Ma CX, Song JC, Zhu Q *et al.* EmotionMap: Visual analysis of video emotional content on a map. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 35(3): 576–591 May 2020. DOI 10.1007/s11390-020-0271-2

EmotionMap: Visual Analysis of Video Emotional Content on a Map

Cui-Xia Ma[1,2,3], *Member*, *CCF*, Jian-Cheng Song[1,2,3], Qian Zhu[1,2,3], Kevin Maher[1,2,4], Ze-Yuan Huang[1,2,3], and Hong-An Wang[1,2,3], *Member*, *CCF*, *IEEE*

[1] *State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China*

[2] *Beijing Key Laboratory of Human-Computer Interaction, Institute of Software Chinese Academy of Sciences, Beijing 100190, China*

[3] *University of Chinese Academy of Sciences, Beijing 100190, China*

[4] *Academy of Arts and Design, Tsinghua University, Beijing 100084, China*

E-mail: cuixia@iscas.ac.cn; {songjiancheng19, zhuqian172}@mails.ucas.ac.cn; kevintmaher@gmail.com huangzeyuan@126.com; hongan@iscas.ac.cn

Received January 6, 2020; revised March 25, 2020.

Abstract Emotion plays a crucial role in gratifying users' needs during their experience of movies and TV series, and may be underutilized as a framework for exploring video content and analysis. In this paper, we present EmotionMap, a novel way of presenting emotion for daily users in 2D geography, fusing spatio-temporal information with emotional data. The interface is composed of novel visualization elements interconnected to facilitate video content exploration, understanding, and searching. EmotionMap allows understanding of the overall emotion at a glance while also giving a rapid understanding of the details. Firstly, we develop EmotionDisc which is an effective tool for collecting audiences' emotion based on emotion representation models. We collect audience and character emotional data, and then integrate the metaphor of a map to visualize video content and emotion in a hierarchical structure. EmotionMap combines sketch interaction, providing a natural approach for users' active exploration. The novelty and the effectiveness of EmotionMap have been demonstrated by the user study and experts' feedback.

Keywords video visualization, emotion analysis, visual analysis, sketch interaction

1 Introduction

While there is much research confirming that emotion plays a crucial role in gratifying users' needs during their experience of movies and TV series, there is little implementation of analytical systems to allow users to understand emotional content. The role of emotion in gratifying users' needs includes sensation seeking, self-reflection, vicarious kinds of experience, and mood management [1]. Depending on the participant and circumstance, intense and even negative emotion can be gratifying. Emotion affects and reflects the development of movie content and the mood of the audiences. Additionally, emotion corresponds to the highlights of movies and can help users quickly grasp the plot and determine whether to watch the entire content or not. Another important requirement is that users need to explore the emotional details of a plot and understand the emotion trends in movies according to their personal interests. Other researchers have observed that a system allowing users to understand emotional content would be valuable for users [2].

Video visualization interfaces assist users in understanding video content and have removed the burden of understanding videos [3]. How to effectively represent, analyze and interact with video content is important for video analysis. Most visual analytic tools for video content focus on the analysis of factual content in videos, such as video surveillance, sports video and movie content analysis [4–6]. However, these researches mainly focus on raising the efficiency and the reliabil-

Regular Paper

Special Section of CVM 2020

This work was supported by the National Key Research and Development Program of China under Grant No. 2016YFB1001200, the National Natural Science Foundation of China under Grant No. 61872346, and the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. 19080102.

ity of analyzing video content using the low-level video features. Less effort has been invested in visualizing the emotional content of videos at a cognitive level. In addition, they did not provide a system for users to explore video content by themselves interactively. Visualization keeps the user in the loop, and is a complementary technology to address the shortcoming of automated video analysis.

In this work, we present EmotionMap, an interactive visual analysis system for expressing video emotional content using a map as a visual metaphor as shown in Fig.1. Movies usually contain rich emotional content, thereby we choose movies as the resource to analyze in this study. First, we develop an annotation tool based on the circumplex model and Ekman's basic emotions [7, 8] that is used to collect users' assessments of movies. We also use facial expression recognition algorithms based on deep learning to get the characters' facial expression in movies. Then, we process the above two kinds of emotional data separately to model the emotion of videos. We map the data to 2D space and visualize video emotional data through a map. Integrating natural sketch interaction in the map, we provide an interactive system for video emotional content analysis. In order to evaluate the novelty and the effectiveness of EmotionMap, we conduct our evaluation from the qualitative and quantitative point of views. The experimental results show that our system performs well in video content analysis especially for the character and emotion analysis.

In particular, the paper makes the following contributions.

- *Emotional Content Extraction and Modeling.* We collect the subjective and the objective emotion data of several movies through users' assessments and automatic recognition algorithms. Then we model the emotional content of these movies based on emotion representation models.
- *Using the Metaphor of a Map for Video Content Analysis.* We create a novel form for video analysis by means of a map. It provides an efficient way for exploring video content, especially emotion. In particular it allows the understanding of the overall emotions at a glance while also giving a rapid understanding of the details.
- *Interactive System with Multi-Views and Natural Sketches.* We integrate natural sketch interaction in our system and construct an interactive visual analysis system with views that are interconnected, mainly including a map view, a character view, a video view, and a timeline view.

This paper is structured as follows. Section 2 reviews the relevant work in the field of video visualization, emotion analysis, and map-like visualization.

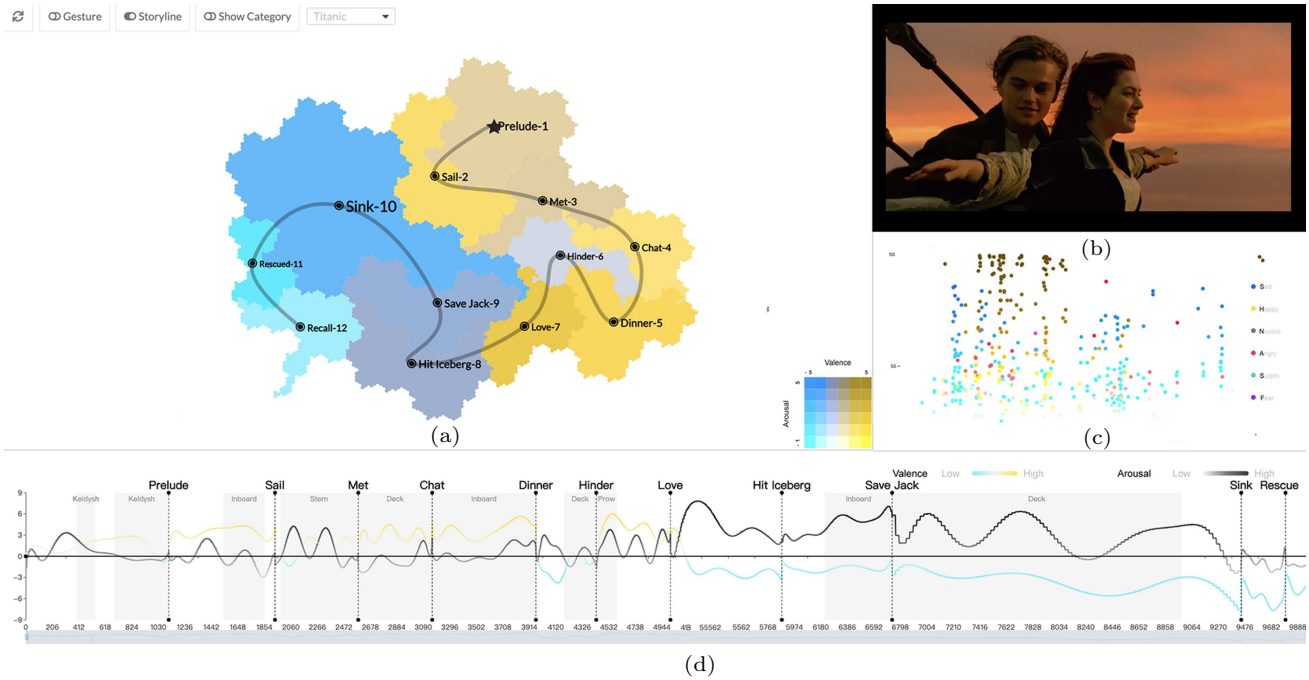


Fig.1. Overview of EmotionMap with four embedded views. (a) Map view, the main view that visualizes movie content using a map as a metaphor. (b) Video view, for video playback. (c) Character view, showing the emotion of selected characters. (d) Timeline view, showing the emotional data in linear time.

Section 3 introduces our design goals and visualization tasks. Section 4 introduces the process of collecting and modeling video emotional content. Section 5 describes the way to generate the map for video content analysis. Section 6 presents the interactive system for video analysis, which is integrated with sketch interaction and multi-view exploration. Section 7 verifies our system through expert interviews and a user study. Section 8 summarizes our conclusions, the limitations of our system, and future work.

2 Related Work

2.1 Video Content Visualization

Research work in video content visualization differs according to the application as well as the means of expression. Some use individual video frames to analyze video events [9, 10] while others use abstract visualizations to analyze contents [11, 12]. Some work integrates computer vision techniques and maps low-level vision features to more advanced semantics. These types of work mostly focus on applying visualization techniques to real-world scenarios such as surveillance videos, or sports videos used to assist people to browse and analyze videos effectively [4–6].

Another kind of research work commonly seen is visual analysis interfaces made for users to explore video content. Due to the development of storytelling in the field of visualization, some work analyzes video content in particular for movies through storytelling methods [15–17]. The recent work StoryCurves [16] proposed a visualization technique for communicating nonlinear narratives. The authors [16] constructed a "storycurve" from a sequence of events in narrative order and chronological order. In addition, Kurzhals *et al.* [17] proposed a system for visual analysis of film content including characters and scenes. Pan *et al.* [18] proposed an interactive video analytics tool for analyzing the contents of role-event videos.

Most of the current work we observe uses timelines as well as storyboard representation to support effective summarization of video content. However, there are also lots of more complex relations, such as character relations and event relations contained in videos. In some situations these relations need to be expressed more effectively. Moreover, in many cases emotion plays an important role in video visualization, and it may not effectively be captured by timelines or storyboards. Thus, we propose EmotionMap, an interactive system that visualizes video content especially for emotion using a metaphor map.

2.2 Emotion Visualization and Analysis

The semantic meaning of a given video clip is ambiguous, as the content of an individual clip can be perceived in many different ways. There are two different basic levels of content perception, a cognitive level and a perceived (or affective) level [19]. Understanding the emotional content is an important dimension for video analysis.

With the development of affective computing, related work about video emotion analysis continues to increase. Wang and Ji divided video affective content analysis into two approaches: direct and implicit [20]. Most of the direct approaches are concentrated in using computer vision methods to analyze emotional features. They extracted a range of features from video frames and learned the features related to emotion [21, 22]. Considering the multi-modal content in videos, many studies also created multi-modal analysis frameworks for video emotion analysis [23, 24]. However, these studies mainly focused on the low-level features to analyze and express video emotion. Less effort has been devoted to visualization of video emotion. Zhao *et al.* [25] presented video affective content in the form of a summary by learning audio-visual features. Lan *et al.* [26] created video summarization from an emotional perspective. However, their work has no detailed analysis of emotional content combined with events and characters. They added emotion expression as a part of a video summary, without visual analysis on emotion. There are also some studies which can be applied to interactive systems in order to communicate emotions. Seemo is a novel embedding framework, which allows mapping of human emotions into vector space representations [27]. Huang *et al.* [28] designed a novel location-based mobile social app, improving people's awareness and regulations of their emotions.

Current work related to video emotional analysis mostly focuses on feature learning, including the features from images, audio, or texts. There is little work that focuses on the visual analysis of video emotional content.

2.3 Map-Like Visualization

The sense of space is an important component of human cognition. There are some visual analytics studies that use a map as a metaphor to present data. Many

studies use maps to represent non-spatial information, especially for network data and user events in social media [29–32]. Some researchers focus on using maps to express hierarchical data, such as file systems and library catalogues [33, 34]. Other work is devoted to generating more natural, realistic maps to carry the map metaphor further [35–37]. Additionally, map-like visualization is also used to visualize large volumes of data and dynamic data. For example, cartography was used to express the vast amount of world knowledge encoded within Wikipedia and created thematic maps of almost any kind of data [38]. Mashima *et al.* [39] explored a way to visualize large-scale dynamic relational data with the help of a geographic map metaphor.

While these studies are good at presenting different kinds of data through map-like visual metaphors, they focus less on video content visualization. Inspired by the work in the field of multimedia [40], which presented video content by taking advantage of the metaphor map, we propose EmotionMap, visualizing video emotional content by means of a map as a metaphor. We develop an interactive system combined with natural sketch interaction and other interface elements to explore video emotional content.

3 Design Goals

The overarching goal of our work is to develop a novel visualization technique to allow users to explore, understand, and search movie content through the angle of emotion. In order to design such techniques, we put forward the following design goals.

*G*1: *let users understand the relation between events, scenes and roles and the corresponding audience and character emotions.* There are many types of information in movies. In addition to the visual and audio information that we are familiar with, movies also contain information such as the key events in the plot, the characters, and the emotion. This information is interrelated and together conveys the central idea (and content) of the film. For example, the whole movie can be divided into different key events, and each event has a centralized theme participated by different movie characters, all occurring in many different scenes. Emotion plays a key role in conveying the content of a movie so that grasping the emotion information will make it easier for users to understand the relationships of the other kinds of information in the film.

*G*2: *show the proportion and magnitude of emotions. The same audiences will feel different emotion when they watch different movies.* Comedy movies convey positive emotion, while horror movies can be frightening and scary. The emotion experienced by the audience can reflect the type and the quality of the movie. If we can visually see the emotional information of the audience by means of its composition and emotional intensity for a given movie, we might be able to more easily explore, choose, and analyze clips we want to watch.

*G*3: *present the emotional change in time.* The change of emotion often corresponds to the development of the movie plot. For example, in the climax of the movie, the audience's emotion is relatively intense. Understanding the emotional changes over time in the film helps to understand the development of the story.

*G*4: *help users retrieve movie content with emotional information* (*data*)*, and play it rapidly.* A movie usually contains one and a half hours to two hours. When rewatching a movie again after a while, it is difficult for us to quickly find the corresponding segment based on our own impressions. Perhaps we forget the specific content in that segment of the movie, but we remember the emotions experienced at that time. With the help of emotional information, we may quickly find the clips we want and immediately view them.

*G*5: *allow users to compare the overall emotional content of different films.* There are lots of movies on the market at present and we often feel troubled to choose the movies we like. If we can quickly understand the emotional tends and intensity of movies, we can choose the movie we want to watch without viewing spoilers. Thus we can get the expected emotional experience we desire. Based on these design goals, we propose several visualization tasks in Table 1 to satisfy these requirements. These goals and tasks are kept in mind for the creation of EmotionMap.

4 Data Collection and Emotion Modeling

One step in video emotional content visualization is to decide how to collect emotion. In this study, we use several movies as sources and collect two types of data to represent the emotion of movies. We use them as the subjective and the objective data. The subjective data is the assessments from 100 trained participants while the objective data is the facial expression recognition results based on a state-of-the-art deep learning method. Subjective emotion can directly reflect audiences' emotion when watching movies, and it well shows the emotion that the film director wants to con-

Table 1. Visualization Task

Task	Visualization Task	Corresponding Goal
VT1	To present temporal proportion and change of emotional data	G2, G3
VT2	To find the relation between temporal emotion data and video content	G3
VT3	To show the relations among events, emotion, roles and scenes	G1
VT4	To enable rapid video browsing guided by emotional information	G4
VT5	To give users information about the upcoming plot	G3
VT6	To display data through different views and multiple scales supporting user interaction	G1, G4
VT7	To identify the movie type and the overall emotion at a glance	G2, G3, G5

vey through the movie. Objective emotion is the emotion shown by the actor during the performance. With the help of objective emotion, users can explore the relationship between the characters. Subjective and objective emotion can sometimes be not consistent. For example, when a villain is happy about his/her success, the subjective emotion could be disgusted. The reason why the emotional data is collected manually is that there is no effective technology to accurately collect people's subjective emotions, and visualization requires high-qualified data. It is believed that the future development of technology will make this process more convenient. We will introduce the process of collecting and modeling emotional data in this section.

4.1 Emotion Expression and User Emotional Assessment

Emotion is different from factual information in that it is difficult to measure accurately. Emotion can be expressed in a discrete way as well as continuous representation based on different emotion representation models. The categorical and the dimensional approaches are the most commonly models for emotion analysis in psychology [41]. Users' assessments are usually used to represent the emotional information of videos in psychological experiments. Movies are commonly used as resources in these experiments. We select several films and divide them into key events, and then use these as our experimental resources. We develop a tool for collection of user assessments, EmotionDisc. EmotionDisk is based on the circumplex model and Ekman's six pancultural basic emotions [7, 8] and it is efficient for collecting the audience's emotional information. The prototype of our tool is the circumplex model in Fig.2.

The emotional content of a given video clip can be defined as the intensity and the type of emotion expected to arise while the user is watching the clip [21]. Arousal can be defined as the intensity of emotion while valence is the type of emotion. In the EmotionDisc shown in Fig.3, valence is extend-

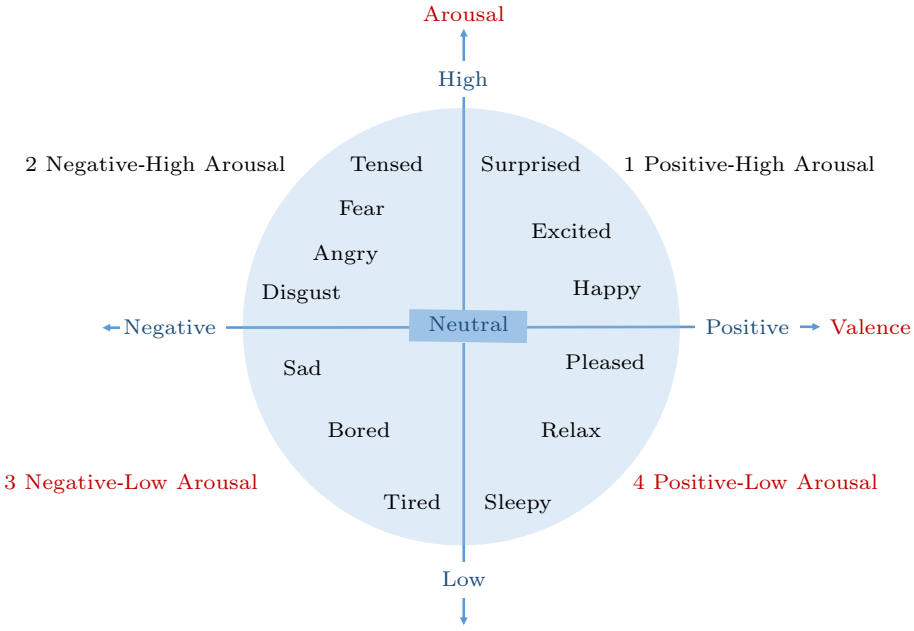


Fig.2. Emotional model of continuous space which includes discrete emotion words.

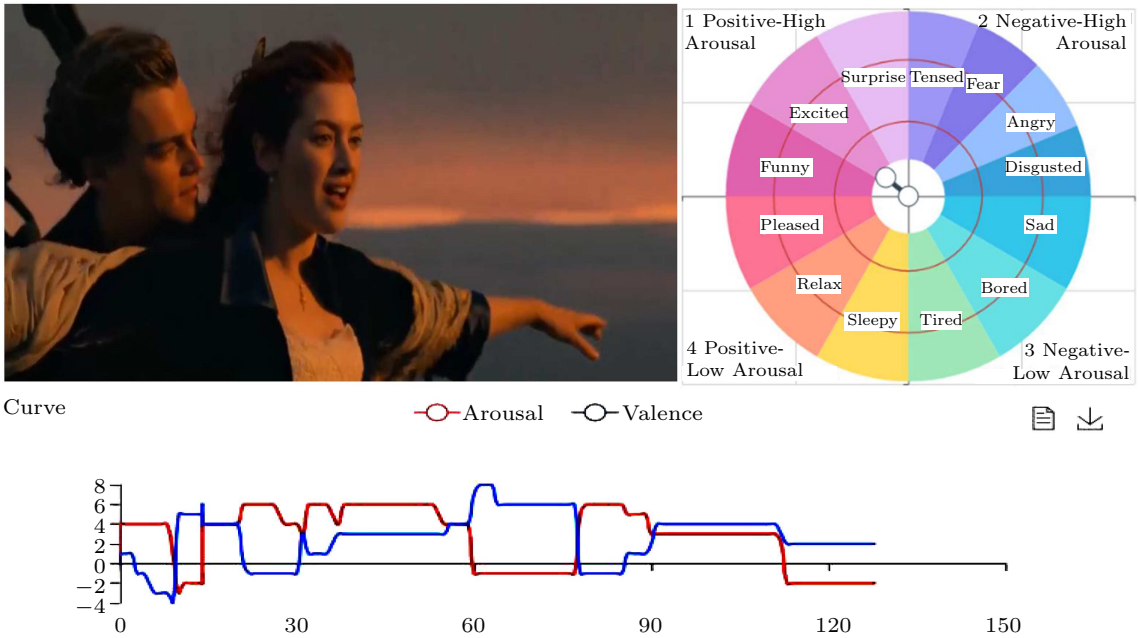


Fig.3. Interface of EmotionDisc.

ing from "pleasant/positive" to "unpleasant/negative", while arousal is ranging on a continuous scale from "energized/excited" to "calm/peaceful". We map emotion types on the EmotionDisc for users to choose. Users can drag the pointer to choose the emotion while they are watching the movie. And they can control the intensity of selected emotions by adjusting pointer radius. The curves below record the values of arousal and valence corresponding to the selected emotion on the disc next to the video. When users finish watching the movie, the data collection process is completed. The tool has been approved by the experts in psychology and some improvements have been made based on the advice of experts. Combining the circumplex model and six basic emotions, we select 14 emotion words including neutral emotion and map them to the dimensional space supported by the arousal and the valence in Fig.3. We use the warm colors to indicate positive emotions while the cold colors indicate negative emotions.

We recruit 100 participants to record the emotion of videos. They are trained about 5 minutes to familiarize the operation and emotion types on the EmotionDisc, and a test video is arranged for them to practice before the experiment. During the experiment, participants are asked to select the emotion type based on their feelings while they are watching the video. Since the emotions on the EmotionDisc correspond to the dimensional emotion space, we can get both the arousal and the valence data at the same time. When in use, EmotionDisc allows researchers to collect the emotional type the users record, as well as the corresponding valence and arousal data. The effectiveness of EmotionDisc is that we can get two types of emotion data at the same time without the participants having a professional understanding of emotional space. As participants have different levels of familiarity and preference about the movie, in order to balance these two parts and ensure the overall quality of data, we require each participate to score their familiarity and preference of the video clip before experiments. We make the following assumptions before data processing.

1) The more familiar participants know about the video, the more accurate the emotional assessment is made.

2) If participants prefer the video content, there may exist a certain personal tendency to make the assessment.

We use F and P to represent the familiarity and preference score for each participate respectively. And C stands for the emotion type that participants chose. The method we use to calculate the emotion type from the 100 participants is the following:

$$E_i = \sum_{n=1}^{100} \left(\left(\frac{F}{P} \right)_n \times C_{ne} \right), \quad C_{ne} = \begin{cases} 1, & \text{if } e = i, \\ 0, & \text{otherwise}, \end{cases}$$

$$T_s = \arg\max_i E_i.$$

The contribution score of the n-th participant can

be expressed as $(F/P)_n \times C_{ne}, n \in [1, 100]$. C_{ne} means the n-th participant chooses the emotion e. And e, i denote the emotions in [surprised, excited, happy, pleased, relax, neutral, tired, bored, sleepy, sad, disgust, fear, angry, tensed]. $(F/P)_n$ is the weight of the n-th participant. E_i represents the count of each emotion by second from 100 participants and T_s means the selected emotion type for the s-th second. Finally, we get a stable emotional category list for each video clip that changes over time. For the arousal and the valence data, we visualize both in the 2D space and use the Polynomial Regression algorithm to fit the curve. We set the Dof (degree of freedom) as 2. Finally, we get the VA (Valence & Arousal) curve and corresponding emotion type over time. Together the data gives a more comprehensive view of the emotional content.

4.2 Facial Expression Recognition

Facial expressions can provide valuable clues to the emotions a character in a movie is experiencing. We use the following methods to obtain character facial expression data.

• Deep learning methods to accumulate all character facial expression data.

• Deep learning facial recognition methods to assign the facial recognition data to each movie character.

For the facial expression emotional classification, we use the model developed by Arriaga *et al.* [42] The model classifies the six pancultural basic emotions proposed by Ekman. The model has accurately classified 66% facial character expressions in the 2013 Kaggle Facial Expression Recognition Challenge (the winning entry achieving the accuracy of 71%).

In the next step we assign emotional data to each of the characters in the film by using the state-of-the-art face recognition built with the deep learning dlib library①. The algorithms have an accuracy of 99.38% on the Labeled Faces in the Wild [43] benchmark.

While subsequent research has incrementally surpassed these accuracy benchmarks in facial expression and facial recognition, we find the accuracy sufficient, and the data collected to be valuable for users in exploring movie content. Furthermore, given the increasing accuracy of emotion recognition algorithms, in future implementations of similar systems, the value of the emotional data collected by automated methods is likely to increase.

5 Visual Design

In order to express a story clearly, we need to consider four aspects: where, when, who, and what. In the context of movies, we need to tell an audience what has happened, where it happened, when it happened, and who was involved. It is difficult for time linear visualization to show these four aspects at the same time. In this section we discuss how the user goals discussed in Section 3 can be aided by an emotion map.

We will show the details of the movie content in our map. The goal of storytelling in visualization tends to integrate complex information in an intuitive form [15]. The cartographic map metaphor has been in use for a while because it is a very intuitive way to represent the information. Many examples have been demonstrated to use this metaphor [31, 44]. A metaphor map is a visual expression which imitates the map of non-spatial information (semantic information) (VT3). A metaphor map takes advantage of the potential human experience to promote the expression and the excavation of data in the form of spatial objects. In addition, map-like visualization saves space and gives an intuitive view of the proportion and the size of the elements (VT1, VT7).

In this work, we propose to visualize movie content using a metaphor map. We expand our EmotionMap based on the gosper map [35], combining with multi-scale semantic zooming and interactive operations, to analyze movie content with emotion (VT6). A gosper map is generated according to fractal rules, is stable and can maintain a certain shape when the content of the movie changes.

5.1 Map Generation

As shown in Fig.4, we give an example of how to build our movie metaphor map according to David's work [37]. Firstly, as shown in Fig.4(a), we divide a movie into a hierarchical structure. We can explore movie content easily by this structure. The movie is divided into many events and each event contains at least one kind of audiences' emotion. Each parent emotion node is composed of leaf nodes that each has the same emotion as the parent, and each leaf node represents one second in the movie. The leaf nodes with the same parent are arranged in chronological order as shown in Fig.4(b). This hierarchical structure lets nodes aggregate according to the priority of events, emotions and time, and to some extent retains the order of time (VT1, VT2). At the same time, emotional clustering allows us

①https://pypi.org/project/face-recognition/, May 2020.

to see more clearly the main emotions in an event. Secondly, as shown in Fig.4(b), we generate a Gosper curve in space. The Gosper curve is composed of 2D space-filling curves, which are often used to generate fractals, and hexagons are positioned along the Gosper curve by the order of leaf nodes so that the map is formed. Details about how a Gosper curve generates the map are illustrated in [37]. Thirdly, we map the nodes in Fig.4(a) to hexagons and construct a metaphor map of the movie. Thereby the movie is divided into seconds and each second is corresponded to a hexagon in the map. The map size of a movie depends on its duration, and it means that a bigger area in the map represents a clip of the movie which has a longer time.

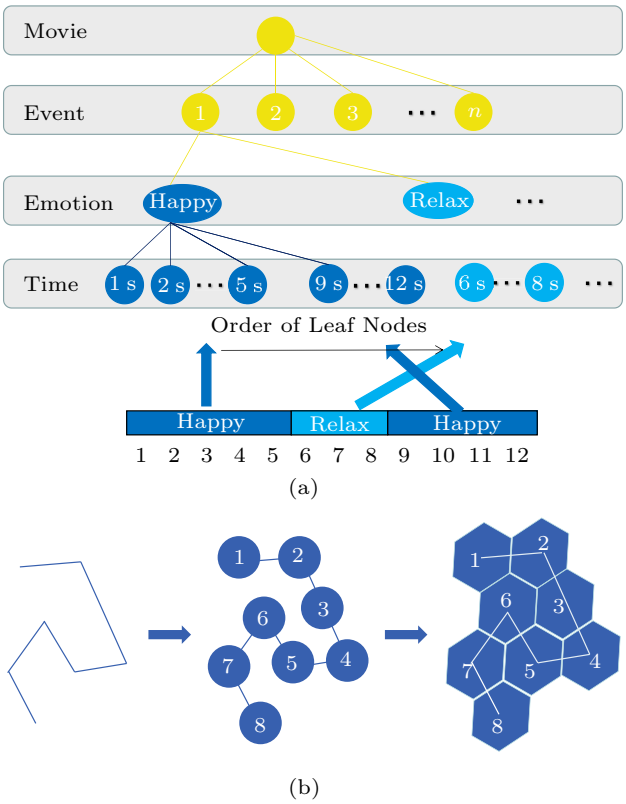


Fig.4. Process for constructing the movie metaphor map in our work. (a) How to convert original emotion data sequence of a movie to hierarchical structures. (b) How to arrange data on the Gosper curve.

As the hexagonal layout is ordered, the video content is represented on the map by events, emotion and time. The method of generating maps from 2D space-filling curves also ensures that the hexagon with the same emotion, mood and adjacent time is adjacent in space. Utilizing this characteristics of the Gosper curve, we can divide a movie into different events by dividing the map. Then we fill the hexagons with different colors according to the valence and the arousal data of the leaf nodes. When users understand the meaning of the valence, the arousal and colors, they can easily identify the emotional information in a movie (VT1). In addition, based on the above structure, the nodes which are adjacent in time but have different emotion may not be adjacent in space, which reflects that the sudden change of emotions corresponds to the change of spatial locations (VT5).

5.2 Color

An ideal mapping would allow an audience to understand without constant reference to the color legend. In Western color symbolism, yellow is commonly associated with positive "happier" feelings and blue with more negative "sadder" feelings, as reflected in many popular emotional color scale② [45]. A challenge to map colors to the valence using yellow and blue is that the two colors as displayed in common colors such as HSB or RGB have lightness values perceived to the human eye that vary widely. This conflicts with the "darker is more" bias noted by researchers in data visualization color mapping, and the effect may mislead observers [46]. We use a color scale CIELAB that is calculated to match the perception of the human eye. For mapping the arousal in an intuitive way, as shown in Fig.5, we decide to use CIELAB's lightness scale from low arousal events as white and high arousal events as dark, as first, the perceived lightness can be applied independently of the valence or hue, and second, high arousal events in the context movies can be emphasized, and on a light background the high arousal events can be seen as fitting the "darker is more" perceptual bias.

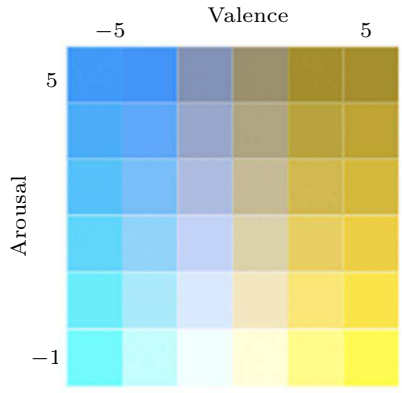


Fig.5. Color mapping used for understanding the valence and the arousal of films, aimed to be intuitively understandable.

②http://atlasofemotions.org, Mar. 2020.

As shown in Fig.6, we find that using this color system was able to provide a strong basis for users to understand the overall emotions in films at a glance. Furthermore, the colors are useful for more detailed analysis (VT7). In Fig.7, maps from three films are compared. From the left are the consistently high valence and varied levels of arousal in Tais-Toi!, and next the predominantly low valence and high arousal of Titanic, followed by the varied valence and arousal of CJ7.

(a) (b) (c)

Fig.6. Three maps of different films. (a) Tais-Toi!. (b) Titanic. (c) CJ7.

5.3 Visualizing a Movie onto a Map for Analysis

In this subsection, we will introduce how we map a movie into a metaphor map and analyze the content of movies with emotion by taking the movie Titanic for example. Maps are generated based on the algorithm described in Subsection 5.1, and their shapes are related to the video's duration based on a Gosper curve that represents time order to a certain extent. As shown in Fig.7, we visualize the movie Titanic in three kinds of map views. They help users to explore video content and emotions in different perspectives.

The event map in Fig.7(a) divides the movie Titanic into events and uses different colors to distinguish them by our method. Events are colored based on the method in Subsection 5.2, and the values of arousal and valence are the average of the whole event. The size of each event represents their duration in the movie. The map is made up by many hexagons that we have introduced in Subsection 5.1 and each point on the map represents the moment of one second length in movies. According to the attribute of a Gosper curve which has been illustrated in [8], the hexagons inside an event are first arranged in terms of similar emotion, and then inside the regions with similar emotion, the hexagons that are adjacent in time order are together. Therefore each event block in a map is composed by a group of adjacent hexagons in time order. In order to let users better grasp the chronological order, events are organized in series through a red story line. Users can explore the narration of the movie by this story line. The position

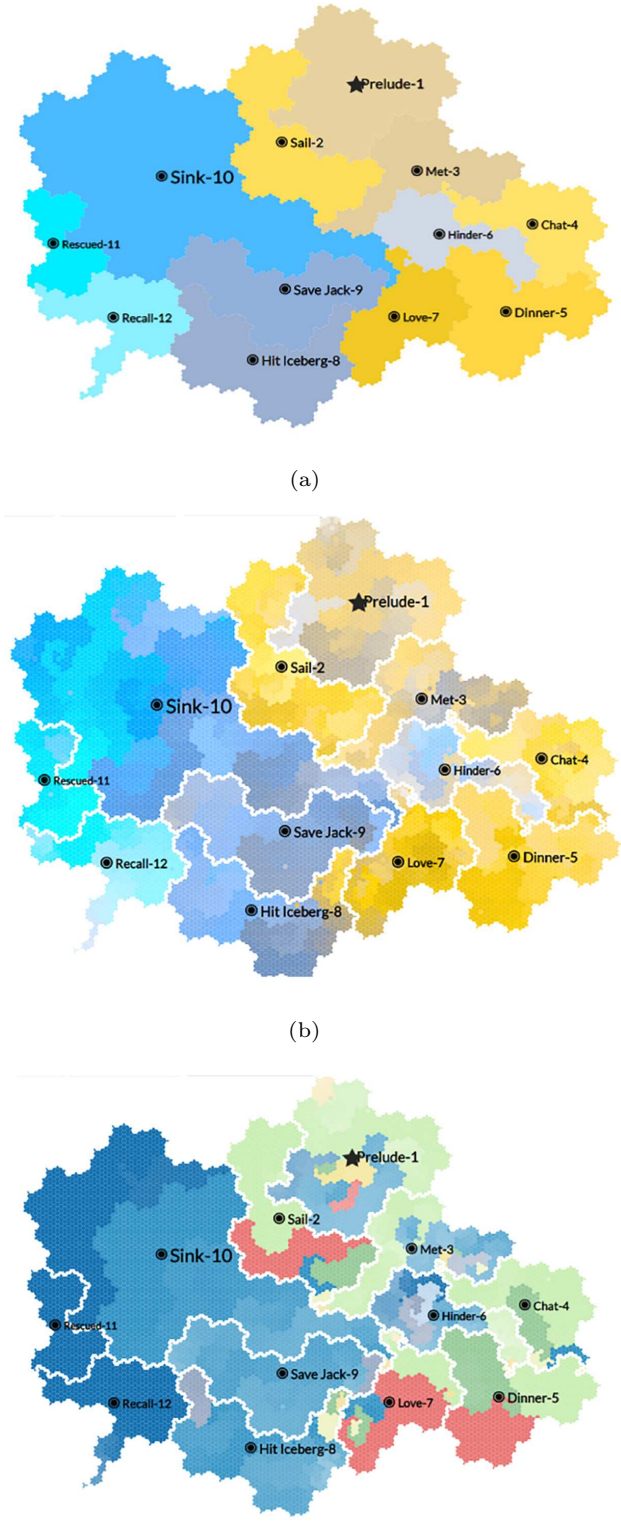


Fig.7. Event map views. There are two ways for exploring video content through event view. (a) The first level that divides a video into events. (b) Detailed level showing valence and arousal in each event. (c) Maps showing the type and intensity of emotion distribution.

of each event point is determined by the closest leaf node occurring chronologically halfway in the corresponding event. Users can explore video content based on the story line and they can choose any event points on a map to browse the corresponding video clips in our system. For a more detailed point of view, we provide the function of semantics zooming like in real maps where we can see the provinces on the macro level and we can see the cities in each province by zooming on the map. Similarly, by zooming the map in Fig.7(a), the map will show the emotional values (valence, arousal) or categories in Figs.7(b) and 7(c) according to users' personal preferences (VT6).

In the view of Fig.7(b), we color the map according to the arousal and the valence results collected in Subsection 4.1. Each hexagon has been colored by their arousal and valence values and a legend is provided next to each map. The map seen in Fig.7(b) expresses the arousal and the valence over time. By observing the distribution of emotions on maps, we can know what happened and the emotional similarity of different parts of the movie.

Visualizing emotions on a map can help users find the highlights of video and analyze the story easily. For instance, in Fig.7(b) we can see that the movie Titanic contains predominantly sad emotions, and can easily identify it as a tragedy (VT7). In detail, the emotion tends to be positive before the event called Hit Iceberg. The incident happened in this event and the emotion after the incident had a negative turn. We can speculate that the mood of people is mostly negative while watching the movies after the incident Hit Iceberg. We can also observe that there is lower arousal before the event Hit Iceberg and it becomes higher after the event by the change in the brightness of color. We can infer that the intensity of audiences' emotion is much higher because the Titanic is going to sink (VT4). We can see that the higher arousal of the movie is concentrated in two events (VT5). One is that Jack and Rose hug at the stern while another is the shipwreck event. The arousal of these two events is high but the valence of them is positive and negative respectively.

To see the emotional categories in more detail, in the view of Fig.7(c), each event is divided by a few color pieces that represent different emotion categories according to the data collected in Subsection 4.1. The colors of different blocks are defined by the emotion results. The bright colors represent positive emotion types and the cold colors stand for the negative emotion relatively. The shade of color represents the intensity of emotions.

5.4 Character Emotions and Metric Visualization

Character facial expressions can provide valuable emotional information about the plot, and be useful for users to explore and retrieve content (VT4). As shown in Fig.8, we plot the categorized character emotions on a scatter plot with the probability of the emotions over time, showing a wide change in character emotions (VT1).

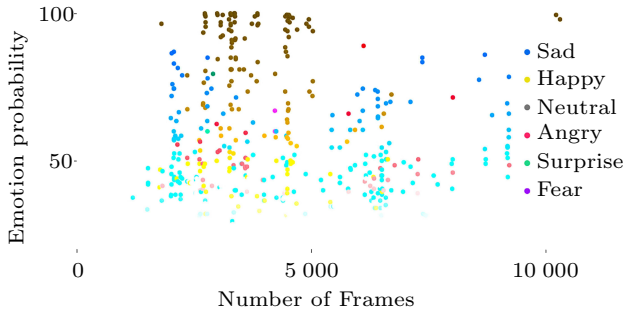


Fig. 8. Overview of the frequency of where characters appear in a film as well as the probability and the frequency of their emotions.

Since character emotion may be useful for retrieving content, we allow the interaction for filtering the emotions. Given that users may need additional visual cues for finding clips, as shown in Fig.9, the faces are plotted on the dots in the scatter plot, so that the visual clues to the content are not lost.

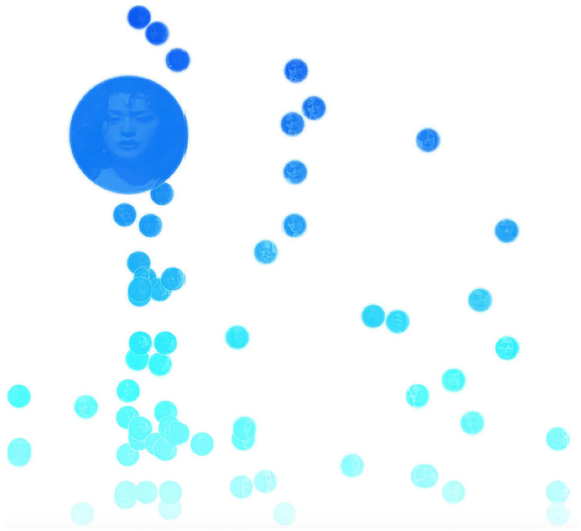


Fig.9. When analyzing Titanic, a user can easily find the scene of Rose's attempting to suicide by filtering and examining the most dense cluster of sad expressions.

Besides emotional information, additional character information may be valuable to users. With the results of the face detection and recognition data, we build a social network for all the characters in the movie. As shown in Fig.10, each person is a node in the network, where two characters are connected if they spoke in the same event. Then we initialize the social network based on the interaction between the characters and use the Pagerank [47] algorithm to calculate the centrality of each node in the network. The algorithm takes the importance of the character in the network into consideration and outputs measurements for the nodes connected to the important central node. Fig.10 shows the character network and the centrality of each character of the movie Titanic. The size of the portrait represents the centrality of each character while the thickness of the connection between two characters represents the intimacy.

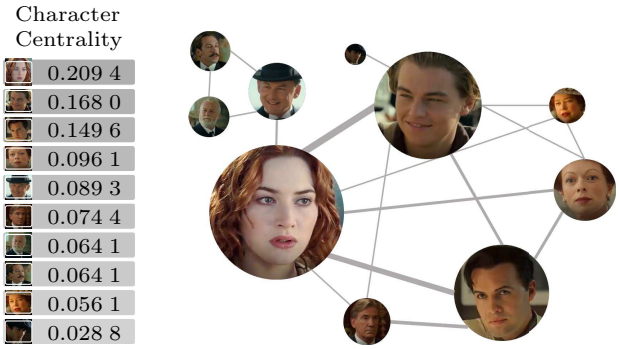


Fig.10. Character network, a social network constructed for the characters of Titanic based on the character centrality.

After the calculation of centrality, we put these characters on the map according to their appearance time in each event. And the size of each character's icon on the map represents the importance of the character. Moreover, we design characters' line on the map for exploring the participation of each character in the story. The position of characters in different events is determined by their occurrence time.

6 Interactive System for Video Analysis

As Kosara and Mackinlay stated: "visualization techniques address the exploration and analysis of data more than presenting data" [48], we aim at proposing an interactive system to explore video content including emotion in a natural and efficient way. Natural sketch interaction and multi-view browsing methods are integrated into our system for video analysis.

6.1 Multi-View for Emotion Analysis

As shown in Fig.1, there are four views in our system. Fig.1(a) is the main view that shows the video content using a map metaphor. The view in Fig.1(b) is a video player. The view in Fig.1(c) is for characters' emotion analysis based on the facial expression recognition results. Users can select the character on the map, and then the emotion of the character in the whole movie is displayed. The emotion of characters can be compared in Fig.1(c) for analysis. The view in Fig.1(d) at the bottom is the emotion curve including valence and arousal values that change over time. The curves are divided by events and show the emotional value and intensity over time according to valence and arousal values. The curve view in Fig.1(d) corresponds to the map view in Fig.1(a). We also mark out the scenes of several periods. With the view in Fig.1(d), users are free to explore video content from a spatial or temporal perspective in our system.

The views of the system are interrelated and interconnected to support multiple scales and views for video analysis (VT6). The operation on the curve in Fig.1(d) will be reflected on the map in Fig.1(a). Users can select an event on the curve which will be highlighted on the map. When users play the video in Fig.1(b) or slide over the curve in Fig.1(d), the hexagon of the corresponding time (second) on the map will be highlighted to indicate the current playback progress and location. Similarly, clicking on the hexagon on the map in Fig.1(a) or the curve in Fig.1(d) will play the video in Fig.1(b) of the corresponding event or time (VT4).

6.2 Sketch Interaction

We integrate sketch interaction for a more natural interaction in our system. We have designed some sketch gestures for different operations that allow users to browse, navigate and query movie content. For example, users can see the events that two characters participated together by drawing a line to connect them. Through natural sketch gestures, users can interact with the stories in a movie more easily and perform more effective visual analysis.

7 Evaluation

7.1 Experts Interviews

We invited three professional practitioners of video creation and video editing (P1, P2, P3). P1 works as

a video editor for post production editing of TV programs. P2 is a director that has created many video ads and movies. P3 is a developer in film processing who has seven years experience in video effects production. All practitioners have at least four years experience in video editing and film production. During the interviews, we introduced EmotionMap and they all used our system for free exploration to get a detailed understanding of all functions embedded in EmotionMap.

All participants agreed that the method of using a metaphor map to analyze video content is very novel. They also felt that it is natural to use sketch for the interaction on the map.

Two participants commented particularly on the usefulness of our work in video understanding and film production. P1: "I think spatial representations are good because it is easier to see relations". P2: "The map-based video expression method provides an overview for video directors. It helps directors consider how to combine the plot content of the story besides chronological order, such as the importance of characters and emotions."

As for the role of emotion in video analysis, they all think that emotion plays an important role. "Whether it is for video creators or video viewers, emotion makes a difference in analyzing video content", P3 stressed. P1 mentioned "People are more likely to focus on the blocks that are darker in color or larger on the map. EmotionMap represents the valence and arousal data in an efficient way for emotion concentration". P2 noted "The color in EmotionMap that includes valence and arousal data is good-looking as well as hierarchical." As a film and television practitioner, P2 stressed "video with great content can mobilize the mood of audience, and according to emotional feedback we can also in turn evaluate and modify the content of the video we create". All of the participants agreed with the application value of emotional analysis. P1 emphasized the important role of emotion analysis in advertising and movie trailer production. P3 hoped to analyze potential user emotions with visualization technology in the future.

7.2 User Study

In consideration that EmotionMap is a novel representation of video content and emotion, we created a user study that evaluates the effectiveness of the system and the user experience.

We listed all functions of the system, from which we selected 11 tasks as shown in Table 2. Fourteen volunteers from different study backgrounds were involved in our system experience task. First we spent about 5–10 minutes introducing EmotionMap and the operations of our system. Then we gave 11 tasks (T1-T11) to them in order, thereby they needed to complete the operations according to the task description. After they completed each task, they needed to answer two questions (TQ1-TQ2) in Table 3 according to the content of the task and usage experience. These tasks allow participants to experience all the functions of our system. Finally, they needed to complete a simple questionnaire that contains three questions (Q1-Q3) in Table 4 about subjective feedback after completing all the tasks. The questionnaire requires them to evaluate the whole system and provide their feedback. Each question has five options, including two positive options, one neutral option, and two negative options. We organized the results from all the participants and visualized the evaluation results in Fig.11.

Table 2. Task for Usage Experience

No.	Description
T1	Please play the video clip of the third event called Met
T2	Please check out the events which Carl was involved in
T3	Please check out the Carl's emotion in the Sink incident
T4	Please check out the emotion of Rose and Jack and observe the relation
T5	Please find out the events that Ruth and Jack were involved in together
T6	Please open the valence map and see the emotion distribution of events
T7	Please open the arousal map and see the emotion intensity of events
T8	Please have a look at the emotion of Rose in the 10th event
T9	Please check the emotion line view and find which event has the highest arousal value
T10	Please check the line view to find which event is the most negative
T11	Please check the line view to count the number of scenes and find the relation between different kinds of emotion

Table 3. Questions for Each Task

No.	Description
TQ1	Do you think it is helpful when you find the answer through our system in this task
TQ2	How do you feel when you use our system to finish this task

Table 4. Subjective Questions in the Questionnaire

No.	Description
Q1	Would you like to use our system to explore movies that you are interested in?
Q2	Do you think it is effective to use our system to explore emotion in movie?
Q3	Do you think it helps to understand movie content such as characters and events with our system?

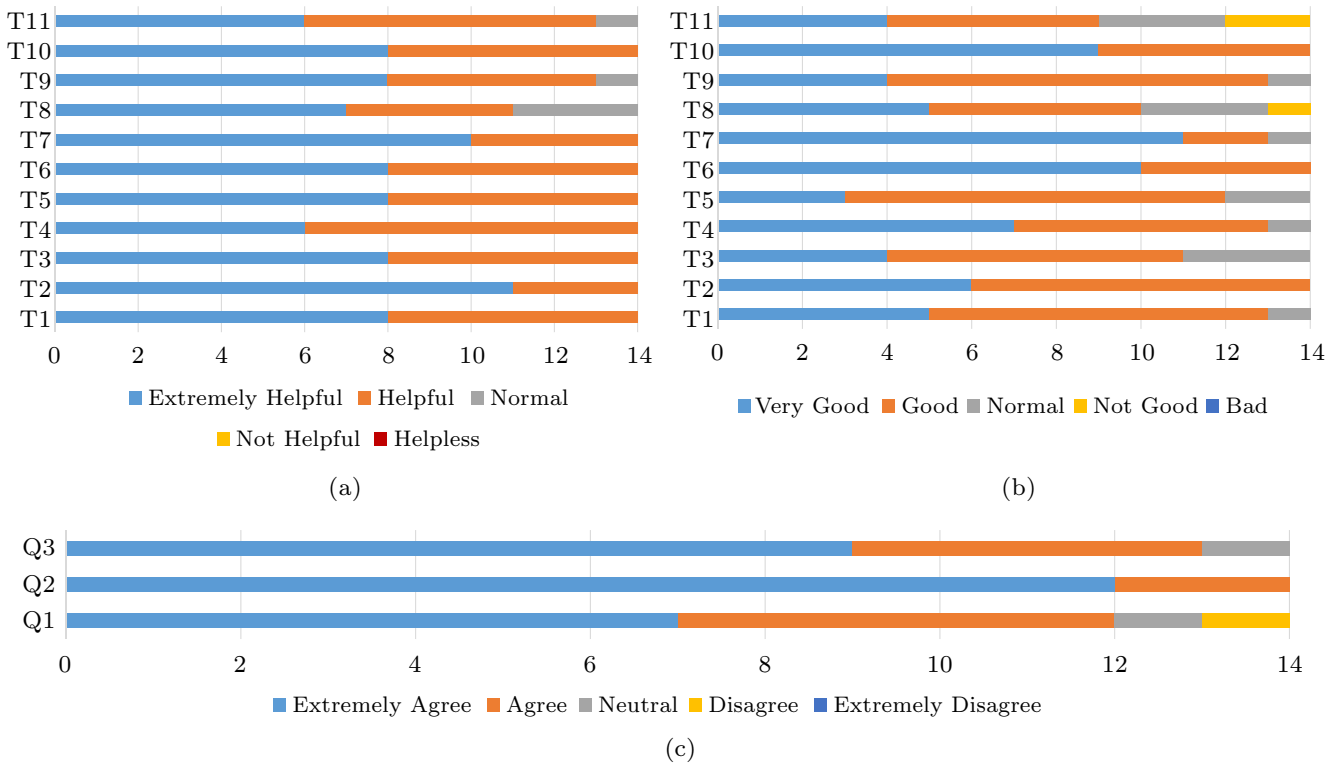


Fig.11. Result of the post-task questionnaire. (a) Help to find answer? (b) How do you feel over the process? (c) Post questionnaire.

The results of each participant's experience during the task are shown in Fig.11(b). Most of the participants think our system is user-friendly and convenient for video content analysis. We conclude the results of the questionnaire.

• EmotionMap has gained all participants' approval ("helpful" or "extremely helpful" to find the answer) for eight out of 11 tasks. The three remaining tasks that received partial approval are tasks T8, T9, and T11.

• According to the results of the participants' experience in completing the tasks, for every task, at least 64% (nine of 14 at least) of participants think EmotionMap is user-friendly.

• About 85% (12 of 14) of participants would like to use our system to explore video content in their life.

• All the participants (100%) of this task agree that EmotionMap is effective for exploring emotional video content.

• More than 90% (13 of 14) of participants think EmotionMap is helpful to analyze video content such as events, characters, and the relationship between characters.

Overall, the above results show that EmotionMap is an efficient, intuitive and user-friendly system for users to explore video content including emotions.

8 Conclusions

In this paper, we proposed EmotionMap, an efficient system for video content analysis. In order to collect the emotional assessments of audiences, we developed EmotionDisc, an efficient tool for video emotion collection. We built a system that allows users to query, navigate and explore the content of videos. EmotionMap is novel and it provides an efficient way for video content exploration, understanding, and searching. Combined with a novel use of a map, it also provides an advantage in exploring aspects of video content. We visualized several movies through our system and evaluated the validity and readability of metaphor map for video analysis.

We collected the feedback from all participants and their suggestions for our system. Some participants felt that it would take time to become familiar with our system. Some people thought the interaction with emotion on a map was not sufficient and additionally we need to try other types of videos. Future improvements to the system will focus more on user-centric analytics. In our future work, we will take the users' psychological signals into account for emotion visualization and analysis, such as pulse and EGG. We will collect more kinds of videos and generate additional maps for comparison. We plan to incorporate multi-modal information con-

tained in videos into automatic algorithms to assist in the analysis of video emotional content. Video visualization is a burgeoning field of study and we are developing new ways of interaction as well as emotion analysis for further exploration of video content.

References

[1] Bartsch A. Emotional gratification in entertainment experience. Why viewers of movies and television series find it rewarding to experience emotions. *Media Psychology*, 2012, 15(3): 267-302.

[2] Zhang S, Tian Q, Huang Q *et al*. Utilizing affective analysis for efficient movie browsing. In *Proc. the 16th IEEE International Conference on Image Processing*, November 2009, pp.1853-1856.

[3] Borgo R, Chen M, Daubney B *et al*. State of the art report on video-based graphics and video visualization. *Computer Graphics Forum*, 2012, 31(8): 2450-2477.

[4] Meghdadi A H, Irani P. Interactive exploration of surveillance video through action shot summarization and trajectory visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(12): 2119-2128.

[5] Stein M, Janetzko H, Lamprecht A *et al*. Bring it to the pitch: Combining video and movement data to enhance team sport analysis. *IEEE Transactions on Visualization and Computer Graphics*, 2017, 24(1): 13-22.

[6] Tanahashi Y, Ma K L. Design considerations for optimizing storyline visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2679-2688.

[7] Ekman P. An argument for basic emotions. *Cognition and Emotion*, 1992, 6(3/4): 169-200.

[8] Russell J A. A circumplex model of affect. *Journal of Personality and Social Psychology*, 1980, 39(6): 1161-1178.

[9] Liang H, Liang R, Sun G. Looking into saliency model via space-time visualization. *IEEE Transactions on Multimedia*, 2016, 18(11): 2271-2281.

[10] Zhang X, Dekel T, Xue T *et al*. MoSculp: Interactive visualization of shape and time. In *Proc. the 31st Annual ACM Symposium on User Interface Software and Technology*, October 2018, pp.275-285.

[11] Bach B, Shi C, Heulot N *et al*. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 2015, 22(1): 559-568.

[12] Parry M L, Legg P A, Chung D H S *et al*. Hierarchical event selection for video storyboards with a case study on snooker video visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 1747-1756.

[13] Liu S, Wu Y, Wei E *et al*. StoryFlow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(12): 2436-2445.

[14] Qiang L, Bingjie C, Haibo Z. Storytelling by the StoryCake visualization. *The Visual Computer*, 2017, 33(10): 1241-1252.

[15] Tong C, Roberts R, Borgo R *et al*. Storytelling and visualization: An extended survey. *Information*, 2018, 9(3): Article No. 65.

[16] Kim N W, Bach B, Im H *et al*. Visualizing nonlinear narratives with story curves. *IEEE Transactions on Visualization and Computer Graphics*, 2017, 24(1): 595-604.

[17] Kurzhals K, John M, Heimerl F *et al*. Visual movie analytics. *IEEE Transactions on Multimedia*, 2016, 18(11): 2149-2160.

[18] Pan Y, Niu Z, Wu J *et al*. InSocialNet: Interactive visual analytics for role-event videos. *Computational Visual Media*, 2019, 5(4): 375-390.

[19] Hanjalic A, Xu L Q. Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 2005, 7(1): 143-154.

[20] Wang S, Ji Q. Video affective content analysis: A survey of state-of-the-art methods. *IEEE Transactions on Affective Computing*, 2015, 6(4): 410-430.

[21] Jung H, Lee S, Yim J *et al*. Joint fine-tuning in deep neural networks for facial expression recognition. In *Proc. the IEEE International Conference on Computer Vision*, December 2015, pp.2983-2991.

[22] Zhao S, Yao H, Jiang X *et al*. Predicting discrete probability distribution of image emotions. In *Proc. the 2015 IEEE International Conference on Image Processing*, Sept. 2015, pp.2459-2463.

[23] Poria S, Cambria E, Bajpai R *et al*. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 2017, 37: 98-125.

[24] Zhalehpour S, Akhtar Z, Erdem C E. Multimodal emotion recognition based on peak frame selection from video. *Signal, Image and Video Processing*, 2016, 10(5): 827-834.

[25] Zhao S, Yao H, Sun X *et al*. Flexible presentation of videos based on affective content analysis. In *Proc. the 19th International Conference on Multimedia Modeling*, January 2013, pp.368-379.

[26] Lan Y, Wei S, Liu R *et al*. Creating video summarization from emotion perspective. In *Proc. the 13th International Conference on Signal Processing*, November 2016, pp.1112-1117.

[27] Liu Z, Xu A, Guo Y *et al*. Seemo: A computational approach to see emotions. In *Proc. the 2018 CHI Conference on Human Factors in Computing Systems*, April 2018, Article No. 364.

[28] Huang Y, Tang Y, Wang Y. Emotion map: A location-based mobile social system for improving emotion awareness and regulation. In *Proc. the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, March 2015, pp.130-142.

[29] Cao N, Lin Y R, Gotz D. UnTangle map: Visual analysis of probabilistic multi-label data. *IEEE Transactions on Visualization and Computer Graphics*, 2015, 22(2): 1149-1163.

[30] Chen S, Chen S, Lin L *et al*. E-map: A visual analytics approach for exploring significant event evolutions in social media. In *Proc. the 2017 IEEE Conference on Visual Analytics Science and Technology*, October 2017, pp.36-47.

[31] Chen S, Chen S, Wang Z *et al*. D-map: Visual analysis of ego-centric information diffusion patterns in social media. In *Proc. the 2016 IEEE Conference on Visual Analytics Science and Technology*, October 2016, pp.41-50.

[32] Watson M C. Time maps: A tool for visualizing many discrete events across multiple timescales. In *Proc. the 2015 IEEE International Conference on Big Data*, October 2015, pp.793-800.

[33] Xin R, Ai T, Ai B. Metaphor representation and analysis of non-spatial data in map-like visualizations. *ISPRS International Journal of Geo-Information*, 2018, 7(6): Article No. 225.

[34] Yang M, Biuk-Aghai R P. Enhanced hexagon-tiling algorithm for map-like information visualisation. In *Proc. the 8th International Symposium on Visual Information Communication and Interaction*, August 2015, pp.137-142.

[35] Auber D, Huet C, Lambert A *et al*. GosperMap: Using a gosper curve for laying out hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(11): 1820-1832.

[36] Gansner E R, Hu Y, Kobourov S. GMap: Visualizing graphs and clusters as maps. In *Proc. the 2010 IEEE Pacific Visualization Symposium*, March 2010, pp.201-208.

[37] Pang P C I, Biuk-Aghai R P, Yang M *et al*. Creating realistic map-like visualisations: Results from user studies. *Journal of Visual Languages & Computing*, 2017, 43: 60-70.

[38] Sen S, Swoap A B, Li Q *et al*. Cartograph: Unlocking spatial visualization through semantic enhancement. In *Proc. the 22nd International Conference on Intelligent User Interfaces*, March 2017, pp.179-190.

[39] Mashima D, Kobourov S, Hu Y. Visualizing dynamic data with maps. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 18(9): 1424-1437.

[40] Ma C X, Liu Y J, Zhao G *et al*. Visualizing and analyzing video content with interactive scalable maps. *IEEE Transactions on Multimedia*, 2016, 18(11): 2171-2183.

[41] Gunes H, Schuller B. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 2013, 31(2): 120-136.

[42] Arriaga O, Valdenegro-Toro M, Ploger P. Real-time convolutional neural networks for emotion and gender classification. arXiv:1710.07557, 2017. https://arxiv.org/abs/1710.07557, March 2020.

[43] Huang G B, Learned-Miller E. Labeled faces in the wild: Updates and new reporting procedures. Technical Report, Univ. Massachusetts, 2014. http://vis-www.cs.umass.edu/lfw/lfw_update.pdf, March 2020.

[44] van Kreveld M, Speckmann B. On rectangular cartograms. In *Proc. the 12th Annual European Symposium on Algorithms*, September 2004, pp.724-735.

[45] Plutchik R. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 2001, 89(4): 344-350.

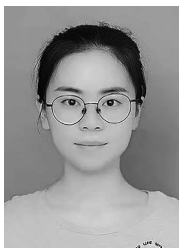
[46] Schloss K B, Gramazio C C, Silverman A T *et al*. Mapping color to meaning in colormap data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2018, 25(1): 810-819.

[47] Page L, Brin S, Motwani R *et al*. The pagerank citation ranking: Bringing order to the web. Technical Report, Stanford InfoLab, 1999. http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf, March 2020.

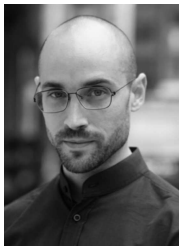
[48] Kosara R, Mackinlay J. Storytelling: The next step for visualization. *IEEE Computer*, 2013, 46(5): 44-50.

Cui-Xia Ma received her B.S. and M.S. degrees in computer application technology from Shandong University, Jinan, in 1997 and 2000, respectively, and her Ph.D. degree in computer application technology from the Institute of Software, Chinese Academy of Sciences, Beijing, in 2003. She was a research associate in the Department of Computer Science, Naval Postgraduate School, Monterey, from 2005 to 2006. She is currently a professor with Institute of Software, Chinese Academy of Sciences, Beijing. Her research interests include human computer interaction and multimedia computing.

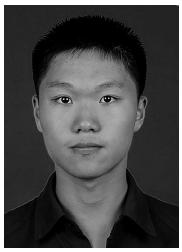
Jian-Cheng Song is a Master student in software engineering at the Institute of Software, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, Beijing. His current research interests include video visual analysis and emotion analysis. He received his B.S. degree in computer science from University of Science and Technology Beijing, Beijing, in 2019.



Qian Zhu is a Master student in computer science at the Institute of Software, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, Beijing. Her research interests focus on video visual analysis and emotion analysis. She received her B.S. degree in digital media from Shandong University, Jinan, in 2017.



Kevin Maher currently is an M.A. candidate in visual communication at Tsinghua University, Beijing. He received his B.A. degree in literature from the University of Louisiana, Lafayette, in 2010. He is currently interning at Institute of Software, Chinese Academy of Sciences, Beijing. His research interests include human computer interaction and data visualization.



Ze-Yuan Huang is a Ph.D. candidate in software engineering at the Institute of Software, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, Beijing. His research interests focus on emotion analysis and human-computer interface.

Hong-An Wang received his Ph.D. degree in computer application technology from the Institute of Software, Chinese Academy of Sciences, Beijing, in 1999. He is a professor with the Institute of Software, Chinese Academy of Sciences. He is currently the director of Beijing Key Laboratory of Human-Computer Interaction. His research interests include human-computer interaction, real-time intelligence, and real-time active database.