphs** [2] provide short annotated clips, selected to demonstrate exactly the situation the authors want to show, for example, scenes with certain kinds of human-centric interactions. However, films can consist of several concatenated shots. Thus, significant information which counts to a cinematographic shot can be lost. Additionally, most of the previously mentioned datasets contain mainly professional feature films. Less research focus is put towards **digitized original historical film** [11, 12, 21]. Similar to [4], HistShotDS [21], contains frame-based shot type annotations of selected frames. However, this dataset lacks due to the limited number of available samples as well as the missing temporal shot information. Historical film documentaries provide unique challenges, such as damaged film reels [8, 12], or overscan areas [8,22] (e.g. sprocket holes, frame lines). Thus, current public datasets are insufficient for developing sustainable film analysis tools for large historical film collections.

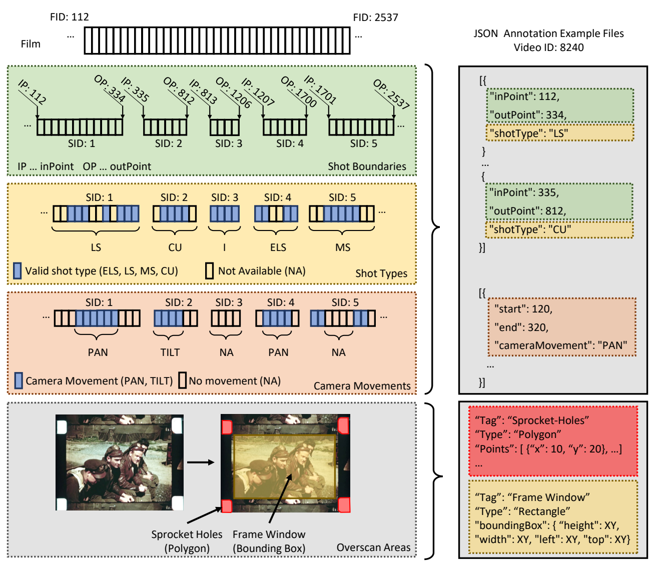

Fig. 1: This Figure illustrates a schematic overview of the annotation process and the resulting JSON files.

3. HISTORICAL FILM DATASET WITH CINEMATOGRAPHIC ANNOTATION (HISTORIAN)

HISTORIAN contains annotations for different cinematographic settings which are shot boundaries [15, 16], shot types [4], and camera movements [10, 18]. The focus of our dataset is on Abrupt Transitions (AT - shot boundaries) as well as the shot type categories: Extreme Long Shot (ELS), Long Shot (LS), Medium Shot (MS), Close Up (CU), Intertitle (I) and Not Available/Not Clear (NA). Compared to state-of-the-art datasets like Movienet [5], or Cinescale [4] the selected categories are slightly different due to the different usage of cinematographic recording techniques and the general film behavior. However, a new category, called Intertitle, is introduced in HISTORIAN. This class includes recordings with text-based information (e.g., inserted name of a person, location signs, etc.). State-of-the-art approaches that are working on camera movement classification [10, 18] are mainly focused on pans, tilts, tracks, zooms, and no-movements. Our investigation focuses on standard movements such as pans, tilts, dollies, trucks, tracks, pedestals, and zooms. Additionally, sprocket holes and the frame window of several analog film formats (16mm and 35mm) are marked for cropping the overscan areas [8, 22]. The dataset includes 98 annotated historical film documentaries (1440 by 1080 pixels, 24fps). All recordings are related to the liberation phase of Nazi concentration camps during the Second World War (1943-1945) and document the situation in concentration camps[1]. The original analog film reels are collected and digitized within the Horizon 2020 project Visual History of the Holocaust (VHH)[2] in collaboration with the National Archive and Records Administration (NARA)[3].

Compared to benchmark datasets [4, 5, 21], HISTORIAN includes the entire unprocessed film sources. Moreover, it captures the challenges and characteristics of historical film documentaries. Quality artifacts such as damages, motion blur or over and underexposure as well as overscan regions [8] are significant characteristics of this dataset. Another difference to benchmark datasets [4, 5] relates to the film setting. HISTORIAN includes recordings of real-world situations. Contrary to that, modern feature films are generally recorded on synthetically generated film sets. Furthermore, the proposed dataset contains shot boundary positions for all films,shot-based shot types, as well as camera movement annotations. Additionally, HISTORIAN is extended with bounding box and polygon annotations to promote research on automatic segmentation of overscan areas such as sprocket holes or the final frame window.

3.1. Annotation Process & Properties

All annotations are gathered by following a fully manual annotation procedure in several cycles involving human annotators from the computer vision and film domain. A schematic illustration of the annotation process is given in Figure 1. The entire dataset can be found on Zenodo [13].

Shots boundaries are marked while stepping through the entire film. The exact start and stop frames are collected if a shot change occurs. The manual annotation process points out that mainly Abrupt Transitions (AT) occurs instead of Gradual Transitions (GT) such as fades, wipes, or dissolves [12]. The reason for that observation is that historical film documentaries are mainly unprocessed and raw films compared to modern film productions. However, the shot boundaries are selected to crop the gradual transition as accurately as possible from the core content.

For each shot, the corresponding **shot type** category is added. There exist no clear borders between consecutive categories such as CU-MS or ELS-LS or LS-MS, which makes shot type annotation not trivial and gives human annotators room for interpretation [21]. Furthermore, shots can start with a camera movement (e.g., tilt-pan) until the subject of interest is in the focus of the camera lens. Therefore, human annotators have to focus on the center part of a frame (spatial) and the middle part of the entire shot related to the time (temporal). In order to filter the significant subject of interest (e.g., person, group of persons, building, etc.), annotators use the ratio between the subject size and the frame size to distinguish between different shot types inspired by [23]. The negative class, Not Clear/Not Available (NA), is used if no decision can be made, e.g., black frames or massive quality restrictions.

[1] Viewer discretion is advised as the visual material contains graphic depictions of atrocities and/or their aftermath.

[2] https://www.vhh-project.eu/ - last accessed: 11/02/2022

[3] https://www.archives.gov/ - last accessed 28/01/2022

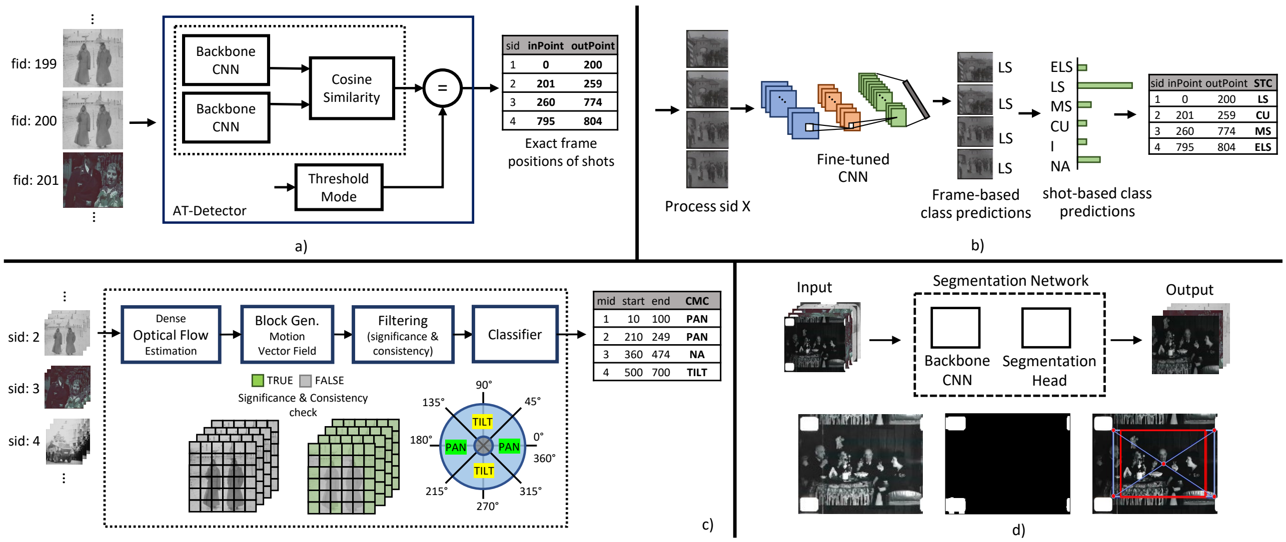

Fig. 2: Schematic illustration of (a) Shot Boundary Detection module (SBD) with dynamic threshold, (b) Shot Type Classification (STC) with shot-based shot type predictions using the final class distribution, (c) Camera Movements Classification (CMC) based on dense optical flow estimation and (d) Overscan Detection (OSD) with the segmentation network DeepLabV3.

A manual assessment points out that one shot may include multiple **camera movements**. Therefore, sequence-based tagging provides more meaningful annotation results instead of classifying the entire shot with one unique movement category. The human annotators step through the entire film during the annotation process and mark start and stop-frame positions of an identified camera movement. Finally, information about the exact frame-based and shot-based location of a camera movement is provided. Thus, a user can directly work on sequences including specified motion categories or explore motion activities on a shot-based level.

The **overscan artifacts** are masked to provide researchers the possibility to work with both cropped and uncropped films. Researchers may also be interested in working on automatic overscan detection tasks [22]. The annotation procedure is done manually by using the Visual Object Tagging Tool (VOTT)[4]. Polygons are drawn to overlap as accurately as possible with the Sprocket Holes (SH). The bounding box coordinates for the frame window are selected by considering the *most inner points* of the SHs for 16mm film reels, while the *most outer point* of the SHs are selected for 35mm reels [22]. For all films showing overscan artifacts, one annotation block containing the segmented sprocket holes (Polygons) as well as the final frame window (Bounding Box) is added. An empirical assessment of the recordings shows that one annotation block per film is enough due to the stable scan conditions. Despite this, the center frames are extracted and annotated for eight randomly selected shots (per film).

3.2. Dataset Summary

HISTORIAN contains 98 films, with over 17 hours and about 1.5 million frames. The films have a length between 4 and 32 minutes, with an average of 11 minutes. Most films run for less than 17 minutes, with a few outliers that run significantly longer. The proposed dataset includes 10593 shot annotations with the corresponding shot type. Figure 3 (right) depicts the distribution of the shot type categories. Furthermore, the dataset contains 838 annotated frame sequences, including pans (302), tilts (131), tracks (273), trucks (33), dollies (76), pan-tilts (18), zooms (4), and pedestal (1). The corresponding class distribution can be observed in Figure 3 (left). Three frames of different shots are extracted for each film, representing the film reels: 16mm with black/white SHs (17) and 35mm with black/white SHs (47). In total, masked overscan regions of 192 positions of 64 different films are masked.

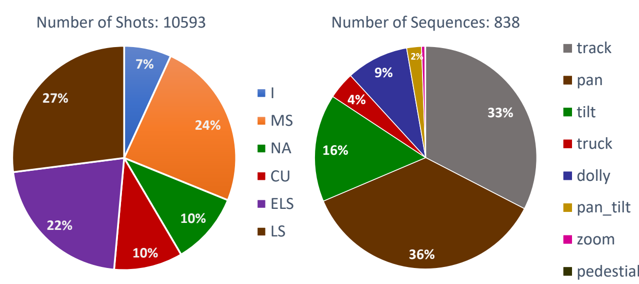

Fig. 3: Overview of the number of shots, their duration and their shot types. The shot boundaries and types were obtained from the automated pipeline.

4. BASELINE METHODOLOGY

In order to evaluate the proposed dataset, a baseline film analysis pipeline is implemented. This pipeline includes the modules: Shot Boundary Detection (SBD), Shot Type Classification (STC), Camera Movements Classification (CMC) and, Overscan Detection (OSD).

Shot Boundary Detection (SBD): The SBD module is inspired by the deep learning-based approach published by Jingwei et al. [24]. However, the focus in our investigation is on Abrupt Transitions (AT) due to the less usage of Gradual Transitions(GT) in unprocessed historical film documentaries [9]. In order to optimize run time and

[4]`https://github.com/microsoft/VoTT` - last accessed: 2022/01/27

resources, SqueezeNet [25] is used to extract meaningful visual features instead of a Resnet50 or 3D-CNN [16]. Zongjie et al. [26] show that using a sliding window in combination with an adaptive threshold mechanism results in more robust detection rates of challenging ATs. In our investigation, a similar strategy using two thresholds to optimize the sensitivity of the AT detector is followed, while the parameters are selected empirically. Details about the approach and thresholds are given in the reference implementation on Github [14]. The final output of this module is the exact frame position where a shot starts (inPoint) and ends (outPoint). A schematic visualization of the shot boundary detection pipeline is shown in Figure 2a.

Shot Type Classification (STC): This module consists of a standard CNN classification model (VGG16) to predict frame-based shot types (inspired by [4]). Compared to Huang et al. [5], shot-based annotations are gathered by predicting the majority class of the frame-based shot type results within one shot, instead of using 3D Convnets [5]. Figure 2b illustrates the pipeline in a schematic visualization. For details about the training procedure, the implementation, and pre-trained models, refer to the Github repository [14].

Camera Movements Classification (CMC): The CMC module is implemented by using the reference implementation of dense optical flow estimation from OpenCV[5]. Camera movements like pan, tilt, or no-movement are derived based on the calculated pixel-based motion vector fields. A significance and consistency check, introduced by [10], is used to filter motion vectors of interest by calculating block-based displacements in combination with a temporal sliding window. Empirically, a valid combination of all necessary parameters and thresholds for the sliding window approach [10] are gathered. In order to distinguish between interesting movements, the filtered angles are quantified to four bins in the range of [0°-360°] [10,18]. In Figure 2c a schematic illustration of the CMC module is shown. A reference implementation of this module, including details about the empirically selected parameters and threshold, are given in the repository [14].

Overscan Detection (OSD): In order to evaluate the overscan region annotations, a standard segmentation network (DeeplabV3) is set up to segment sprocket holes in digitized analog films, including overscan areas (inspired by [22]). An overview of the pipeline is illustrated in Figure 2d, and details can be found in the Github repository [14].

5. BASELINE EVALUATION

An evaluation process of the automatic film analysis pipeline is conducted to assess the proposed dataset.

For the **Shot Boundary Detection** module, the automatically predicted shot boundaries of 98 films are compared with the human-annotated results. While 9547 shots are automatically detected, 10593 shots are marked by the human annotators. The SBD evaluation is focused on the exact overlap of inPoint and outPoint of a shot instead of using Intersection over Union (IoU) [15] which has a significant influence on the detection performance. In our evaluation, the harmonic mean of Precision and Recall, $F_{1,Score} = 87\%$, demonstrates the results of the introduced SBD module (see Table 1). Observe that the SBD predictions have an accuracy of about 99% since significantly more frame pairs without ATs exist than with ATs.

The **Shot Type Classification** evaluation is done by exploring the shot-based shot type categories based on the results of the manual SBD annotation. Thus, the automatically created results of the SBD module do not influence the performance evaluation of the STC module. In order to train the CNN classifier, the dataset is split into a training (80%), validation (10%), and test set (10%). The training and validation set is generated by extracting the center frame as a representative of each shot (inspired by [21]). Table 1 illustrates the classification performance, which shows an overall accuracy of 74.5% on the final test set (shot-based results).

Since the focus of the **Camera Movements Classification (CMC)** module is on classifying the camera movements, pan, and tilt, all manually annotated sequences including the corresponding categories are extracted from the entire dataset. Additionally, shots including no movements (NA) are extracted and added to the test set in order to show that the baseline implementation is able to distinguish between motion activity and stationary. Finally, the test set includes 302 pans, 131 tilts, and 83 NA's. The classification performance of the CMC module demonstrates a $F_{1,Score}$ of 76% (see Tab.1).

Table 1: Classification results for Shot Boundary Detection (SBD), Shot Type Classification (STC) and Camera Movements Classification (CMC).

	Precision	Recall	$F_{1,Score}$	Accuracy
SBD	0.92	0.83	0.87	0.99
STC	0.76	0.75	0.75	0.75
CMC	0.78	0.75	0.76	0.82

The **Overscan Detection (OD)** module consists of a standard segmentation approach containing a trained backbone CNN as well as a segmentation head. In this investigation, a DeepLabV3 in combination with a Resnet101 is explored. The model is trained on the annotated samples provided in HISTORIAN and tested on a separate test set, including frames with synthetically generated sprocket holes (inspired by [22]). The overall results are evaluated by analyzing the Intersection over Union (IoU) score and the Dice coefficient, which gives an understanding of how accurate the model can segment sprocket holes on a pixel-based level. In this investigation, an IoU score of 80.8% and a Dice Coefficient of 88.1% is reached on the test set.

6. CONCLUSION

In this paper, a novel film shot dataset is presented and published on Zenodo [13] and Github [14]. HISTORIAN contains cinematographic annotations such as shot boundaries, shot type categories, and camera movements. Moreover, this dataset includes annotations to mask overscan areas such as sprocket holes and the core frame window. Compared to state-of-the-art benchmark datasets, our dataset provides a digitized analog film collection related to the Second World War as well as the corresponding annotations of the raw films. Furthermore, this dataset captures the challenges and characteristics of original historical footage. The data sampling and annotation process are done manually involving human annotators of the computer vision and film domain. An automatic film analysis pipeline is demonstrated and evaluated to provide a baseline. The published dataset provides a fundamental base for further research on automatic approaches and archival tools for historical films. Finally, the dataset can be extended in different ways, such as object detection (specific time-related objects) or relation detection (long-term relations in large historical film collections).

[5] `https://docs.opencv.org/4.x/d4/dee/tutorial_optical_flow.html` - last accessed: 11/02/2022

7. REFERENCES

[1] Yu Xiong, Qingqiu Huang, Lingfeng Guo, Hang Zhou, Bolei Zhou, and Dahua Lin, "A graph-based framework to bridge movies and synopses," in *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[2] Paul Vicol, Makarand Tapaswi, Lluis Castrejon, and Sanja Fidler, "MovieGraphs: Towards Understanding Human-Centric Situations from Videos," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, December 2017.

[3] Luca Canini, Sergio Benini, and Riccardo Leonardi, "Classifying cinematographic shot types," *Multimedia Tools and Applications*, vol. 62, no. 1, pp. 51–73, 2013.

[4] Mattia Savardi, András Bálint Kovács, Alberto Signoroni, and Sergio Benini, "CineScale: A dataset of cinematic shot scale in movies," *Data in Brief*, vol. 36, pp. 107002, 2021.

[5] Qingqiu Huang, Yu Xiong, Anyi Rao, Jiaze Wang, and Dahua Lin, "Movienet: A holistic dataset for movie understanding," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12349 LNCS, pp. 709–727, 2020.

[6] Alan F. Smeaton, Paul Over, and Aiden R. Doherty, "Video shot boundary detection: Seven years of trecvid activity," *Computer Vision and Image Understanding*, vol. 114, no. 4, pp. 411–418, 2010.

[7] Giovanna Fossati, *From Grain to Pixel - The Archival Life of Film in Transition*, Amsterdam University Press, Amsterdam, 2018.

[8] Barbara Flückiger, David Pfluger, Giorgio Trumpy, Tunç Aydin, and Aljoscha Smolic, "Film material-scanner interaction," Tech. Rep., University of Zurich, Zurich, February 2018.

[9] Daniel Helm and Martin Kampel, "Shot boundary detection for automatic video analysis of historical films," in *New Trends in Image Analysis and Processing – ICIAP 2019*, Marco Cristani, Andrea Prati, Oswald Lanz, Stefano Messelodi, and Nicu Sebe, Eds., Cham, 2019, pp. 137–147, Springer International Publishing.

[10] Muhammad Hasan, Min Xu, Xiangjian He, and Yi Wang, "A camera motion histogram descriptor for video shot classification," *Multimedia Tools and Applications*, vol. 74, pp. 1–26, August 2014.

[11] Matthias Zeppelzauer, Dalibor Mitrovic, and Christian Breiteneder, "Analysis of historical artistic documentaries," in *Ninth International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2008, Klagenfurt, Austria, May 7-9, 2008*, Klagenfurt, 2008, pp. 201–206, IEEE Computer Society.

[12] Markus Seidl, Matthias Zeppelzauer, Dalibor Mitrović, and Christian Breiteneder, "Gradual transition detection in historic film material—a systematic study," *Journal on Computing and Cultural Heritage*, vol. 4, no. 3, pp. 1–18, 2011.

[13] Daniel Helm, Fabian Jogl, and Martin Kampel, "Dataset: HISTORIAN," `https://doi.org/10.5281/zenodo.6644516`, 2022, [Online; last accessed 2022/06/29].

[14] Daniel Helm, Fabian Jogl, and Martin Kampel, "Sourcecode," `https://github.com/dahe-cvl/historian_dataset/`, 2022, [Online; last accessed 2022/06/23].

[15] Ahmed Hassanien, Mohamed Elgharib, Ahmed Selim, Sung-Ho Bae, Mohamed Hefeeda, and Wojciech and Matusik, "Large-scale, fast and accurate shot boundary detection through spatio-temporal convolutional neural networks," *CoRR*, vol. abs/1705.03281, 2017.

[16] Shitao Tang, Litong Feng, Zhanghui Kuang, Yimin Chen, and Wei Zhang, "Fast video shot transition localization with deep structured models," in *Computer Vision – ACCV 2018*, C. V. Jawahar, Hongdong Li, Greg Mori, and Konrad Schindler, Eds., Cham, 2019, pp. 577–592, Springer International Publishing.

[17] Lorenzo Baraldi, Costantino Grana, and Rita Cucchiara, "Shot and scene detection via hierarchical clustering for re-using broadcast video," in *Computer Analysis of Images and Patterns*, George Azzopardi and Nicolai Petkov, Eds., Cham, 2015, pp. 801–811, Springer International Publishing.

[18] Pawin Prasertsakul, Toshiaki Kondo, Hiroyuki Iida, and Teera Phatrapornnant, "Camera operation estimation from video shot using 2D motion vector histogram," *Multimedia Tools and Applications*, vol. 79, no. 25-26, pp. 17403–17426, 2020.

[19] Qingqiu Huang, Wentao Liu, and Dahua Lin, "Person search in videos with one portrait through visual and temporal links," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 425–441.

[20] Anyi Rao, Linning Xu, Yu Xiong, Guodong Xu, Qingqiu Huang, Bolei Zhou, and Dahua Lin, "A local-to-global approach to multi-modal movie scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10146–10155.

[21] Daniel Helm, Florian Kleber, and Martin Kampel, "Histshot: A shot type dataset based on historical documentation during wwii," in *Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods - ICPRAM*. INSTICC, 2022, pp. 636–643, SciTePress.

[22] Daniel Helm and Martin Kampel, "Overscan detection in digitized analog films by precise sprocket hole segmentation," in *Advances in Visual Computing*, George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Daniela Ushizima, Sek Chai, Shinjiro Sueda, Xin Lin, Aidong Lu, Daniel Thalmann, Chaoli Wang, and Panpan Xu, Eds., Cham, 2020, Springer International Publishing.

[23] Ines Cherif, Vassilios Solachidis, and Ioannis Pitas, "Shot type identification of movie content," in *2007 9th International Symposium on Signal Processing and Its Applications*, 2007, pp. 1–4.

[24] Xu Jingwei, S. Li, and X. Rong, "Shot boundary detection using convolutional neural networks," *VCIP 2016 - 30th Anniversary of Visual Communication and Image Processing*, pp. 1–4, 2017.

[25] Forrest N. Iandola, Matthew W. Moskewicz, Khalid Ashraf, Song Han, William J. Dally, and Kurt Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and ¡1mb model size," *ArXiv*, vol. abs/1602.07360, 2016.

[26] Zongjie Li, Xiabi Liu, and Shuwen Zhang, "Shot Boundary Detection based on Multilevel Difference of Colour Histograms," *Proceedings - 2016 1st International Conference on Multimedia and Image Processing, ICMIP 2016*, pp. 15–22, 2016.