Automatic Color Scheme Extraction from Movies

Conference Paper · June 2020

DOI: 10.1145/3372278.3390685

Automatic Color Scheme Extraction from Movies

Suzi Kim
School of Computing, KAIST
Republic of Korea
kimsuzi@kaist.ac.kr

Sunghee Choi
School of Computing, KAIST
Republic of Korea
sunghee@kaist.edu

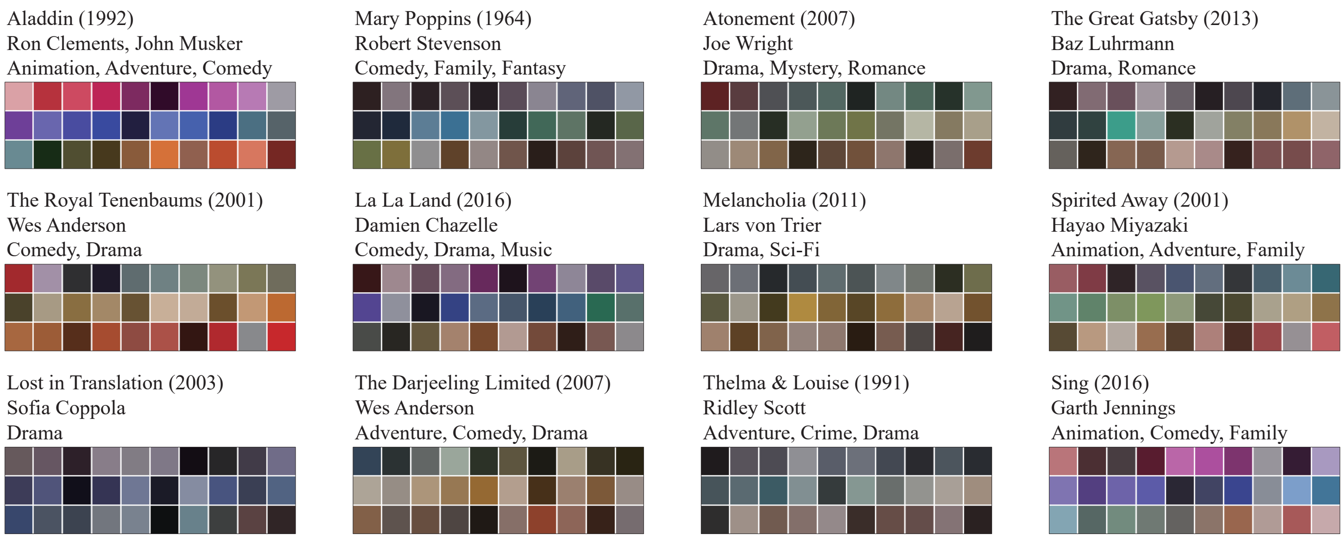


Figure 1: Examples of color scheme extraction to represent a movie's visual identity.

ABSTRACT

A color scheme is an association of colors, i.e., a subset of all possible colors, that represents a visual identity. We propose an automated method to extract a color scheme from a movie. Since a movie is a carefully edited video with different objects and heterogeneous content embodying the director's messages and values, it is a challenging task to extract a color scheme from a movie as opposed to a general video filmed at once without distinction of shots or scenes. Despite such challenges, color scheme extraction plays a very important role in film production and application. The color scheme is an interpretation of the scenario by the cinematographer and it can convey a mood or feeling that stays with the viewer after the movie has ended. It also acts as a contributing factor to describe a film, like the metadata fields of a film such as a genre, director, and casting. Moreover, it can be automatically tagged unlike metadata, so it can be directly applied to the existing movie database without much effort. Our method produces a color scheme from a movie in a bottom-up manner from segmented shots. We formulate the color extraction as a selection problem where perceptually important colors are selected using saliency. We introduce a semi-master-shot, an alternative unit defined as a combination of contiguous shots taken in the same place with similar colors. Using real movie videos, we demonstrate and validate the plausibility of the proposed technique.

CCS CONCEPTS

• **Information systems → Video search**; • **Computing methodologies → Video segmentation**; *Interest point and salient region detections*; *Image processing*; *Perception.*

KEYWORDS

color scheme; color palette; color extraction; video retrieval; color clustering; scene segmentation; shot segmentation

ACM Reference Format:
Suzi Kim and Sunghee Choi. 2020. Automatic Color Scheme Extraction from Movies. In *Proceedings of the 2020 International Conference on Multimedia Retrieval (ICMR '20), June 8–11, 2020, Dublin, Ireland.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3372278.3390685

ICMR '20, June 8–11, 2020, Dublin, Ireland

ACM ISBN 978-1-4503-7087-5/20/06...$15.00
https://doi.org/10.1145/3372278.3390685

1 INTRODUCTION

IMDb [22] is the largest and most popular movie database, comprising more than 5 million movies. Since 2000, more than 10,000 films have been released in the U.S. and Canada alone, and the number of releases has been increasing steeply [64]. Consequently, movie recommendations based on user preference have been researched more extensively. Most recommendation systems use social graphs based on movie metadata, such as genre, director, keyword, and cast. However, manually tagging metadata to each movie requires considerable labor, and it is extremely time-consuming to re-tag

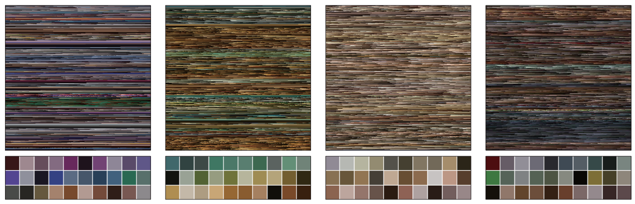


Figure 2: Metadata is not enough to quantify a movie's mise-en-scène. (From top to bottom) A set of colors that encompass the entire movie and color scheme extracted by our method.

every previous movie when a new metadata field is required. Therefore, extracting a unique descriptor from the video itself is very important for efficient recommendation systems.

Since a movie combines various media, it can be represented by multimodal descriptors, including visual, textual and audio features. Color is particularly significant for visual media [29, 42], affecting the viewer's perceptions and preferences. This paper proposes a system to automatically extract a color scheme from a movie to use as a descriptor.

A color scheme is an association of colors expressed as a subset of all possible colors to represent a visual identity. As revealed by previous studies [19, 35, 44], color has a strong influence on human perception to elicit emotional responses. Visual elements are the first thing that human perceives when they watch a video, and the color is the most basic element in visual aspects which influences the human's impression and emotion.

Film production strongly considers the color tone that dominates a movie, with a cinema colorist adjusting the overall movie color. Directors leverage the colors to support the narrative of the movie and generate a unified fictional space. According to the *Cinematography for Directors* [12], the color scheme is an interpretation of the scenario by the cinematographer and it can convey a mood or feeling that stays with the viewer after the movie has ended. It is because the color scheme is not just a result shot by a camera, but a combination of various elements of film production, including backgrounds and sets created by a production designer [17], lightings set by the gaffer [15], and costumes created by a wardrobe designer [28].

Each field of metadata, such as genre and director, cannot be a primary key that separates all movies individually. In the same manner, we do not intend to distinguish all movies only with the proposed color scheme. We attempt to show that the color scheme is not a unique characteristic for each movie, but a contributing factor to cluster the movies. For example, *La La Land* (2016) and *Whiplash* (2014) are drama musical films written and directed by Damien Chazelle. They share similar metadata, i.e., director, genre, and casting, but give different impressions due to the intensity of colors dominating the whole duration, as shown in Figure 2. Whiplash should also be linked to *Constantine* (2005) and *Léon: The Professional* (1994), which maintain similar color tones, but for now there is no common metadata to connect these films. A color scheme can be a very simple and accurate descriptor to quantify a movie's mise-en-scène.

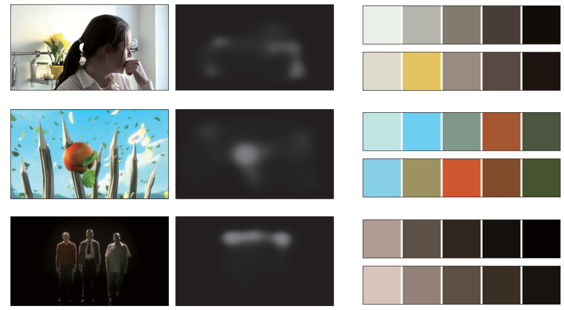


Figure 3: Differences in color scheme extraction according to the existence of the saliency map. Images taken from 1000 Days ©copyright Revival Film, Big Buck Bunny ©copyright 2008, Blender Foundation | www.bigbuckbunny.org, and Meridian ©copyright 2016, Virgin Soil Pictures Inc.

The colors in the movie are not just made but are the result of the intentional composition of film production and color grading. Using the color scheme as a descriptor can be a more fundamental approach than an emotion descriptor, which is frequently used in movie recommendation systems. Users and designers often discuss and share a list of the iconic colors of movies in IMDb [48–51], YouTube [56, 57], and personal websites [23, 41, 52–55, 59]. Every commercial film goes through different film production and color grading process, so each film has distinctive color scheme characteristics. Recently, the extraction of script-based or semantic context-based descriptors has been actively conducted, and color scheme-based descriptor has also begun to attract people's attention.

Several previous studies have considered color scheme extraction from images [20, 21, 24, 37] but little attention has been paid to the extraction from a video [75], and particularly from a movie. A movie is an elaborate compilation by the director, embodying their message and values. In contrast to the general videos, which are filmed continuously without shot or scene distinctions, movies are carefully edited with many different objects and heterogeneous contents. Movies are generally longer than the general videos, although not usually exceeding three hours, and include 200,000–250,000 images (assuming 24 fps). Although it is a challenge to extract major colors from so many images with complicated contents, there exists a dominant color scheme, by design, as you can imagine color palettes after watching Wes Anderson's movies.

As described in Figure 3, we cannot obtain a movie's overall color scheme through simple color clustering methods, due to interference of colors that appear repeatedly but meaninglessly, such as sky, walls, or black clutter. A saliency map is a solution to ensure that color selection reflects object importance in the movie. Saliency maps represent pixel significance within each image by following human fixation points. Since major pixels dominate the color impression, rather than all pixels, we employ saliency maps to obtain a color scheme from each frame.

The major contributions of this paper are as below:

- To the best of our knowledge, this is the first work to generate a color scheme from a video. Extracting a color scheme from images has been well studied in computer graphics because the color scheme is the most basic unit for image recoloring and vectorization. The reason why color scheme extraction is difficult for video is that it should consider the overall colors of the video while avoiding being driven by a less significant long-shot. So, we split a video into small units and choose the final color scheme in a bottom-up manner.

- We define a semi-master-shot, which is a new unit to combine contiguous shots taken in the same location with similar colors. The semi-master-shot can be used in video processing, which has been actively studied for decades, such as video highlight detection and video thumbnail generation.
- Beyond the simple saliency adoption, we conduct a deep consideration of how to use the saliency map properly. We measure the importance at three levels, namely, the importance of each pixel in the frame, and the importance of each frame in the shot and the importance of each shot in the movie.
- We demonstrate the proposed color scheme's plausibility and functionality as a descriptor using real movie videos.

2 RELATED WORK

2.1 Color Scheme Extraction

Color quantization from images has been extensively studied in computer vision and graphics fields to address color loss due to image compression [14]; display device limitations [18, 43, 76], which allow only a limited number of colors to be displayed; and color printing [58, 65]. It aims to express the original image with a smaller number of colors.

Recent studies have proposed a number of color scheme extraction methods for various purposes, including image recolorization [10, 38, 68, 80], image decomposition [26, 69, 70], image abstraction [13, 27], digital art [34], and image indexing [63]. Color scheme extraction is typically achieved by color clustering [10, 21, 25, 32, 47, 66, 80], histogram thresholding [31, 37, 63, 67, 71], image segmentation [1, 13, 40], and convex hull enclosure [69, 70].

Clustering approaches group the pixels based on distance in color space, where cluster centroids are selected as the representative color. Histogram thresholding converts an image into a color space histogram and identifies dominant colors by dividing the histogram into multiple sections. Convex hull enclosure converts an image into convex hull geometry for a specific color space and selects the convex hull boundary vertices as representative colors. Clustering and histogram methods always select colors that occur within the image, whereas convex hull enclosure can select colors that express the original image well but may not necessarily exist in the image.

We propose a two-stage color scheme extraction: base palette extraction from a semi-master-shot (Section 5) and merging palettes into a single color scheme (Section 6). We employ both color clustering and convex hull enclosure methods within the proposed pipeline to take advantage of each method's strengths.

2.2 Movie Processing

2.2.1 Shot Segmentation. A movie consists of a linear sequence of scenes, where each scene consists of several shots. A scene is a sequence of interrelated shots that share a common semantic thread, whereas a shot is a sequence of frames filmed by a single camera without interruption. Since scenes are segmented according to the semantic context [81], scene segmentation is relatively less accurate than shot segmentation. It is also quite inefficient to collect color palettes directly from the shot because many duplicated shots occur due to video editing. Therefore, to extract color schemes as we wish, we need a new unit that simultaneously satisfies desired accuracy and efficiency.

A master shot is a single shot that contains all the characters, representing the atmosphere of all the space being filmed. Modern movies use master shots during the production stage, but they tend not to be included in the actual movie due to their unappealing style. Therefore, we define a *semi-master-shot* rather than a true master shot, combining contiguous shots taken in the same location with similar colors.

Shot changes can occur abruptly or through a gradual transition. Abrupt transitions occur over a single frame due to camera switch; whereas gradual transitions, such as dissolve, fade-in, fade-out, and wipe, stretch over several frames with a variety of video effects. Shot segmentation aims to detect these transitions by grouping frames using image similarity.

Although most scene segmentation methods use multimodal features to reduce ambiguity, i.e., video, audio, and text, it is usually possible to detect shot segmentation transitions using video features alone, because the shot is filmed as a single take. Several video features rely on color histograms [4, 7, 11, 72] and local descriptors, such as SURF [2] and SIFT [9, 16]. We adopt Baraldi's shot segmentation methods [7] to define the movie's semi-master-shot.

2.2.2 Keyframe Extraction. The keyframe is the most representative among the set of frames covering the overall shot or scene. Similar principles of keyframe extraction is applied to the video thumbnail [33] and summarization [39, 79], which have become popular with the recent increase in machine learning applications.

The main issues for keyframe extraction are how many keyframes should be selected from a shot or scene and how to select these keyframes from the set. The appropriate number of keyframes can be as few as one [46, 61] or several [8, 73], depending on the extraction method. Simple and fast keyframe selection methods sample frames uniformly [8, 73] or randomly [62]. However, these sampling methods produce unstable selection results and are difficult to determine optimal frame sampling. Cost function thresholding [16, 77] addresses the drawbacks of sampling methods but requires considerable extra computational effort to calculate the cost function for a typical movie (~200,000 frames). We select a fixed number of keyframes from a semi-master-shot to reduce computational cost. Instead of uniform and random sampling, we introduce an objective function evaluating importance, clarity, and representativeness for each frame to ensure that the keyframe captures the entire shot contents well.

3 OVERVIEW

It is somewhat more difficult to extract a color scheme from video compared with a single image, and particularly for movies with hundreds of shots. Even if each image within the video or movie is down-sampled to (for example) 240×135 pixels, there exist ~30,000 colors in each frame and ~6 billion colors in a movie. Moreover, movies include complex sets of shots with wide ranges of color

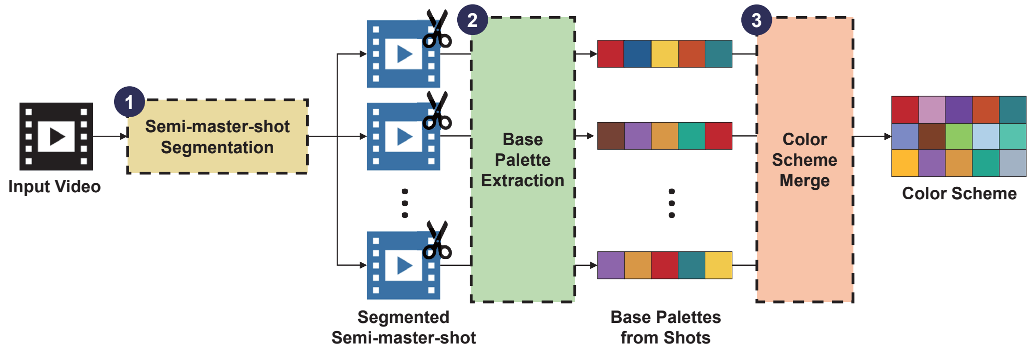


Figure 4: Proposed color scheme extraction modules: (1) semi-master-shot segmentation, (2) base palette extraction, and (3) color schemer merge

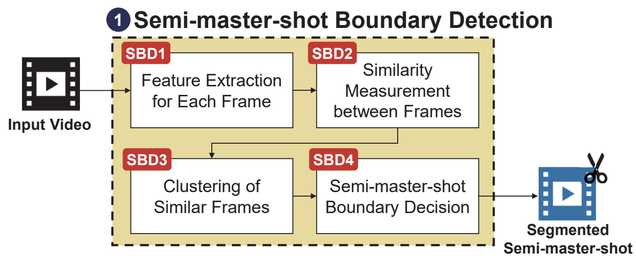


Figure 5: Semi-master-shot boundary detection (SBD)

tones. Therefore, we must consider color combinations appearing simultaneously in a shot as well as combinations throughout the movie. Previous color scheme extraction from movies has generally relied upon personal input from experienced designers. Therefore, we exploit the tendency for designers to focus on costumes, main objects, or salient backgrounds.

Previous color scheme extraction methods are generally based on evenly clustering pixels in all frames. However, as discussed in Section 1, simple clustering tends to promote darker colors or even black. This problem persists even when only keyframe pixels are evenly clustered, tending to select meaningless backgrounds and clutters. Therefore, we propose a color scheme extraction that incorporates three saliencies: the relative importance of the considered shot among the set of shots, the importance of the frame to the shot, and the importance of the pixel to the frame.

Thus, the proposed method includes three modules, as shown in Figure 4: (1) semi-master-shot segmentation (Section 4), (2) base palette extraction (Section 5), and (3) color scheme merge (Section 6). The resulting color scheme is clustered bottom-up from the segmented semi-master-shots.

4 SEMI-MASTER-SHOT BOUNDARY DETECTION

Semi-master-shot boundary detection (SBD) segments a movie into shot groups, combining contiguous shots taken in the same place with similar colors, to enhance segmentation accuracy and efficiency. SBD divides the video into a disjoint set of semi-master-shots, as shown in Figure 5.

Semi-master-shots are generally clustered by color difference using local descriptors for similarity factors, such as SIFT or SURF, which requires considerable computational overhead. In contrast, we adopt the Imagelab Shot Detector (ILSD) [7] segmentation method, which only considers RGB colors. ILSD measures the similarity between frames as the sum of two color difference metrics: squared difference between every corresponding pixel in two frames, and chi-squared distance of RGB color histograms. Similar frames are clustered using a sliding window to compare frame differences centered on the current frame, shifting in one direction.

Generally, ILSD detects abrupt and gradual transitions separately. The i-th frame, f_i, is regarded as an abrupt transition if the difference between f_i and f_{i+1} exceeds some threshold, T, and differences between neighboring shots exceed $T/2$. Gradual transitions are identified by repeating the process for detecting abrupt transition with increasing window size up to the maximum size of W. After shot detection, ILSD groups adjacent shots into scenes using hierarchical clustering. To prevent duplicate detection of the same transition, the two adjacent transitions are separated at frame intervals of more than a constant T_s, which is called the safe zone.

Although Baraldi et al. [7] group shots, which are segmented by ILSD, into a scene by clustering based on the color comparison, we cannot use the scene as a semi-master-shot for two reasons. First, they perform scene clustering using a fixed number of clusters, i.e., assuming it already knows the total number of scenes. Second, since the semi-master-shot does not require perfect scene segmentation, scene clustering increases computational overhead. Therefore, we acquire semi-master-shots mitigating T to determine the color difference between shots. To enhance the function of the safe zone, we use T_s^*, which is proportional to the average length of shots, instead of the fixed value of T_s (see Section 7 for details).

5 BASE PALETTE EXTRACTION

Base palette extraction (BPE) extracts the base palette from a segmented semi-master-shot, as shown in Figure 6. Although segmented, the semi-master-shot is still video, and it is challenging to extract a limited number of colors from a video containing too many frames. Therefore, BPE extracts a color palette in two stages: keyframe selection (Section 5.1) and color clustering (Section 5.2).

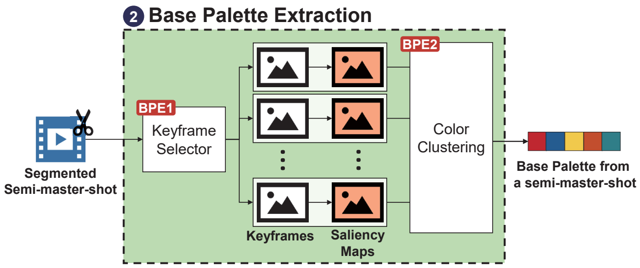


Figure 6: Base palette extraction (BPE)

We adopt a saliency map related to human visual attention for keyframe selection and color clustering. The saliency map represents the significance of each pixel as human fixation points. Since major pixels dominate movie impressions, rather than all pixels in every frame [78], we use the saliency map to help identify the optimal color palette from each frame image. Previous studies have applied saliency maps to color theme generation [30] or color theme extraction for fabric images [31]. We use the saliency map from a given image automatically [45], i.e., without pre-knowledge.

5.1 Keyframe Selection

The keyframe best represents a set of frames covering a shot or scene's overall contents. Using keyframes considerably reduces the amount of data to be clustered. We select a fixed number of keyframes from a semi-master-shot to further reduce computational cost. Instead of selecting keyframes uniformly or randomly, we introduce an objective function $C(f)$ to estimate importance, clarity, and representativeness for each frame f to ensure that the keyframe captures the entire shot contents well,

$$C(f) = \alpha_s C_s(f) + \alpha_c C_c(f) + \alpha_r C_r(f). \tag{1}$$

$C(f)$ is defined as a weighted sum of three terms: saliency ($C_s(f)$), clarity ($C_c(f)$), and representativeness ($C_r(f)$). Proper values for coefficients α_s, α_c and α_r will be suggested later based on our experiments to lead to a balanced result.

The saliency term $C_s(f)$ measures the importance of the frame with the average of saliency values of pixels in it as below:

$$C_s(f) = \frac{\sum_{p \in f} \mu_p}{|f|}, \tag{2}$$

where μ_p is a saliency of pixel p and $|f|$ is the total number of pixels in f. The saliency term enforces frames with high visual attention to be scored high.

The representativeness term $C_r(f)$ grades the coverage of f among all frames in the semi-master-shot. The representativeness can be expressed as a similarity in relation to other frames existing in the same semi-master-shot. The representativeness term is computed as below:

$$C_r(f) = \frac{\sum_{\forall f^* \in S} Sim(f, f^*)}{|S|}, \tag{3}$$

where S is the semi-master-shot that f belongs to and $|S|$ is the total number of frames in S. Similarity between frames f and f^*, $Sim(f, f^*)$, is defined by HSV histogram comparison and increased as f is more similar to many frames. We compute the pairwise

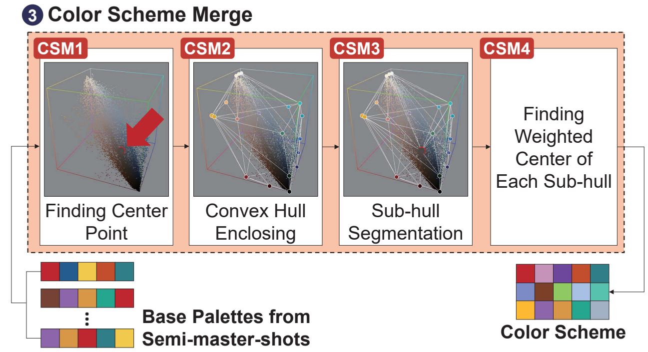


Figure 7: Color scheme merge (CSM)

distance between f and all other frames in the same semi-master-shot using correlation distance metrics of HSV histograms with 50 bins for each channel.

Clarity, $C_c(f)$ measures the clarity of f, i.e., the degree of blurring, using the blind/referenceless image spatial quality evaluator (BRISQUE) [36],

$$C_c(f) = 1 - 0.01 * BRISQUE(f). \tag{4}$$

The BRISQUE score is usually in the range [0, 100] and a smaller $C_c(f)$ indicates better image quality. We examine costs for all frames in the semi-master-shot and select the best three as keyframes.

5.2 Color Clustering

Color clustering extracts a base color palette from a set of selected keyframes by clustering pixels based on their distance in the color space, and selecting the cluster centroids as representative colors. Rather than clustering pixel coloring equally, we weight the cluster according to the pixel's visual attention using saliency maps from the keyframes.

We express saliency as a probability rather than the value itself to prevent overfitting the color palette from the saliency map. Pixel p from a keyframe is included in the clustering target depending on the probability $\psi(\mu_p)$, a weighted random function generating 0 or 1 with weight μ_p, providing the pixel saliency. This encourages including pixels with higher saliency in the clustering set, although low weighed pixels may sometimes be selected. We perform k-means clustering using pixels' RGB colors from all keyframes as a single set. The number of cluster centroids, k, is equal to the number of colors in a base palette, and we set $k = 5$ for our experiments.

6 COLOR SCHEME MERGE

Combining all base palettes extracted from semi-master-shots into the scheme raises two problems. We retain a large number of colors even in the reduced color set, due to the large number of semi-master-shots segmented from the movie; and shots taken in the same environment are sometimes separated in the movie during editing, causing overlapping palettes from different semi-master-shots. Therefore, we need a merge process to reduce the number of colors from the base palettes.

Base palette colors are cluster centroids, i.e., they frequently appear in semi-master-shots. Therefore, we perform additional clustering on base palette colors to derive a final color scheme that tends to ignore colors that may be distinctive but are distant from the centroids. The colors from each shot are not negligible because it is also selected in a large part of the film. Therefore, we achieve the merging using convex hull enclosing (CHE) [70] to generate a color scheme that includes all colors in the base palettes. We place the given colors in three-dimensional RGB space and generate a convex hull to enclose all color points. The convex hull is then simplified to a fixed number of vertices, while enclosing all color points.

However, if the convex hull vertices are used directly as the final color scheme, it is highly likely that the saturated colors are selected and the colors located in the middle of the convex hull are ignored. So, rather than using the vertices of the convex hull directly as a color palette, we segment the convex hull into sub-hulls and select the representative color of each sub-hull as the final color scheme, as shown in Figure 7. The center color is obtained by averaging all colors of the base palettes with the same weight, and this vertex is called v. We simplify the convex hull so that the number of faces in the convex hull mesh is equal to the final palette size π. We connect v and each triangular face to create π sub-hulls having the shape of a triangular pyramid. One vertex of sub-hull, which is closest to the weighted center of the sub-hull, becomes the representative color of the sub-hull. A weighted center is derived by weighting each color node of the sub-hull according to the length of the semi-master-shot where the node belongs. The center of each sub-hull becomes the color that forms the color scheme. This strategy mitigates the saturation of colors and also prevents the selection of non-existing colors.

7 RESULTS

We evaluated the proposed method on open video scene detection (OVSD) [60] and a commercial movie dataset. Current video datasets, e.g. RAI [5], BBC Planet Earth [6], and TRECVID 2019 [3], contain not only movies but also various other videos, including news, sports, documentary, surveillance and home video. OVSD is an open dataset of Creative Commons licensed videos freely available for download and use, comprising 21 short or full-length movies from various genres, e.g. drama, animation, crime, sci-fi, etc. OVSD is designed for scene detection, and the composition of the movie is aesthetically insufficient to evaluate the color scheme extraction. To compare the final color scheme results, we use a commercial movie dataset instead of the OVSD.

Although movies contain richer narrative patterns of shots and scenes compared with those general videos, there is no dataset targeting commercial movies due to copyright issues. We collect the commercial movie dataset particularly to compare current works by artists or descriptor roles. The dataset contains 53 movies from various genres, including *Inception, Atonement, Charlie and the Chocolate Factory, Days of Being Wild, Forrest Gump, Eternal Sunshine of the Spotless Mind, Her, La La Land, Life of Pi, Mary Poppins, Melancholia, Oldboy, Punch-Drunk Love, Spirited Away, The Avengers, The Grand Budapest Hotel, The Matrix, Your Name*, etc. The movies were encoded at 24 fps, resized 320 pixels wide, and opening and closing credits were discarded. All experiments were performed using a single machine running Ubuntu 16.04.1 LTS with i9-9900K Intel Core CPU, 32GB RAM, and NVIDIA GeForce RTX 2080 Ti graphics card.

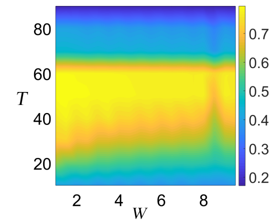


Figure 8: Average F1 score for W and T of OVSD. T_s^* differs for each OVSD movie depending on the movie length. We selected a combination of constants with the highest F1 score: $T = 50$ and $W = 2.5$.

	F1 score	Precision	Recall
ILSD [7]	0.738	0.623	0.905
Ours w/o T_s^*	0.771	0.642	0.964
Ours w/ T_s^*	**0.793**	**0.662**	**0.990**

Table 1: Average F1 score on OVSD. The proposed method including T_s^* outperforms ILSD segmentation and the proposed method without T_s^*.

7.1 Semi-master-shot Boundary Detection Accuracy

We used the F1 score [7, 74] to evaluate SBD accuracy, i.e., the harmonic mean of precision and recall. Precision measures correct detection coverage, i.e., whether the same semi-master-shot frames are grouped together, whereas recall measures semi-master-shot overflow, i.e., whether frames not belonging to the same semi-master-shot are erroneously included. We labeled OSVD semi-master-shots manually to calculate the F1 score, and experimentally determined optimal $T = 50$ and $W = 2.5$, as shown in Figure 8. Following the discussion in Section 4, T_s^* was determined by the input movie average shot length. Table 1 compares F1 scores calculated using ILSD, and the proposed method with and without T_s^*, and the proposed method including T_s^* outperforms the other two methods.

7.2 Keyframe Selection

We evaluate the significance of each term from Equation 1 by omitting each term in turn. We sampled a sequence of frames in the same semi-master-shot from OVSD videos, and examined frame costs to select the best three keyframes. Figure 9 shows the selected keyframes according to C_c, C_r, and C_s. The three coefficients are determined by trial and error as $\alpha_c = 0.5$, $\alpha_s = 0.6$, and $\alpha_r = 0.8$.

Figure 9(a) shows that omitting C_c selects blurry frames, which are inappropriate as keyframes. C_r encourages keyframes to have close visual characteristics with the same semi-master-shot frames. Figure 9(b) shows that a frame far from the whole flow is often selected without C_r. The middle frame in Figure 9(c) shows that low saliency images are likely to be selected even if the frames are selected from a similar region.

7.3 Color Clustering

We considered color clustering to extract a base palette from keyframes using saliency maps, which provide weighting of dominant color in the keyframes. To evaluate saliency map effectiveness, we compared color clustering outcomes with and without saliency maps,

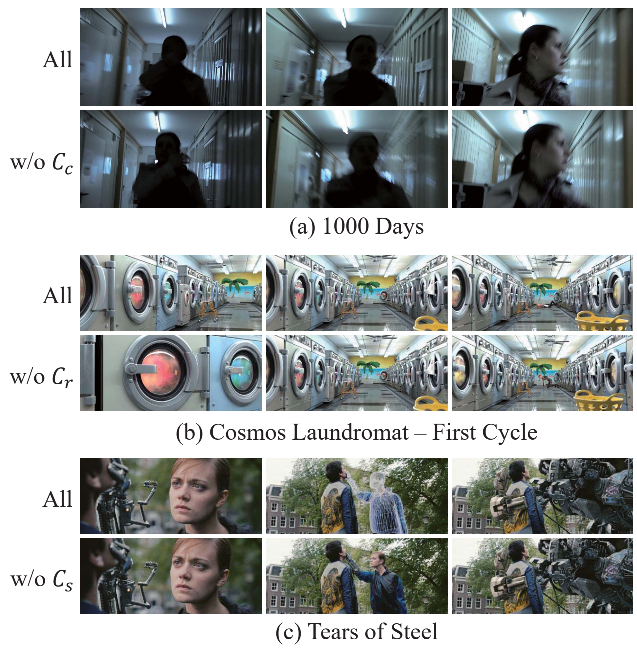


Figure 9: Keyframe selection results: (top row in each pair) selection results reflecting all terms, and (bottom row) results of omitting each term. Images taken from (a) 1000 Days ©copyright Revival Film, (b) Cosmos Laundromat - First Cycle ©copyright Blender Foundation | gooseberry.blender.org, and (c) Tears of Steel ©copyright Blender Foundation | mango.blender.org

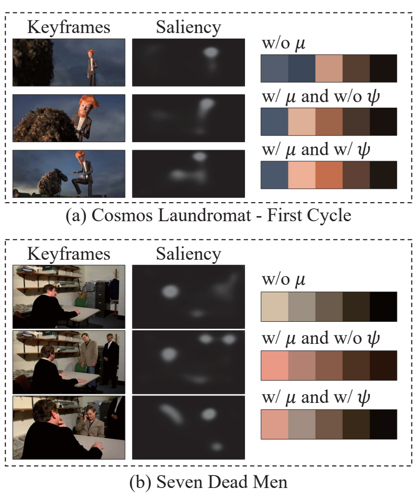


Figure 10: Color clustering based on saliency (μ) and probability function (ψ). Images taken from (a) Cosmos Laundromat - First Cycle ©copyright Blender Foundation | gooseberry.blender.org and (b) Seven Dead Men ©copyright 2009, Brett Koonce

and also by the probability function while applying the saliency map.

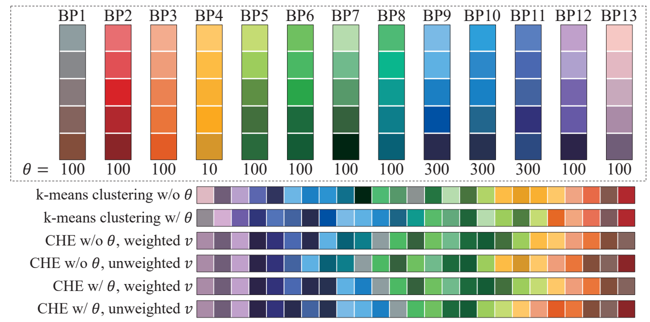


Figure 11: Color merging effects from semi-master-shot length(θ) and center point (v). Color schemes in the top two rows are generated from k-means clustering, and bottom four schemes are from the convex hull enclosing.

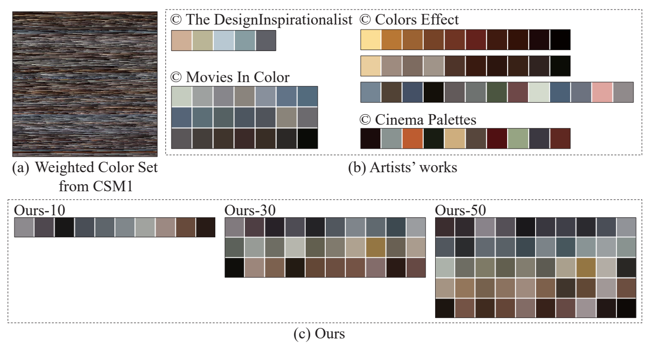
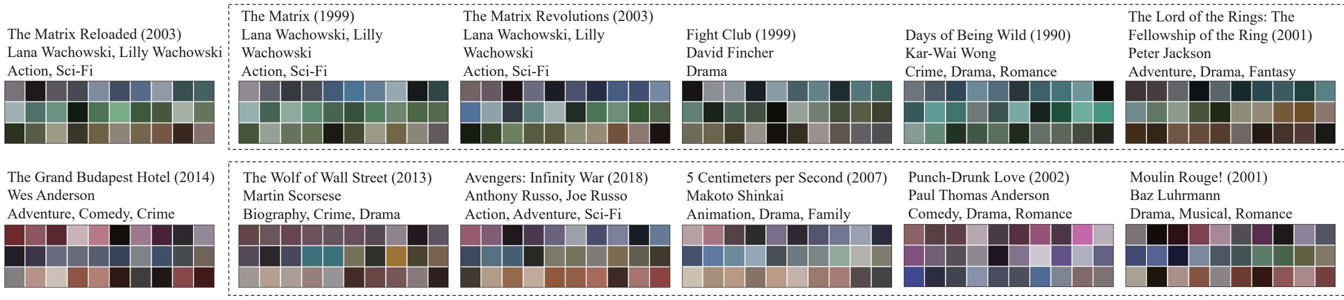


Figure 12: Designer [41, 54, 55, 59] and proposed method extracted color schemes from *Inception (2010), Christopher Nolan*, for various scheme sizes. (a) shows a color set to be plotted in RGB space by reflecting shot length in the CSM stage. The proposed schemes include all colors in the movie scene, whereas designer generated color schemes are from a specific scene, rather than the overall movie, and hence several important colors may be omitted, e.g. yellow series for *Movies In Color*, and blue series for *Cinema Palettes*. Although the artist has to repeat tasks to increase scheme color counts, the proposed method achieves the expansion easily due to its high scalability, i.e., the number of colors can be easily expanded by adjusting the number of vertices converging on the convex hull.

Figure 10 shows color clustering according to saliency (μ) and probability (ψ) functions from the given keyframes and their corresponding saliency maps, extracted from the OVSD videos. The color palettes were sorted by HSL color space in descending order. Since the absence of saliency makes all keyframe pixels evenly clustered, character colors, which tend to have high saliency, are improperly weighted (Figure 10(a)). If saliency is applied without the probability function, too many character colors overwhelm other backgrounds (Figure 10(b)). Thus, combining saliency and probability function produces more plausible results balancing characters and noticeable backgrounds.



(a) Query Movie (b) Top-5 Results

Figure 13: Top five movie retrieval results using color schemes. Movies unconnected through their meta information, e.g., genre or director, are commonly selected from their color scheme similarities.

7.4 Color Scheme Merging

We employed CHE to merge multiple base palettes to a single color scheme, addressing various merging issues, e.g. numerous colors to be reduced and overlapping palettes from different semi-master-shots, by weighting colors according to the related semi-master-shot importance. We adopted sub-hull clusters in RGB space and weighted each color vertex by its shot length, as shown in Figure 11.

Without θ, all colors in base palettes are equally reflected in the RGB space, so both k-means clustering and CHE generate a color scheme that is not relevant to the importance of the semi-master-shots. Even if the θ is reflected, k-means clustering generates a color scheme that ignores the color of the base palette with low θ, such as BP4 in Figure 11.

CHE tends to represent all colors of base palettes as possible. However, there is a difference between the results calculated by treating θ as weight in deriving v and the results that are not. If v is derived by weighting with θ, CHE extracts a color scheme that does not reflect the significance of palettes properly, despite the large proportion of blue (BP9-BP11) palettes. We suppose that this is because v is biased towards the position of colors with high θ, so the nodes with high θ are segmented into different sub-hulls. Therefore, the proposed method uses θ and unweighted v to merge the color scheme with color distribution considering the shot length while covering all base palette colors.

7.5 Color Scheme Extraction

Figure 1 shows typical outcomes from the proposed color scheme extraction on the commercial movie dataset. Figure 12 compares color schemes from the proposed and conventional designer schemes, confirming the proposed scheme has comparable plausibility compared to designer schemes, which require significantly more time and effort. Although the artist has to repeat the task to increase the number of colors, the proposed method expands the number of colors easily due to its high scalability, just by adjusting the number of vertices converging on the convex hull.

Figure 13 shows the proposed color scheme efficacy as a descriptor, considering evaluation of similarity from the commercial movie dataset. The similarity was measured as the sum of pairwise Euclidean distances between all pairs of colors from the two color schemes. The proposed scheme was able to select movies with unrelated metadata, e.g. genre or director, from their color scheme similarities. Our commercial movie dataset contains a series with similar colors, such as the Matrix Trilogy, to evaluate how well the color schemes match the mood of different movies. The query of *The Matrix Reloaded* successfully retrieved the remaining two series, and it implies that our method can play a role as a descriptor for movie retrieval.

8 CONCLUSION

Color reveals the overall mood of a movie by affecting human perception to elicit emotional responses. To the best of our knowledge, there is no system supporting the color scheme descriptor. This is because of the difficulty of extraction, not because of the lack of effectiveness. For example, users frequently share lists of movies with iconic color palettes in IMDb, and professional artists often produce and share color schemes of popular movies.

We proposed an automated color extraction scheme using saliency maps bottom-up from semi-master-shots. Since no single descriptor can ideally represent a movie, the best practice is to use as many descriptors as possible. In addition to the commonly used visual, textual and audio descriptors, color schemes also play an important role as a descriptor for movie summarization, retrieval, and indexing. The color schemes provide basic color information for producing posters, graphic designs, and teasers that capture the characteristics of a movie. It is also useful for video processing such as video recoloring, vectorization, and segmentation. It can also be used to check the color tone during the actual film production process.

Experiments involving dozens of real movies verified that the proposed scheme generates perceptually plausible color schemes. Future work will investigate weighting the relationship between colors in the same base palette during merging, and using hidden information to extract movie descriptors aside from colors, e.g. time series editing, costume design, cinematography, production design, etc.

ACKNOWLEDGMENTS

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2019-0-01158, Development of a Framework for 3D Geometric Model Processing)

REFERENCES

[1] Yağız Aksoy, Tunç Ozan Aydın, Aljoša Smolić, and Marc Pollefeys. 2017. Unmixing-Based Soft Color Segmentation for Image Manipulation. *ACM Trans. Graph.* 36, 2 (2017), 19:1–19:19.

[2] Evlampios E. Apostolidis and Vasileios Mezaris. 2014. Fast shot segmentation combining global and local visual descriptors. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*. 6583–6587. https://doi.org/10.1109/ICASSP.2014.6854873

[3] George Awad, Asad A. Butt, Keith Curtis, Yooyoung Lee, Jonathan G. Fiscus, Afzal Godil, Andrew Delgado, Jesse Zhang, Eliot Godard, Lukas L. Diduch, Alan F. Smeaton, Yvette Graham, Wessel Kraaij, and Georges Quénot. 2019. TRECVID 2019: An evaluation campaign to benchmark Video Activity Detection, Video Captioning and Matching, and Video Search & retrieval. In *Proceedings of TRECVID 2019*. National Institute of Standards and Technology (NIST).

[4] Lorenzo Baraldi, Costantino Grana, Guido Borghi, Roberto Vezzani, and Rita Cucchiara. 2016. Shot, Scene and Keyframe Ordering for Interactive Video Re-use. In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP*. 626–631. https://doi.org/10.5220/0005768706260631

[5] Lorenzo Baraldi, Costantino Grana, and Rita Cucchiara. 2015. Analysis and Re-Use of Videos in Educational Digital Libraries with Automatic Scene Detection. In *Digital Libraries on the Move - 11th Italian Research Conference on Digital Libraries, IRCDL*. 155–164. https://doi.org/10.1007/978-3-319-41938-1_16

[6] Lorenzo Baraldi, Costantino Grana, and Rita Cucchiara. 2015. A Deep Siamese Network for Scene Detection in Broadcast Videos. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, MM*. 1199–1202. https://doi.org/10.1145/2733373.2806316

[7] Lorenzo Baraldi, Costantino Grana, and Rita Cucchiara. 2015. Shot and Scene Detection via Hierarchical Clustering for Re-using Broadcast Video. In *Computer Analysis of Images and Patterns - 16th International Conference, CAIP*. 801–811. https://doi.org/10.1007/978-3-319-23192-1_67

[8] Lorenzo Baraldi, Costantino Grana, and Rita Cucchiara. 2017. Recognizing and Presenting the Storytelling Video Structure With Deep Multimodal Networks. *IEEE Trans. Multimedia* 19, 5 (2017), 955–968. https://doi.org/10.1109/TMM.2016.2644872

[9] Tamires T. S. Barbieri, Tiago Henrique Trojahn, Moacir Ponti, and Rudinei Goularte. 2015. Shot-HR: a video shot representation method based on visual features. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*. 1257–1262. https://doi.org/10.1145/2695664.2695841

[10] Huiwen Chang, Ohad Fried, Yiming Liu, Stephen DiVerdi, and Adam Finkelstein. 2015. Palette-based photo recoloring. *ACM Trans. Graph.* 34, 4 (2015), 139:1–139:11. https://doi.org/10.1145/2766978

[11] Tamires Tessarolli de Souza and Rudinei Goularte. 2013. Video shot representation based on histograms. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing, SAC*. 961–966. https://doi.org/10.1145/2480362.2480547

[12] J.B. Frost. 2009. *Cinematography for Directors: A Guide for Creative Collaboration.* Michael Wiese Productions.

[13] Timothy Gerstner, Doug DeCarlo, Marc Alexa, Adam Finkelstein, Yotam Gingold, and Andrew Nealen. 2012. Pixelated image abstraction. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering, NAPR*. Eurographics Association.

[14] Michael Gervautz and Werner Purgathofer. 1988. A Simple Method for Color Quantization: Octree Quantization. In *New Trends in Computer Graphics*. Springer, 219–231. https://doi.org/10.1007/978-3-642-83492-9_20

[15] M. Goodridge and T. Grierson. 2012. *FilmCraft: Cinematography.* Octopus.

[16] Rachida Hannane, Abdessamad Elboushaki, Karim Afdel, P. Nagabhushan, and Mohammed Javed. 2016. An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram. *International Journal of Multimedia Information Retrieval IJMIR* 5, 2 (2016), 89–104. https://doi.org/10.1007/s13735-016-0095-6

[17] F. Hannigan. 2013. *Filmcraft: Production Design.* Taylor & Francis.

[18] Paul S. Heckbert. 1982. Color image quantization for frame buffer display. In *Proceedings of the 9th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH*. 297–307. https://doi.org/10.1145/800064.801294

[19] Michael Hemphill. 1996. A Note on Adults' Color–Emotion Associations. *The Journal of Genetic Psychology* 157, 3 (1996), 275–280. https://doi.org/10.1080/00221325.1996.9914865

[20] Shih-Wen Hsiao and Cheng-Ju Tsai. 2015. Transforming the natural colors of an image into product design: A computer-aided color planning system based on fuzzy pattern recognition. *Color Research & Application* 40, 6 (2015), 612–625. https://doi.org/10.1002/col.21929

[21] Yu-Chen Hu and Ming-Gong Lee. 2007. K-means-based color palette design scheme with the use of stable flags. *J. Electronic Imaging* 16, 3 (2007), 033003. https://doi.org/10.1117/1.2762241

[22] IMDb. 2019. IMDb Statistics. https://www.imdb.com/pressroom/stats/ [Online; accessed 05-Jan-2020].

[23] Jason Hellerman. 2019. How a Film Color Palette Can Make You a Better Filmmaker. https://nofilmschool.com/Film-color-theory-and-color-schemes [Online; accessed 05-Jan-2020].

[24] EunJin Kim and Hyeon-Jeong Suk. 2015. Hue extraction and tone match: generating a theme color to enhance the emotional quality of an image. In *ACM Special Interest Group on Computer Graphics and Interactive Techniques Conference, SIGGRAPH, Posters Proceedings*. 7:1. https://doi.org/10.1145/2787626.2787657

[25] EunJin Kim and Hyeon-Jeong Suk. 2016. Key Color Generation for Affective Multimedia Production: An Initial Method and Its Application. In *Proceedings of the 2016 ACM Conference on Multimedia Conference, MM*. 1316–1325. https://doi.org/10.1145/2964284.2964323

[26] Yuki Koyama and Masataka Goto. 2018. Decomposing Images into Layers with Advanced Color Blending. *Comput. Graph. Forum* 37, 7 (2018), 397–407. https://doi.org/10.1111/cgf.13577

[27] Jan Eric Kyprianidis and Jürgen Döllner. 2008. Image Abstraction by Structure Adaptive Filtering. In *EG UK Theory and Practice of Computer Graphics*. 51–58. https://doi.org/10.2312/LocalChapterEvents/TPCG/TPCG08/051-058

[28] Deborah Nadoolman Landis. 2012. *FilmCraft: Costume Design.* Focal Press.

[29] Yanqiang Lei, Weiqi Luo, Yuan-Gen Wang, and Jiwu Huang. 2012. Video Sequence Matching Based on the Invariance of Color Correlation. *IEEE Trans. Circuits Syst. Video Techn.* 22, 9 (2012), 1332–1343. https://doi.org/10.1109/TCSVT.2012.2201670

[30] Sharon Lin and Pat Hanrahan. 2013. Modeling how people extract color themes from images. In *ACM SIGCHI Conference on Human Factors in Computing Systems, CHI*. 3101–3110. https://doi.org/10.1145/2470654.2466424

[31] Shiguang Liu, Yaxi Jiang, and Huarong Luo. 2018. Attention-aware color theme extraction for fabric images. *Textile Research Journal* 88, 5 (2018), 552–565. https://doi.org/10.1177/0040517516685278

[32] Shiguang Liu and Huarong Luo. 2016. Hierarchical emotional color theme extraction. *Color Research & Application* 41, 5 (2016), 513–522. https://doi.org/10.1002/col.21988

[33] Wu Liu, Tao Mei, Yongdong Zhang, Cherry Che, and Jiebo Luo. 2015. Multi-task deep visual-semantic embedding for video thumbnail selection. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. 3707–3715. https://doi.org/10.1109/CVPR.2015.7298994

[34] Yi-Hsiang Lo, Ruen-Rone Lee, and Hung-Kuo Chu. 2019. Generating Color Scribble Images using Multi-layered Monochromatic Strokes Dithering. In *Computer Graphics Forum (Proc. Eurographics)*, Vol. 38.

[35] Frank H. Mahnke. 1996. *Color, Environment, and Human Response: An Interdisciplinary Understanding of Color and Its Use as a Beneficial Element in the Design of the Architectural Environment.* Wiley.

[36] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Processing* 21, 12 (2012), 4695–4708. https://doi.org/10.1109/TIP.2012.2214050

[37] Bryan S. Morse, Daniel R. Thornton, Qing Xia, and J. Uibel. 2007. Image-Based Color Schemes. In *Proceedings of the International Conference on Image Processing, ICIP*. 497–500. https://doi.org/10.1109/ICIP.2007.4379355

[38] Yuki Nagai, Yusuke Uchida, Emi Myodo, and Shigeyuki Sakazawa. 2013. A color transformation method based on color theme that takes constraints on color ratio and spatial coherence into consideration. In *IEEE International Conference on Image Processing, ICIP*. 108–112. https://doi.org/10.1109/ICIP.2013.6738023

[39] Chong-Wah Ngo, Yu-Fei Ma, and HongJiang Zhang. 2005. Video summarization and scene detection by graph modeling. *IEEE Trans. Circuits Syst. Video Techn.* 15, 2 (2005), 296–305. https://doi.org/10.1109/TCSVT.2004.841694

[40] Pere Obrador. 2006. Automatic color scheme picker for document templates based on image analysis and dual problem. In *Proc. SPIE*, Vol. 6076. https://doi.org/10.1117/12.647075

[41] Melissa Oconnor. 2012. The Design Inspirationalist. https://thedesigninspirationalist.com/color-in-films-the-design-inspirationalist/ [Online; accessed 05-Jan-2020].

[42] Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. 2011. Color compatibility from large datasets. *ACM Trans. Graph.* 30, 4 (2011), 63. https://doi.org/10.1145/2010324.1964958

[43] Michael T. Orchard and Charles A. Bouman. 1991. Color quantization of images. *IEEE Trans. Signal Processing* 39, 12 (1991), 2677–2690. https://doi.org/10.1109/78.107417

[44] Li-Chen Ou, M. Ronnier Luo, Andrée Woodcock, and Angela Wright. 2004. A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application* 29, 3 (2004), 232–240. https://doi.org/10.1002/col.20010

[45] Junting Pan, Elisa Sayrol, Xavier Giró i Nieto, Kevin McGuinness, and Noel E. O'Connor. 2016. Shallow and Deep Convolutional Networks for Saliency Prediction. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. 598–606. https://doi.org/10.1109/CVPR.2016.71

[46] Rameswar Panda, Sanjay K. Kuanar, and Ananda S. Chowdhury. 2018. Nyström Approximated Temporally Constrained Multisimilarity Spectral Clustering Approach for Movie Scene Detection. *IEEE Trans. Cybernetics* 48, 3 (2018), 836–847. https://doi.org/10.1109/TCYB.2017.2657692

[47] Huy Q. Phan, Hongbo Fu, and Antoni B. Chan. 2018. Color Orchestra: Ordering Color Palettes for Interpolation and Prediction. *IEEE Trans. Vis. Comput. Graph.* 24, 6 (2018), 1942–1955. https://doi.org/10.1109/TVCG.2017.2697948

[48] posted to IMDb by user "ur31876393". 2015. 100 Most Influential Films of All Time. https://www.imdb.com/list/ls074063990/ [Online; accessed 05-Jan-2020].
[49] posted to IMDb by user "ur62332483". 2018. Most Iconic Use of Colors in Movies. https://www.imdb.com/list/ls024889906/ [Online; accessed 05-Jan-2020].
[50] posted to IMDb by user "ur62332483". 2019. Poll: Most Iconic Use of Colors in Movies. https://www.imdb.com/poll/Jd1sBb5kq3g/ [Online; accessed 05-Jan-2020].
[51] posted to IMDb by user "ur65884801". 2017. Color in cinema. https://www.imdb.com/list/ls064713143/ [Online; accessed 05-Jan-2020].
[52] posted to Instagram by user "cinemapalettes". 2016. CINEMA PALETTES. https://www.instagram.com/cinemapalettes/ [Online; accessed 05-Jan-2020].
[53] posted to Instagram by user "colorpalette.cinema". 2016. Color Palette Cinema. https://www.instagram.com/colorpalette.cinema/ [Online; accessed 05-Jan-2020].
[54] posted to Instagram by user "colorseffect". 2017. Colors Effect. https://www.instagram.com/ColorsEffect/ [Online; accessed 05-Jan-2020].
[55] posted to Twitter by user "cinemapalettes". 2015. Cinema Palette. https://twitter.com/CINEMAPALETTES [Online; accessed 05-Jan-2020].
[56] posted to YouTube by user "CineFix". 2017. 10 Best Uses of Color of All Time. https://www.youtube.com/watch?v=tILIeNjbH1E [Online; accessed 05-Jan-2020].
[57] posted to YouTube by user "Criswell". 2015. Colour In Storytelling. https://www.youtube.com/watch?v=aXgFcNUWqX0 [Online; accessed 05-Jan-2020].
[58] Jan Puzicha, Marcus Held, Jens Ketterer, Joachim M. Buhmann, and Dieter W. Fellner. 2000. On spatial quantization of color images. *IEEE Trans. Image Processing* 9, 4 (2000), 666–682. https://doi.org/10.1109/83.841942
[59] Roxy Radulescu. 2013. Movies In Color. https://moviesincolor.com/films [Online; accessed 05-Jan-2020].
[60] Daniel Rotman, Dror Porat, and Gal Ashour. 2017. Robust video scene detection using multimodal fusion of optimally grouped features. In *19th IEEE International Workshop on Multimedia Signal Processing, MMSP*. 1–6. https://doi.org/10.1109/MMSP.2017.8122267
[61] Daniel Rotman, Dror Porat, Gal Ashour, and Udi Barzelay. 2018. Optimally Grouped Deep Features Using Normalized Cost for Video Scene Detection. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, ICMR*. 187–195. https://doi.org/10.1145/3206025.3206055
[62] Klaus Schoeffmann, Manfred del Fabro, Tibor Szkaliczki, László Böszörményi, and Jörg Keckstein. 2015. Keyframe extraction in endoscopic video. *Multimedia Tools Appl.* 74, 24 (2015), 11187–11206. https://doi.org/10.1007/s11042-014-2224-7
[63] Martin Solli and Reiner Lenz. 2010. Color Semantics for Image Indexing. In *5th European Conference on Colour in Graphics, Imaging, and Vision and 12th International Symposium on Multispectral Colour Science, CGIV*. 353–358.
[64] Statista. 2019. Number of movies released in the United States and Canada from 2000 to 2018. https://www.statista.com/statistics/187122 [Online; accessed 05-Jan-2020].
[65] Eric J. Stollnitz, Victor Ostromoukhov, and David Salesin. 1998. Reproducing Color Images Using Custom Inks. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH*. 267–274. https://doi.org/10.1145/280814.280889
[66] Siti Noraini Sulaiman and Nor Ashidi Mat Isa. 2010. Adaptive fuzzy-K-means clustering algorithm for image segmentation. *IEEE Trans. Consumer Electronics* 56, 4 (2010), 2661–2668. https://doi.org/10.1109/TCE.2010.5681154
[67] Shamik Sural, Gang Qian, and Sakti Pramanik. 2002. Segmentation and histogram generation using the HSV color space for image retrieval. In *Proceedings of the 2002 International Conference on Image Processing, ICIP*. 589–592. https://doi.org/10.1109/ICIP.2002.1040019
[68] Jianchao Tan, Stephen DiVerdi, Jingwan Lu, and Yotam I. Gingold. 2018. Pigmento: Pigment-Based Image Analysis and Editing. *IEEE Trans. Vis. Comput. Graph.* (2018), 1–1. https://doi.org/10.1109/TVCG.2018.2858238
[69] Jianchao Tan, Jose I. Echevarria, and Yotam I. Gingold. 2018. Efficient palette-based decomposition and recoloring of images via RGBXY-space geometry. *ACM Trans. Graph.* 37, 6 (2018), 262:1–262:10. https://doi.org/10.1145/3272127.3275054
[70] Jianchao Tan, Jyh-Ming Lien, and Yotam I. Gingold. 2017. Decomposing Images into Layers via RGB-Space Geometry. *ACM Trans. Graph.* 36, 1 (2017), 7:1–7:14. https://doi.org/10.1145/2988229
[71] Khang Siang Tan and Nor Ashidi Mat Isa. 2011. Color image segmentation using histogram thresholding - Fuzzy C-means hybrid approach. *Pattern Recognition* 44, 1 (2011), 1–15. https://doi.org/10.1016/j.patcog.2010.07.013
[72] Dalton Meitei Thounaojam, Khelchandra Thongam, Kh. Manglem Singh, and Sudipta Roy. 2016. A Genetic Algorithm and Fuzzy Logic Approach for Video Shot Boundary Detection. *Comp. Int. and Neurosc.* 2016 (2016), 8469428:1–8469428:11. https://doi.org/10.1155/2016/8469428
[73] Tiago Henrique Trojahn, Rodrigo Mitsuo Kishi, and Rudinei Goularte. 2018. A new multimodal deep-learning model to video scene segmentation. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web, WebMedia*. 205–212. https://doi.org/10.1145/3243082.3243108
[74] Jeroen Vendrig and Marcel Worring. 2002. Systematic evaluation of logical story unit segmentation. *IEEE Trans. Multimedia* 4, 4 (2002), 492–499. https://doi.org/10.1109/TMM.2002.802021
[75] Holger Winnemöller, Sven C. Olsen, and Bruce Gooch. 2006. Real-time video abstraction. *ACM Trans. Graph.* 25, 3 (2006), 1221–1226. https://doi.org/10.1145/1141911.1142018
[76] Xiaolin Wu. 1992. Color Quantization by Dynamic Programming and Principal Analysis. *ACM Trans. Graph.* 11, 4 (1992), 348–372. https://doi.org/10.1145/146443.146475
[77] Yun Zhai and Mubarak Shah. 2006. Video Scene Segmentation Using Markov Chain Monte Carlo. *IEEE Trans. Multimedia* 8, 4 (2006), 686–697. https://doi.org/10.1109/TMM.2006.876299
[78] Fang-Lue Zhang, Xian Wu, Ruilong Li, Jue Wang, Zhao-Heng Zheng, and Shi-Min Hu. 2018. Detecting and Removing Visual Distractors for Video Aesthetic Enhancement. *IEEE Trans. Multimedia* 20, 8 (2018), 1987–1999. https://doi.org/10.1109/TMM.2018.2790163
[79] Ke Zhang, Wei-Lun Chao, Fei Sha, and Kristen Grauman. 2016. Video Summarization with Long Short-Term Memory. In *14th European conference on computer vision, ECCV*. 766–782. https://doi.org/10.1007/978-3-319-46478-7_47
[80] Qing Zhang, Chunxia Xiao, Hanqiu Sun, and Feng Tang. 2017. Palette-Based Image Recoloring Using Color Decomposition Optimization. *IEEE Trans. Image Processing* 26, 4 (2017), 1952–1964. https://doi.org/10.1109/TIP.2017.2671779
[81] Yingying Zhu and Dongru Zhou. 2003. Scene change detection based on audio and video content analysis. In *Proceedings Fifth International Conference on Computational Intelligence and Multimedia Applications, ICCIMA*. IEEE, 229–234. https://doi.org/10.1109/ICCIMA.2003.1238130