Distant Viewing of the Harry Potter Movies via Computer Vision

Alina El-Keilany, Thomas Schmidt and Christian Wolff

Media Informatics Group, University of Regensburg, Germany

Abstract

We present an exploratory study performing distant viewing via computer vision methods in the genre of fantasy movies. As a case study we use 10 modern fantasy movies of the *Harry Potter* franchise (also referred to as *Wizarding World* franchise). We apply methods and state-of-the-art models for color and brightness analysis, object detection, location classification as well as facial emotion recognition. We present descriptive results as well as inference statistics. Furthermore, we discuss the results and the quality of the methods for this unique use case and give examples. We were able to find significant differences in our statistical analysis in the results of the methods across the movies with the movies of the *Harry Potter* series getting darker and negative emotional expressions on faces becoming more frequent.

Keywords

computer vision, film studies, distant viewing, harry potter, object detection, emotion recognition

1. Introduction

Digital film analysis has gained a lot of interest and popularity in digital humanities (DH) in recent years. Although movies are a multimodal medium, research often focuses on one specific modality. A lot of research uses the text channel as it is more accessible and methods are more established in DH [1, 2, 3]. However, due to advances in machine learning and computer vision (CV), scholars have also started investigating the visual image channel of movies, for example, to analyze shot lengths [4], colors [5, 6, 7], contrast [8] or sentiment [9, 10]. However, current CV methods offer possibilities beyond basic visual parameters like the method of object detection which has been used in DH for various tasks [11, 12, 13, 14] and emotion recognition which has been used in theater studies [15]. To get a larger overview of potential CV methods and tools, we recommend the survey paper by Pustu-Iren et al. [16].

To give this research branch a theoretical grounding, Arnold and Tilton [11] defined the term "distant viewing" for this kind of computational quantitative analysis of movies and other video

The 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022), Uppsala, Sweden, March 15-18, 2022.

alina.el-keilany@stud.uni-regensburg.de (A. El-Keilany); thomas.schmidt@ur.de (T. Schmidt); christian.wolff@ur.de (C. Wolff)

https://www.uni-regensburg.de/sprache-literatur-kultur/medieninformatik/sekretariat-team/thomas-schmidt/index.html (T. Schmidt); https://www.uni-regensburg.de/sprache-literatur-kultur/medieninformatik/sekretariat-team/christian-wolff/index.html (C. Wolff)

0000-0001-7171-8106 (T. Schmidt); 0000-0001-7278-8595 (C. Wolff)

material in DH. In this paper, we extend previous work on five case studies [14] and present a project in the line of distant viewing research for the specific case study of modern "fantasy" movies, more precisely 10 cinema movies of the *Wizarding World* (*Harry Potter*) franchise. We selected various popular CV methods and applied them on the movies: Color and brightness analysis, object detection, location classification and emotion recognition. Our approach is predominantly exploratory. We investigate if these methods uncover certain characteristics of the movies that can be validated statistically and if we can identify diachronic developments across the movies with the metrics given by the CV methods (similar to research on websites by [17]). By doing so, we want to reflect upon the advantages, disadvantages and limitations of the specific methods for digital film studies and which methods to pursue for further research.

2. Corpus and Preprocessing

The movie corpus for our analysis, consists of ten released movies of the *Wizarding World* franchise, consisting of the two subseries *Harry Potter* and *Fantastic Beasts*. The *Harry Potter* Series is based on J.K. Rowling's books of the same title, and follows the eponymous Harry Potter, a student at Hogwarts School of Witchcraft and Wizardry, on his journey of coming of age and his fight against the main antagonist Voldemort. In 2016 a new series in the *Wizarding World* franchise begun with *Fantastic Beasts and Where to Find Them* and continued with the release of *Fantastic Beasts: The Crimes of Grindelwald* in 2018. In general, the movies are prototypical for the fantasy genre.

The titles, short titles, and abbreviations (as we use them in this paper), release years, directors as well as the run times of the movies are shown in table 1. All movies have a frame rate of 25 frames per second with each frame having 32 bits per sample and a 720x576 resolution. The technical prerequisites of all CV methods are met. As a sample for our analysis we regard one frame per second of each movie. Therefore, we extracted a single frame for every second of the movie, keeping its temporal integrity, while reducing the data we process drastically. We do regard this sample as sufficient and representative of the movies. The number of frames we effectively worked with is presented in the column "frames" in table 1. Overall, we collected 77,192 frames which we will refer to as the corpus.

Considering the results, we will first present descriptive data and then inference statistics via significance tests for the methods we gathered numeric data. As significance test, we performed a one-way Welch's ANOVA except for one setting with nominal data for which we use Pearson's chi-squared test. Our data meets all necessary requirements for these test. We speak of significant differences for $p < 0.05$ and refer to Cohen ([18]) to interpret the effect in the case of ANOVAs. Cohen defines η^2 >0.01 as weak, >0.06 as moderate and >0.14 as strong effect. Furthermore, while we did not perform rigorous systematic evaluations, we will report upon the general impression about the quality of the methods.

Title	Year	Director	Runtime (mins.)	Frames
Harry Potter and the Philosopher's Stone (Harry Potter 1; HP1)	2001	Chris Columbus	152	8,293
Harry Potter and the Chamber of Secrets (Harry Potter 2; HP2)	2002	Chris Columbus	161	8,606
Harry Potter and the Prisoner of Azkaban (Harry Potter 3; HP3)	2004	Alfonso Cuarón	142	7,465
Harry Potter and the Goblet of Fire (Harry Potter 4; HP4)	2005	Mike Newell	157	8,270
Harry Potter and the Order of the Phoenix (Harry Potter 5; HP5)	2007	David Yates	138	7,388
Harry Potter and the Half-Blood Prince (Harry Potter 6; HP6)	2009	David Yates	153	8,278
Harry Potter and the Deathly Hallows – Part 1 (Harry Potter 7; HP7)	2010	David Yates	146	7,750
Harry Potter and the Deathly Hallows – Part 2 (Harry Potter 8; HP8)	2011	David Yates	130	6,791
Fantastic Beasts and Where to Find Them (Fantastic Beasts; FB1)	2016	David Yates	133	7,147
Fantastic Beasts: The Crimes of Grindelwald (Fantastic Beasts 2; FB2)	2018	David Yates	134	7,204

Table 1
General information on the movie corpus. We extracted one frame per second for each movie.

3. Color Analysis

3.1. Approach

We analyzed the movies' visual parameters color and brightness using OpenCV [19]. For the color analysis, we focus on "movie barcodes", a method already applied in color analysis for digital film studies [5, 6, 7]. To get an average color value for each frame, we imported the frames as arrays of RGB-values and calculated a mean value for all three color-channels over all pixels. These mean color values extracted per frame can be utilized to visualize the movies by generating a so-called "movie barcode", in which each frame is represented by a vertical line of its mean color [5]. The barcodes can be used to view the movies from a distance and let us perceive the diachronic progression of colors across a movie and multiple movies.

3.2. Results

The movie barcodes show significant artistic scenes considering color usage (fig. 1): For example, the light blue strips in the middle of HP3 consist of scenes playing in winter; the large field of light in the otherwise rather dark HP8 is due to a specific scene in which Harry Potter spends time in a state of limbo. The movie barcodes show a lot of warm browns and beige tones in the first two Harry Potter movies as well as larger areas of dark blue, green and cyan colors in HP3. Overall, the movies tend to get darker and less colorful which is in line with the plot of the movies getting more serious and less light-hearted. Reflecting upon the benefits of this method, we conclude that movie barcodes do offer an interesting analysis method for the overall style

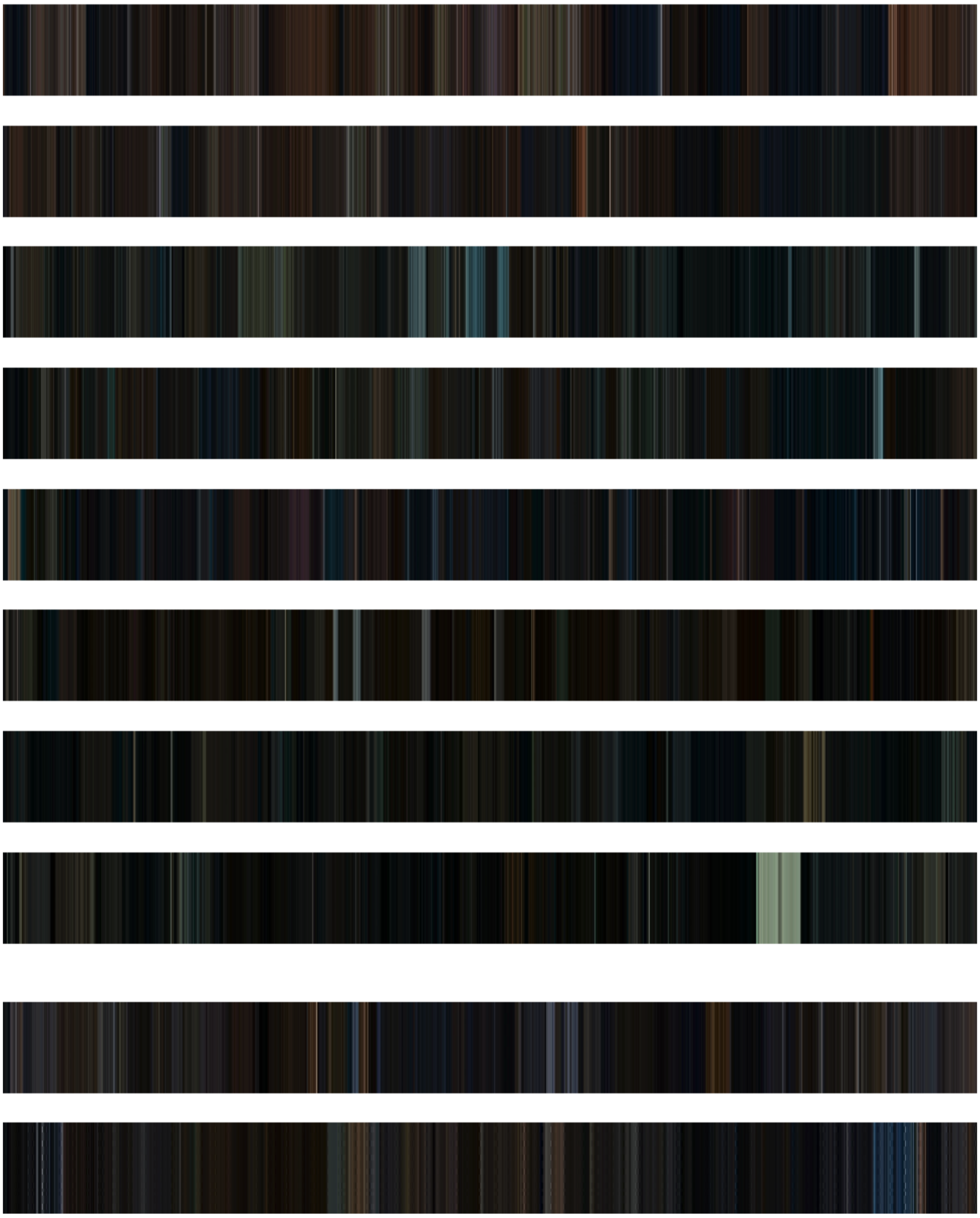

Figure 1: "Movie barcodes" for all movies (HP1-HP8, FB1-FB2, from top to bottom).

and presentation of a movie. However, the limitation is that the analysis is done in a rather qualitative way consisting of interpretation of the barcodes which is always a process that is prone to subjectivity.

movie	mean	SD	median	max
HP1	0.19	0.12	0.17	0.91
HP2	0.15	0.08	0.13	0.98
HP3	0.16	0.12	0.13	0.96
HP4	0.13	0.09	0.11	0.85
HP5	0.12	0.09	0.10	0.97
HP6	0.10	0.09	0.07	0.87
HP7	0.10	0.08	0.07	0.81
HP8	0.13	0.16	0.07	0.96
FB1	0.15	0.10	0.13	0.97
FB2	0.14	0.10	0.12	0.94
Overall	0.14	0.11	0.11	0.98

Table 2
Descriptive statistics for brightness for each movie and overall. Mean is the average across all frames, SD is the standard deviation, max the maximum.

4. Brightness Analysis

4.1. Approach

We calculated the brightness value for each frame by converting it to a grayscale image and then calculating the mean value over all pixels representing the image's brightness on a scale of 0 to 1 (with 0 being a solid black and 1 being a solid white image).

4.2. Results

Table 2 summarizes the statistics for the brightness values. The highest brightness value can be found for HP1 (M = 0.19) and the lowest for HP7 (M = 0.10). Indeed the brightness becomes consistently lower throughout the series. The *Fantastic Beasts* movies are of average brightness. We performed a one-way Welch's ANOVA to assess the significance of the difference among the movies. We did receive a significant result (F = 680.21, $p < 0.001$). Post-hoc tests (using Holm correction to adjust p) showed the largest effects regarding the difference between HP1 and HP6 ($\eta^2 = 0.15$), and HP1 and HP7 ($\eta^2 = 0.16$), which are large effects according to Cohen ([18]).

These results are in line with the plot of the movie becoming more serious and darker. Thus, with the brightness analysis and the inference statistics we show that more recent movies differ significantly to the older movies considering this metric although the absolute values are rather similar. Subsequently, we see brightness analysis as a beneficial method for digital film studies. However, it is hard to point to specific scenes and frames since the summarized value is an overall calculation over all frames of a movie. Nevertheless, we can also look for maximum values to find interesting stylistic scenes and frames for in-depth analysis (e.g. fig 2).

Figure 2: Frame with the highest brightness value in HP3. (0.94)

5. Object Detection

5.1. Approach

Object detection is the task to predict object classes and their positions in images. We performed object detection with the Detectron2 API[1] [20] which is regarded as state-of-the-art for object detection. We used a mask-RCNN model pretrained on the well-known COCO dataset [21]. The model can predict 80 common everyday objects like cars, animals or furniture. The predictor takes frames as input and delivers the detected object, its respective location mask, and the confidence of the prediction on a scale of 0 to 1. We set the threshold for the confidence score to 0.5 for a prediction. This rather low value allows for an exploratory assessment of the results, while cutting off the model's too uncertain predictions. To compare the movies regarding the objects occurring in them, we counted the objects for every frame and summed up the total number of occurrences for each object over all frames. Additionally, we calculated the percentage of frames an object is detected in.

5.2. Results

To analyze the results of this method we focus on frequency distributions across movies. Tables 3, 4 and 5 illustrate the 10 most frequently detected objects for each movie and overall. The overall impression is that the distributions are rather homogeneous. Persons are the most frequent objects in all movies by a wide margin which is likely a general characteristic of movies (fig. 3). Other common objects are ties (as they are part of the school uniforms), chairs and books. Objects that uncover specific characteristics of the movies are rare except for the suitcase object in FB1 (see table 5). This object does appear in a high frequency for this movie since it is an important part of the protagonist and the plot in general.

We did not perform exact evaluations but we analyzed the detection results heuristically by scanning through multiple examples across all movies. We gained the impression that the person detection and the detection of furniture does work quite accurate. However, we did

[1] https://github.com/facebookresearch/detectron2

Figure 3: Frame with the most frequent persons as determined by the object detection (HP2).

identify problems with objects in the movies that are not part of the COCO-class set. For example, many of the detected animals are actually fantasy creatures for which (of course) no predefined class is set in the used model (fig. 4). On the other hand, we also identified false classifications for objects that are in the model but actually not part of the movies like wands being classified as smartphones. While all these problems are understandable, we conclude from this that the method of object detection has its greatest potential when adapting the models to the unique domain of a movie genre so that the model does deal with the objects that are important for the specific genre.

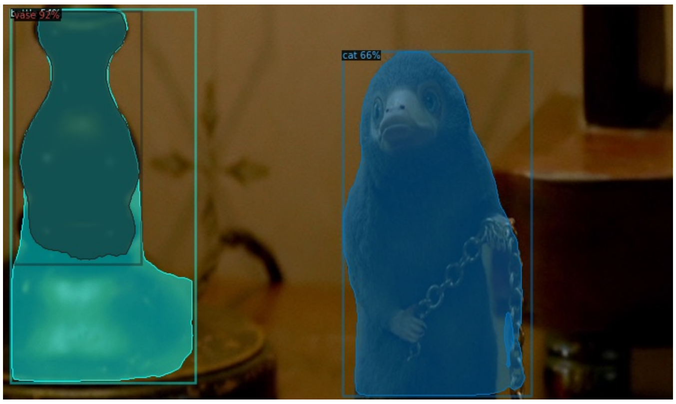

Figure 4: Frame with a fantasy creature "falsely" classified as cat by the object detection (FB1).

Harry Potter 1			Harry Potter 2			Harry Potter 3			Harry Potter 4		
object	#	%	**object**	#	%	**object**	#	%	**object**	#	%
person	23,138	88.0%	person	23,839	87.5%	person	22,842	82.1%	person	31,750	83.8%
tie	2,941	19.5%	tie	3,508	25.8%	tie	2,536	15.6%	tie	2,693	17.3%
chair	824	6.6%	book	2,459	4.2%	chair	1,140	9.7%	chair	598	5.7%
book	820	2.3%	chair	1,050	8.6%	book	586	3.8%	book	307	1.9%
cup	492	3.3%	vase	356	3.2%	bottle	498	4.3%	handbag	299	3.3%
dining table	297	2.7%	cup	325	2.8%	bird	430	4.3%	bottle	244	1.8%
bird	284	2.1%	dog	298	3.3%	dining table	430	4.1%	horse	219	2.4%
horse	284	3.2%	bottle	230	2.1%	cup	396	3.9%	cup	201	1.7%
wine glass	245	2.1%	handbag	223	2.4%	horse	320	0.3%	dog	195	2.3%
vase	211	2.2%	dining table	218	2.1%	wine glass	302	2.4%	wine glass	193	1.7%

Table 3
Distribution of top 10 detected objects for each movie (part 1). # is the absolute frequency for this object class. % is the percentage of frames containing the specific object at least once.

Harry Potter 5			Harry Potter 6			Harry Potter 7			Harry Potter 8		
object	#	%	**object**	#	%	**object**	#	%	**object**	#	%
person	25,984	87.9%	person	19,377	82.3%	person	17,762	84.9%	person	21,739	83.7%
tie	3,491	23.5%	book	1,716	4.3%	chair	1,612	9.6%	tie	1,315	8.5%
chair	1,026	9.5%	tie	1,637	14.1%	book	1,257	3.2%	chair	328	3.9%
cup	749	5.6%	cup	1,210	6.5%	tie	906	9.2%	book	207	1.4%
bottle	644	3.5%	chair	1,144	9.2%	dining table	479	3.3%	handbag	155	2.2%
book	326	3.4%	bowl	634	4.0%	bottle	338	2.5%	bottle	148	1.3%
handbag	315	3.4%	wine glass	603	4.3%	cup	274	2.6%	horse	135	1.7%
dining table	302	3.3%	dining table	589	5.4%	bird	265	1.0%	cup	112	1.6%
vase	301	3.3%	vase	466	4.6%	wine glass	252	1.7%	dog	108	1.2%
wine glass	269	2.6%	bottle	442	3.9%	car	229	1.3%	wine glass	90	0.7%

Table 4
Distribution of top 10 detected objects for each movie (part 2).

Fantastic Beasts 1			Fantastic Beasts 2			Overall		
object	#	%	**object**	#	%	**object**	#	%
person	21,070	87.6%	person	20,961	86.2%	person	228,462	85.4%
tie	3,850	35.6%	tie	3,611	29.6%	tie	26,488	19.8%
chair	947	10.2%	chair	1,730	11.7%	chair	10,399	8.4%
book	713	4.1%	book	634	3.2%	book	9,025	3.2%
cup	456	3.8%	bottle	390	3.0%	cup	4,481	3.4%
handbag	297	3.6%	dining table	263	2.4%	bottle	3,295	2.5%
suitcase	277	3.4%	cup	230	2.5%	dining table	3,021	2.9%
bottle	270	2.1%	handbag	227	2.8%	wine glass	2,469	2.1%
wine glass	220	1.6%	Dog	221	2.7%	vase	2,317	2.5%
dining table	200	2.5%	Vase	209	2.4%	handbag	2,295	2.6%

Table 5
Distribution of top 10 detected objects for each movie (part 3) and overall.

6. Location Classification

6.1. Approach

Location classification (also often called place or scene classification) does not refer to the geographical location of an image but the overall setting which an image depicts, e.g. a forest,

movie	indoor	outdoor
HP1	78.9%	21.1%
HP2	84.2%	15.8%
HP3	60.5%	39.5%
HP4	75.1%	24.9%
HP5	82.8%	17.2%
HP6	85.2%	14.8%
HP7	70.2%	29.8%
HP8	75.4%	24.6%
FB1	73.7%	26.3%
FB2	74.8%	25.2%
Overall	76.3%	23.7%

Table 6
Distribution of frames classified as predominantly indoor or outdoor for each movie and overall.

an indoor-room, a street etc. To detect locations and the setting of a scene, we used places365[2], which offers a residual neural network (ResNet) pretrained on the Places2[3] dataset [22]. The ResNet can predict 365 location categories, including rather exotic ones like "airfields" or "zen gardens" based on what the overall image resembles the most. The 365 classes are structured in a hierarchical order summing up to differ between *indoor* and *outdoor* on the highest level. Using the model on preprocessed images yields the most likely location as well as the prediction confidence on a scale of 0 to 1. The default mode of the location classifier is assigning every image with the most likely location, but the probabilities of these predictions are often very low. Therefore, we introduced a threshold of 0.7 to keep only rather certain predictions of the model. This resulted in 14,263 classified frames (18.5% of all frames). For each movie we summed up the number of times the location is predicted and calculated the percentage of frames it is detected in. Additionally, we categorized each frame into the groups *indoor* and *outdoor*, using the model's 5 most likely predictions and majority voting.

6.2. Results

First, table 6 presents the distribution of frames classified as rather indoor and outdoor for all movies. We can consistently identify that the majority of frames across all movies are classified as indoor. This is in line with the content of the movies that usually take place inside of a castle. We performed a Pearson's chi-squared test, which showed significant differences between the movies (χ^2 = 243.9, p < 0.001). The effect size measured by Cramér's V (0.13) shows a weak effect ([18]). We can see that the indoor-percentage decreases for the last two movies which makes sense plot-wise since the main characters travel throughout the movies. However, the large outdoor-percentage for HP3 is mostly due to misclassifications. This movie is shot with a lot of blue lightning and effects due to artistic reasons which are constantly misclassified as *underwater* (see fig. 5).

Table 7, 8 and 9 illustrate the distribution of the subcategories across all movies. Similar to

[2]https://github.com/CSAILVision/places365
[3]http://places2.csail.mit.edu/

Figure 5: Frame falsely classified as *underwater* (HP3).

Harry Potter 1			Harry Potter 2			Harry Potter 3			Harry Potter 4		
location	#	%	**location**	#	%	**location**	#	%	**location**	#	%
jail cell	235	21.1%	jail cell	498	37.8%	jail cell	557	50.1%	jail cell	375	27.7%
catacomb	229	20.6%	catacomb	291	22.1%	catacomb	68	6.1%	catacomb	348	25.7%
nursing home	62	5.6%	archive	72	5.5%	elevator shaft	65	5.8%	aquarium	116	8.6%
aquarium	55	4.9%	pub/ indoor	63	4.8%	ocean deep	50	4.5%	ocean deep	73	5.4%
stage/ indoor	45	4.0%	elevator shaft	51	3.9%	aquarium	41	3.7%	discotheque	59	4.4%
staircase	41	3.7%	aquarium	31	2.4%	sky	37	3.3%	elevator shaft	40	3.0%
elevator shaft	39	3.5%	bookstore	29	2.2%	train interior	24	2.2%	sky	39	2.9%
conference center	36	3.2%	sky	21	1.6%	staircase	21	1.9%	auditorium	20	1.5%
archive	23	2.1%	hospital room	20	1.5%	hospital room	20	1.8%	throne room	18	1.3%
sky	22	2.0%	Slum	15	1.1%	crevasse	18	1.6%	staircase	18	1.3%

Table 7
Distribution of top 10 detected locations for each movie (part 1).

the object detection, the distribution is overall homogeneous. However, the detected classes are often rather exotic. The frequent *jail* classifications are surprising. While some scenes do play in jails, most of these classifications are due to the lattice-like windows in the Hogwarts castle in which most of the movies take place (see fig. 6). While many classifications are understandable, the method suffers from the fact that the model is trained for the classification of nature photographs and not for movies. Close shots pose a lot of challenges to the model due to the missing surroundings and landscapes. In future work we intend to segment these shots from wide shots including landscapes to focus on the rather correctly classified frames.

7. Emotion Recognition

7.1. Approach

Emotion recognition is the method to detect emotions on human faces and employed in various use cases in computer science ([23, 24, 25, 26] but, to the best of our knowledge, rarely on the image channel in DH [15] but predominantly on text, e.g. plays [27, 28, 29] or social media

Harry Potter 5			Harry Potter 6			Harry Potter 7			Harry Potter 8		
location	**#**	**%**	**location**	**#**	**%**	**location**	**#**	**%**	**location**	**#**	**%**
jail cell	374	31.8%	catacomb	1,113	51.1%	jail cell	515	34.7%	jail cell	858	51.1%
discotheque	149	12.7%	jail cell	762	35.0%	catacomb	396	26.7%	catacomb	518	30.9%
catacomb	125	10.6%	archive	61	2.8%	basement	113	7.6%	elevator shaft	92	5.5%
pub/indoor	76	6.5%	elevator shaft	52	2.4%	bamboo forest	78	5.3%	church/ indoor	34	2.0%
aquarium	71	6.0%	sky	30	1.4%	elevator shaft	45	3.0%	aquarium	23	1.4%
elevator shaft	53	4.5%	alley	17	0.8%	sky	33	2.2%	sky	19	1.1%
stage/indoor	44	3.7%	igloo	15	0.7%	campsite	32	2.2%	staircase	14	0.8%
underwater/ ocean deep	30	2.6%	aquarium	14	0.6%	elevator/ door	26	1.8%	escalator/ indoor	12	0.7%
medina	28	2.4%	stable	12	0.6%	wheat field	21	1.4%	subway station/ platform	11	0.7%
playground	24	2.0%	cemetery	11	0.5%	alley	20	1.3%	basement	9	0.5%

Table 8
Distribution of top 10 detected locations for each movie (part 2).

Fantastic Beasts 1			Fantastic Beasts 1			Overall		
location	**#**	**%**	**location**	**#**	**%**	**location**	**#**	**%**
jail cell	487	39.8%	jail cell	914	56.0%	jail cell	5,575	39.1%
catacomb	142	11.6%	catacomb	210	12.9%	catacomb	3,440	24.1%
bamboo forest	55	4.5%	sky	60	3.7%	elevator shaft	548	3.8%
elevator shaft	54	4.4%	elevator shaft	57	3.5%	aquarium	435	3.0%
pub/indoor	41	3.3%	aquarium	39	2.4%	sky	309	2.2%
sauna	32	2.6%	medina	39	2.4%	discotheque	272	1.9%
bank vault	30	2.5%	igloo	33	2.0%	pub/indoor	247	1.7%
aquarium	28	2.3%	throne room	26	1.6%	archive	175	1.2%
sky	27	2.2%	burial chamber	24	1.5%	underwater/ocean deep	171	1.2%
igloo	26	2.1%	crevasse	18	1.1%	crevasse	20	1.3%

Table 9
Distribution of top 10 detected locations for each movie (part 3) and overall.

Figure 6: Frame falsely classified as *jail cell* (HP8).

content [30, 31]. We used the Python module FER[4][32] to recognize the characters‘ emotions. In a first step, the faces must be detected. We used a multitask cascaded convolutional networks (MTCNN; [33]) and the Haar Cascade facial recognition algorithm proposed by Viola and Jones [34]. For the emotion analysis, we used a CNN trained on the FER-2013[32] data set that can predict the seven emotional categories *anger, disgust, fear, happiness, neutral, sadness*

[4]https://pypi.org/project/fer;https://github.com/justinshenk/fer

	HP1		HP2		HP3		HP4		HP5		HP6	
emotion	**mean**	**%**	**mean**	**%**	**mean**	**%**	**mean**	**%**	**mean**	**%**	**mean**	**%**
angry	0.16	11.4%	0.16	10.2%	0.17	12.7%	0.21	18.6%	0.15	8.5%	0.19	14.2%
disgust	0.00	0.1%	0.00	0.1%	0.00	0.0%	0.00	0.0%	0.00	0.0%	0.00	0.0%
fear	0.11	5.1%	0.11	4.7%	0.10	2.9%	0.10	2.5%	0.10	2.5%	0.09	2.0%
happy	0.10	9.0%	0.09	7.9%	0.11	7.8%	0.12	10.1%	0.10	8.1%	0.09	8.2%
neutral	0.23	24.3%	0.25	27.8%	0.24	27.2%	0.18	15.2%	0.27	31.4%	0.24	24.5%
sad	0.29	41.2%	0.30	43.2%	0.31	45.3%	0.32	49.2%	0.31	46.6%	0.32	47.6%
surprise	0.10	8.8%	0.08	6.1%	0.06	4.1%	0.07	4.4%	0.06	3.0%	0.07	3.4%

Table 10
Results for the emotion recognition across all movies (part 1). *Mean* is the average of this emotion across all frames (with detected faces), % is the proportion of frames with this specific emotion as maximum value across all these frames.

	HP7		HP8		FB1		FB1		Overall	
emotion	**mean**	**%**	**mean**	**%**	**mean**	**%**	**mean**	**%**	**mean**	**%**
angry	0.18	10.9%	0.21	17.5%	0.18	14.0%	0.19	15.8%	0.18	12.9%
disgust	0.00	0.0%	0.01	0.4%	0.00	0.0%	0.00	0.0%	0.00	0.1%
fear	0.09	2.0%	0.10	3.9%	0.12	5.5%	0.11	3.3%	0.10	3.6%
happy	0.07	4.7%	0.08	5.1%	0.09	7.4%	0.07	5.9%	0.09	7.5%
neutral	0.20	17.6%	0.20	19.2%	0.23	24.9%	0.21	21.9%	0.23	24.1%
sad	0.37	59.7%	0.34	50.7%	0.30	42.7%	0.33	47.9%	0.32	46.9%
surprise	0.08	5.1%	0.07	3.3%	0.08	5.5%	0.08	5.1%	0.08	5.0%

Table 11
Results for the emotion recognition across all movies (part 2) and overall.

and *surprise*. For every face multiple emotions can be predicted simultaneously in varying percentages, summing up to 1. If more than one face was detected in a frame, we calculated a mean value for the emotions. Additionally, we assigned the highest scoring emotion as the dominant emotion for a frame, which allows us to explore what the most dominant emotion is for every movie.

7.2. Results

For the statistical analysis, we averaged the means for all frames of a movie on which at least one face is detected to get an overall value. Furthermore, we calculated the percentage of frames having a specific emotion as maximum value of all emotions (on the same set of frames). In tables 10 and 11, we present the results.

The generally low mean values are due to the fact that the emotion classes often have a value of 0 since the values do need to sum up to 1. Overall, we identified *sadness* as the most frequently detected emotion (46.9%) (see fig. 7 for an example) followed by neutral (24.1%) and anger (12.9%). The *sadness* value increases up to HP8 reaching the maximum in HP7 (59.7%) while the *happy* value decreases. Again, this points to the increased dramatic seriousness in the plot throughout the movies. The emotion *disgust* was rarely detected. We performed a Welch's ANOVA test and found that, indeed, the difference among the movies for each emotion class is significant ($p < 0.001$). However, the effect size is rather small for most emotions ($\eta^2 < 0.01$)

except for *angry* with a moderate effect ($\eta^2 = 0.02$). Nevertheless, this shows that the movies are rather homogeneous concerning the emotional tone. Analyzing the results, we found that the emotion detection generally works quite well. However, the face detection has problems dealing with faces that are not looking directly towards the camera (fig. 8). This is due to the fact that the training material of the model predominantly consists of such faces. We conclude for the face detection that it needs domain adaptation for the complex angles that movies consist of.

Figure 7: Frame with a maximum *sad* value (HP8).

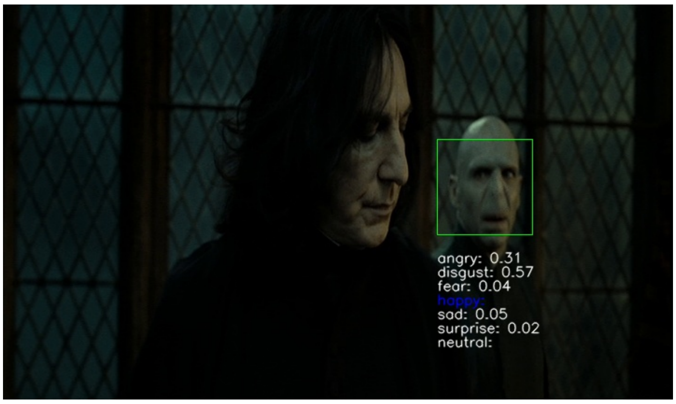

Figure 8: Frame with a face not looking directly towards the camera, thus not being detected by the face detection (HP8).

8. Discussion

One of our research goals was to identify if the applied methods can uncover specific characteristics and differences across the movies. Indeed, the color and brightness analysis showed descriptive differences and in the case of brightness differences that could be supported by significance tests. The movies of the *Harry Potter* series tend to get darker. These general visual results are in line with the results of the emotion analysis which also shows an increase in *sadness* classifications and a decrease for the average *happiness* value. However, for most of the other methods, we found significant differences but with rather low effects. Most of the methods behaved rather homogeneous across the movies with some punctual exceptions. One reason for this might be that the movies belong to the same series, franchise and genre. Therefore, the stylistic and content-based differences might be too small to become apparent via these kind of methods. We want to explore this assumption in future work by conducting case studies with movies of different genres and decades.

We did not perform an exact evaluation. We plan to do so in future work for some of the methods by systematically evaluating a subset of the corpus against human-made annotation to get a precise overview about the quality of the methods. However, we did sporadically explore the quality of the results while conducting our research. While we do think that all of the methods in many cases work surprisingly well, mistakes and misclassifications are not rare. Many of these problems are connected to the fantasy genre and the behavior of the model is understandable. We conclude that this is the general main challenge of the research. All of the CV methods are not intended for artistic movies and therefore need domain-adaptation which is possible and has been a common research branch in machine learning in recent years. However, domain-adaption needs large amounts of correctly annotated frames which is very resource-intensive and challenging for similar narrative content like plays [35, 36]. Nevertheless, we intend to further this process by starting annotation studies for one of the most promising methods, object detection, which we will then use to train and extend general purpose models for the specific use case of fantasy movies.

Despite the problems, we could show that many of the methods offer a lot of possibilities for large-scale distant viewing research in digital film studies. We see a great potential in combining the methods to explore correlations, for example if certain locations in genre-based movies appear more frequent with specific objects. At the same time, we also see potential in analyzing diachronic developments or in comparing different genres via CV methods.

References

[1] E. Hoyt, K. Ponto, C. Roy, Visualizing and Analyzing the Hollywood Screenplay with ScripThreads, Digital Humanities Quarterly 008 (2014).

[2] A. Hołobut, J. Rybicki, The Stylometry of Film Dialogue: Pros and Pitfalls, Digital Humanities Quarterly 014 (2020).

[3] J. Byszuk, The Voices of Doctor Who – How Stylometry Can be Useful in Revealing New Information About TV Series, Digital Humanities Quarterly 014 (2020).

[4] M. Baxter, D. Khitrova, Y. Tsivian, Exploring cutting structure in film, with applications to the films of D. W. Griffith, Mack Sennett, and Charlie Chaplin, Digital Scholarship in the Humanities 32 (2017) 1–16. URL: https://doi.org/10.1093/llc/fqv035. doi:10.1093/llc/fqv035.

[5] M. Burghardt, K. Hafner, L. Edel, S.-L. Kenaan, C. Wolff, An information system for the analysis of color distributions in moviebarcodes, in: M. Gäde (Ed.), Everything changes, everything stays the same? Understanding information spaces : Proc.15th Int. Symp. of Information Science (ISI 2017), Berlin, Germany, 13th-15th March 2017, volume 70 of

Schriften zur Informationswissenschaft, Verlag Werner Hülsbusch, Glückstadt, 2017, pp. 356–358. URL: https://epub.uni-regensburg.de/35682/.
[6] B. Flueckiger, G. Halter, Methods and Advanced Tools for the Analysis of Film Colors in Digital Humanities, Digital Humanities Quarterly 014 (2020).
[7] N. Redfern, Colour palettes in US film trailers: a comparative analysis of movie barcode, Umanistica Digitale (2021) 251–270. URL: https://umanisticadigitale.unibo.it/article/view/12468. doi:10.6092/issn.2532-8816/12468, number: 10.
[8] Pause, Johannes, Walkowski, Niels-Oliver, Dead and Beautiful: The Analysis of Colors by Means of Contrasts in Neo-Zombie Movies, Digital Humanities 2017. Conference Abstracts (2017).
[9] T. Schmidt, D. Halbhuber, Live sentiment annotation of movies via arduino and a slider, in: Digital Humanities in the Nordic Countries 5th Conference 2020 (DHN 2020). Late Breaking Poster., 2020. URL: https://epub.uni-regensburg.de/49300/.
[10] T. Schmidt, I. Engl, D. Halbhuber, C. Wolff, Comparing live sentiment annotation of movies via arduino and a slider with textual annotation of subtitles., in: DHN Post-Proceedings, 2020, pp. 212–223. URL: https://epub.uni-regensburg.de/50811/.
[11] T. Arnold, L. Tilton, Distant viewing: analyzing large visual corpora, Digital Scholarship in the Humanities (2019). URL: https://doi.org/10.1093/digitalsh/fqz013. doi:10.1093/digitalsh/fqz013.
[12] G. Howanitz, B. Bermeitinger, E. Radisch, S. Gassner, M. Rehbein, S. Handschuh, Deep Watching - Towards New Methods of Analyzing Visual Media in Cultural Studies, 2019. doi:10.5281/zenodo.3326470.
[13] T. Schmidt, S. Kurek, Der Einsatz von Computer Vision-Methoden für Filme - Eine Fallanalyse für die Kriminalfilm-Reihe Tatort, in: DHd 2022 Kulturen des digitalen Gedächtnisses. 8. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum" (DHd 2022), Potsdam, Germany, 2022. URL: https://zenodo.org/record/6328167. doi:10.5281/zenodo.6328167.
[14] T. Schmidt, A. El-Keilany, J. Eger, S. Kurek, Exploring Computer Vision for Film Analysis: A Case Study for Five Canonical Movies, in: 2nd International Conference of the European Association for Digital Humanities (EADH 2021), Krasnoyarsk, Russia, 2021. URL: https://epub.uni-regensburg.de/50867/. doi:10.5283/epub.50867.
[15] T. Schmidt, C. Wolff, Exploring Multimodal Sentiment Analysis in Plays: A Case Study for a Theater Recording of Emilia Galotti, in: Proceedings of the Conference on Computational Humanities Research 2021 (CHR 2021), Amsterdam, The Netherlands, 2021, pp. 392–404. URL: http://ceur-ws.org/Vol-2989/short_paper45.pdf.
[16] K. Pustu-Iren, J. Sittel, R. Mauer, O. Bulgakowa, R. Ewerth, Automated Visual Content Analysis for Film Studies: Current Status and Challenges, Digital Humanities Quarterly 014 (2020).
[17] T. Schmidt, A. Mosiienko, R. Faber, J. Herzog, C. Wolff, Utilizing html-analysis and computer vision on a corpus of website screenshots to investigate design developments on the web, Proceedings of the Association for Information Science and Technology 57 (2020) e392. URL: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/pra2.392. doi:10.1002/pra2.392.
[18] J. Cohen, Statistical power analysis for the behavioral sciences, 2nd ed ed., L. Erlbaum Associates, Hillsdale, N.J, 1988.

[19] G. Bradski, The OpenCV Library., Dr. Dobb's Journal: Software Tools for the Professional Programmer 25 (2000). URL: https://elibrary.ru/item.asp?id=4934581.
[20] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, R. Girshick, Detectron2, 2019. URL: https://github.com/facebookresearch/detectron2.
[21] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, P. Dollár, Microsoft COCO: Common Objects in Context, arXiv:1405.0312 [cs] (2015). URL: http://arxiv.org/abs/1405.0312, arXiv: 1405.0312.
[22] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, A. Torralba, Places: A 10 Million Image Database for Scene Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (2018) 1452–1464. URL: https://ieeexplore.ieee.org/document/7968387/. doi:10.1109/TPAMI.2017.2723009.
[23] A.-M. Ortloff, L. Güntner, M. Windl, T. Schmidt, M. Kocur, C. Wolff, Sentibooks: Enhancing audiobooks via affective computing and smart light bulbs, in: Proceedings of Mensch Und Computer 2019, MuC'19, Association for Computing Machinery, New York, NY, USA, 2019, p. 863–866. URL: https://doi.org/10.1145/3340764.3345368. doi:10.1145/3340764.3345368.
[24] D. Halbhuber, J. Fehle, A. Kalus, K. Seitz, M. Kocur, T. Schmidt, C. Wolff, The mood game - how to use the player's affective state in a shoot'em up avoiding frustration and boredom, in: Proceedings of Mensch Und Computer 2019, MuC'19, Association for Computing Machinery, New York, NY, USA, 2019, p. 867–870. URL: https://doi.org/10.1145/3340764.3345369. doi:10.1145/3340764.3345369.
[25] P. Hartl, T. Fischer, A. Hilzenthaler, M. Kocur, T. Schmidt, Audiencear - utilising augmented reality and emotion tracking to address fear of speech, in: Proceedings of Mensch Und Computer 2019, MuC'19, Association for Computing Machinery, New York, NY, USA, 2019, p. 913–916. URL: https://doi.org/10.1145/3340764.3345380. doi:10.1145/3340764.3345380.
[26] T. Schmidt, M. Schlindwein, K. Lichtner, C. Wolff, Investigating the relationship between emotion recognition software and usability metrics, i-com 19 (2020) 139–151. URL: https://doi.org/10.1515/icom-2020-0009. doi:10.1515/icom-2020-0009.
[27] T. Schmidt, K. Dennerlein, C. Wolff, Emotion Classification in German Plays with Transformer-based Language Models Pretrained on Historical and Contemporary Language, in: Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, Association for Computational Linguistics, Punta Cana, Dominican Republic (online), 2021, pp. 67–79. URL: https://aclanthology.org/2021.latechclfl-1.8. doi:10.18653/v1/2021.latechclfl-1.8.
[28] T. Schmidt, K. Dennerlein, C. Wolff, Using Deep Learning for Emotion Analysis of 18th and 19th Century German Plays, in: M. Burghardt, L. Dieckmann, T. Steyer, P. Trilcke, N.-O. Walkowski, J. Weis, U. Wuttke (Eds.), Fabrikation von Erkenntnis. Experimente in den Digital Humanities, 2021. doi:10.26298/melusina.8f8w-y749-udlf.
[29] T. Schmidt, K. Dennerlein, C. Wolff, Towards a Corpus of Historical German Plays with Emotion Annotations, in: D. Gromann, G. Sérasset, T. Declerck, J. P. McCrae, J. Gracia, J. Bosque-Gil, F. Bobillo, B. Heinisch (Eds.), 3rd Conference on Language, Data and Knowledge (LDK 2021), volume 93 of *Open Access Series in Informatics (OASIcs)*, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2021, pp. 9:1–9:11.

doi:10.4230/OASIcs.LDK.2021.9.
[30] T. Schmidt, P. Hartl, D. Ramsauer, T. Fischer, A. Hilzenthaler, C. Wolff, Acquisition and analysis of a meme corpus to investigate web culture., in: Digital Humanities Conference 2020 (DH 2020), Ottawa, Canada, 2020. URL: https://epub.uni-regensburg.de/49294/. doi:10.17613/mw0s-0805.
[31] T. Schmidt, F. Kaindl, C. Wolff, Distant reading of religious online communities: A case study for three religious forums on reddit., in: DHN, Riga, Latvia, 2020, pp. 157–172. URL: http://ceur-ws.org/Vol-2612/paper11.pdf.
[32] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, Y. Bengio, Challenges in Representation Learning: A report on three machine learning contests, arXiv:1307.0414 [cs, stat] (2013). URL: http://arxiv.org/abs/1307.0414, arXiv: 1307.0414.
[33] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks, IEEE Signal Processing Letters 23 (2016) 1499–1503. doi:10.1109/LSP.2016.2603342.
[34] P. Viola, M. J. Jones, Robust Real-Time Face Detection, International Journal of Computer Vision 57 (2004) 137–154. URL: https://doi.org/10.1023/B:VISI.0000013087.49260.fb. doi:10.1023/B:VISI.0000013087.49260.fb.
[35] T. Schmidt, B. Winterl, M. Maul, A. Schark, A. Vlad, C. Wolff, Inter-rater agreement and usability: A comparative evaluation of annotation tools for sentiment annotation, in: C. Draude, M. Lange, B. Sick (Eds.), INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge), Gesellschaft für Informatik e.V., Bonn, 2019, pp. 121–133. doi:10.18420/inf2019_ws12.
[36] T. Schmidt, M. Burghardt, K. Dennerlein, C. Wolff, Sentiment annotation for lessing's plays: Towards a language resource for sentiment analysis on german literary texts, in: T. Declerck, J. P. McCrae (Eds.), 2nd Conference on Language, Data and Knowledge (LDK 2019), 2019, pp. 45–50. URL: http://ceur-ws.org/Vol-2402/paper9.pdf.