A Proposal of Video Evaluation Method Using Facial Expression for Video Recommendation System

Masashi Okubo[✉] and Shun Tamura

Doshisha University, 1-3 Tatara-Miyakodani,
Kyotanabe, Kyoto 610-0321, Japan
mokubo@mail.doshisha.ac.jp, shun.tamura@outlook.com

Abstract. Recently, video sharing services and video on demand services like YouTube, Netflix and Amazon Prime Video have come to prominence world-wide. These services try to get the users' evaluation of videos for the better recommendation. For example, some of them prepare the thumbs-up/down buttons and star buttons in their interface to get users' explicit evaluation. However, the effect of this method is doubts. The reason is these method needs user's conscious operation. Therefore, we need efficient method for getting users' implicit evaluation of videos. In this study, we focus on facial expressions especially smile of users watching videos. In this paper, we have developed the system that can get user's expression of users watching videos by some APIs, and investigated the relation between evaluation of videos by scoring and facial expression by using the proposed system. As a result of the experiment, positive correlations between the ranking by the user's scoring and the ranking based on the amount of smile of users watching videos are shown in 22 of the 28 participants. In addition, strong positive correlations were found among 13 of them. This result suggests possibility of using the implicit evaluation like smiling as well as the explicit evaluation like a scoring. So, it is suggested that recommendation method based on facial expression is effective.

Keywords: Facial expression · Recommendation method · Video on demand

1 Introduction

In recent years, video-on-demand websites like Netflix, Amazon Prime Video, and even user-submitted video sites like YouTube have taken the world by storm. These websites utilize a keyword search engine to better help users find videos they are looking for without having to manually navigate the massive video libraries these sites offer. These search engines pull information from a video's title or synopsis that matches the input keyword. However, in websites where the synopsis can be freely edited by the up loader, there are many times where the information provided does not match the video's content. This can understandably prevent users from finding the video they are looking for. It is also common for searches to produce a vast number of pages of results, so finding a video that piques the interest of the user may be a bit daunting. In order to solve this dilemma, many websites offer a video recommendation

system which can be fine-tuned to suggest videos the user may be interested in. Netflix, one of the world's biggest video-on-demand retailers, shows us just how valuable this recommendation system is since nearly 80% of the videos watched are the result of recommendations [1].

The standard method in which videos are recommended is through collaborative filtering. The degree of similarity is calculated between users who vote on a video, and videos are recommended to other users that fall into that degree of similarity whenever a user with similar tastes approves of a video. Recently as well, recommendation systems that utilize machine learning have come into the spotlight. YouTube, for example, has created a neural network for its recommendation system which is built up of tens of billions of points of training data stemming from nearly a billion parameters [2]. Likewise, the act of user input in the form of voting on videos is an integral part of this method as well. A user's video evaluation index is expanded whenever the user performs explicit acts like pressing the "thumbs up" button on the webpage or subscribing to a content creator. Acts the user may not even know about add information to this index like watching a video to the end or watching the same video multiple times.

However, explicit actions like actively rating a video may be troublesome for the user, and so there is reason to believe that not many users do this. Further, even if a user were to watch a video to the end, it still does not necessarily mean they think highly of it, and so may not reflect exactly how the user feels. It is believed to be for these reasons that there is a need for an implicit rating system which can accurately reflect how the user feels about a video.

2 Video Recommendation System from Facial Expressions

2.1 Research Regarding Implicit Video Rating

In reference to the dilemma stated earlier, M. Suganuma and associates devised a method which utilized retinal information and brain waves [3]. During their experiments, it was suggested that reaction patterns from the participants in regards to video stimulation could be sorted into types. Conversely, D. Natsume and associates proposed an alternative video recommendation method based on tracking eye movement during a video. There, a user's video recommendation would be built up upon other similar data points taken from other users [4]. In their experiment, 20 participants were made to watch a multitude of videos while having their eye movement tracked. At the same time, participants were required to rate each video on a 1–4 scale which would create a degree of interest. A recommendation rank would then be defined through the recommendation target user and other user's similar points of interest. After calculating the rank correlation created from actual participant input degrees of interest, it turned out that it had a correlation average of 0.61.

Regardless of the method, it is indeed true that special instruments would be needed to analyze brain waves and eye movement. Furthermore, equipping users with such instruments would be no easy feat. Therefore, there is much reason to believe that a more natural and easy to use system is necessary.

2.2   Video Recommendation System from Facial Expressions

The main objective of this research is to obtain an index which can appropriately reflect a user's implicit evaluation of a video. Moreover, this method is meant to be utilized for standard video scenes, and ideally the use of special equipment should not be necessary. Essentially, this research deals with facial expressions, as facial expressions play an essential role in a person's emotional expression [5]. However, with recent advancements in facial expression recognition technology, facial expressions taken from a portrait photo can be recognized easily even if taken from a standard camera. Therefore, it is suggested that by filming a user's face while they watch a video, it is possible for a method to read their implicit evaluation of the video based solely on their facial expressions. Additionally, this research aims to develop a video recommendation system which utilizes a user's facial features to find videos they may find interesting. An outline of the proposed system can be found in Fig. 1.
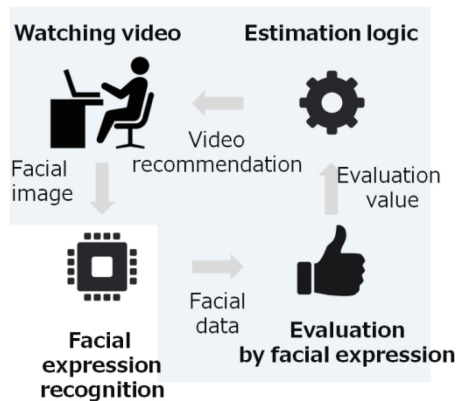

Fig. 1. Outline of proposed system

3   Relationship Between Facial Expressions During a Video and Its Evaluation

3.1   Outline

An experiment was carried out with 8 participants at Doshisha University with the aim of verifying the relationship between facial features made by a viewer during a video and video evaluation.

3.2   Test System

Since the goal of this research is to track the facial expressions of a viewer during a video while at the same time not putting too much of a physical burden on them, I developed a system which would record the facial expressions analyzed through a web

camera loaded with a facial expression recognition API and which would then list them chronologically. The proposed system outline can be found in Fig. 2.
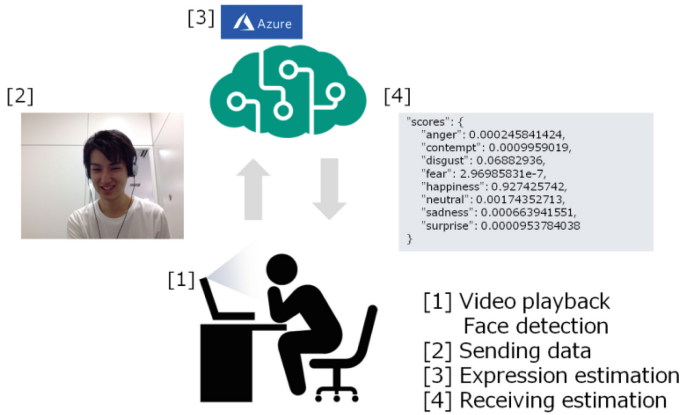

"scores": {
    "anger": 0.000245841424,
    "contempt": 0.0009959019,
    "disgust": 0.06882936,
    "fear": 2.96985831e-7,
    "happiness": 0.927425742,
    "neutral": 0.001174352713,
    "sadness": 0.000663941551,
    "surprise": 0.0000953784038
}

[1] Video playback
    Face detection
[2] Sending data
[3] Expression estimation
[4] Receiving estimation

Fig. 2. Outline of test system

There are 4 videos within this test system. The web camera automatically begins photographing the user's face as soon as they press the video's play button. The web camera continues to take a photo of the user's face every three seconds. After the video is finished, the photos are sent to an outside web API. The facial expression recognition results are then sent from the web API and are stored inside the system.

Emotion API, which is a web API part of the Cognitive Services API pack developed though Microsoft's machine learning, was used in recognizing the facial expressions during this experiment [6]. Emotion API analyzes the received images of a person's face, and then performs facial expression recognition and emotion estimations. The feelings it can detect are 7 in total. These are anger, contempt, disgust, fear, happiness, sadness, and surprise. If it cannot ascertain a feeling from the image, then it analyzes it as a neutral feeling. The API's emotion estimation results appear as a confidence, where each feeling can receive a max confidence score of 1.

3.3 Experiment Method

A participant is directed into a room and is instructed to watch 4 different videos, as required in the test system. The content of the videos is explained in Table 1. After the first video is finished, the participant is then asked to answer a survey where they will state how they felt after it finished and to what degree of interest they are currently at. Survey questions can be found in Table 2. All the participants watched the videos in the order of A, B, C, and then D. After all the videos were finished, the Emotion API was sent the photographed facial expressions where they were then assigned an estimated emotion value. It was hypothesized that there are not many users who actively watch videos which cause them to feel feelings like anger, fear, and contempt, so only feelings like neutral, happiness, sadness, and surprise were given an estimated emotion value.

Table 1. Video contents

Video#	Contents
A	Natural scenery
B	Flash mob
C	Painful separation
D	Magic show

Table 2. Survey after viewing video

Question#	Question	Answer
Q1	Do you know the video?	Yes/No
Q2	What kind of feelings sprang out?	Happiness/Surprise/Sadness/and so on
Q3	Are you interested in the video?	Yes/No
Q4	Why do you think so?	Free answer
Q5	Do you want the similar video?	Yes/No

3.4   Experiment Results

Figure 3 features an example result taken from participant 1. It displays the time scale of the emotion "happiness" which was analyzed during video D viewing. From this data we can see that happiness hits the confidence threshold multiple times. The participant stated they felt a mix of happiness and shock, and that they ranked the video a 5 on the degree of interest scale.
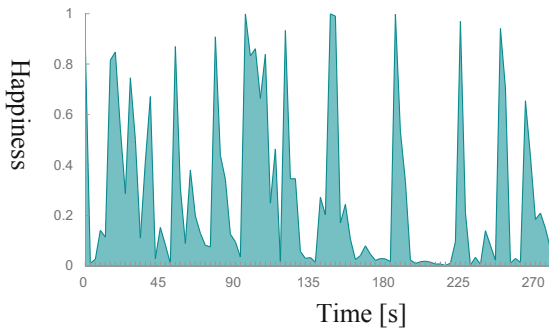

Fig. 3. Transition of happiness scores for participant 1 (video D)

Throughout the experiment, participant 1 changed facial expressions many times according to which videos they found high on their degree of interest scale. The other 4 participants also showed a similar reaction to a couple videos as well. However, it was not possible to attain a video evaluation reading based on the changing facial expressions of 3 different participants. Further, since "happiness" was the emotion which had the most changes in terms of confidence across all participants, there is reason to believe that it easy for the algorithm to recognize.

4 Relationship Between a Viewer's Smile and Video Evaluation

4.1 Outline

From the experiment mentioned in Sect. 3, the smiles made during the videos allowed the algorithm to better recognize them, so one could then assume that there is indeed a relation between video evaluation and facial features. Doshisha University conducted another experiment for even further testing. This time the target facial expression was restricted to smiles only, and 28 students were observed to see if there was a relation between smiling during a comedy video and video evaluation.

4.2 Test System

This test works in the same way as the experiment in Sect. 3. Participants' facial expressions were tracked during a video while at the same time not putting too much of a physical burden on them. The system I developed would record the facial expressions through a web camera and send the images to the facial expression recognition component. The difference between this test and the one mentioned in Sect. 3 is the facial expression recognition component. On the assumption that real-time analysis would be integral to an actual recommendation system, I incorporated an SDK into the system which will help alleviate possible network lag issues due to the web-based API. More on the SDK will be explained below. The system can be accessed through a web interface, where the 9 comedy videos to be used will be hosted. On the starting screen, buttons will be displayed in random order. Each button linked to a video and its survey. Once the play button is pressed on the video page, the web camera will automatically begin taking photos of the face. Photos of the face during the video will be sent to the facial expression recognition component, where facial expression recognition protocols will be carried out until the video ends. After the video is over, the facial expression recognition results will then be stored on the PC. An overview of this system is shown in Fig. 4.
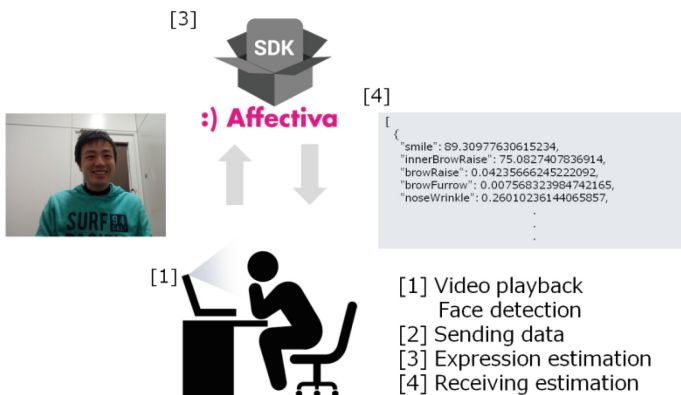

Fig. 4. Overview of the test system

Affectiva's Affectiva Emotion SDK is being used as the built-in type facial expression recognition SDK. When images are input into the Affectiva Emotion SDK, it starts by extracting the facial boundaries from each image. Then it extracts and confirms features that make up the face, like eyes, eyebrows, and lips. By utilizing Affectiva's own model, it can determine if parts on the face like eyes are open or if one's eyebrows are raised. After that, those conditions are compared with the Facial Action Coding System's (FACS) Action Units (AU), where then facial expression recognition and an estimation of emotion begins [7].

FACS is an analysis tool part of Ekman and their associates' proposed facial expression theory [8]. FACS takes anatomical knowledge about the face and applies it to classify relationships between the movement of mimetic muscles and how those muscles express emotion. AU is a term used to classify one codified way of moving for a mimetic muscle, and facial expressions are built upon numerous AU components working together. For example, combining AU6's "raises cheeks" with AU12's "raises corners of mouth" creates a smile, which is generally understood as signifying happiness or joy. The Affectiva Emotion SDK evaluates the submitted facial photos combined with AU data, allowing facial expression recognition and emotion estimations to take place. The results are then split into two separate JSON files: one for expressions and one for emotions. Expressions contains information on physical indicators like smiles or frowns, while emotions contains data on expressed emotion like joy or anger. There, both expressions and emotions are scaled 0 to 100 in terms of confidence.

4.3   Experiment Method

The process is carried out according to the above-mentioned test system. A participant is ushered into a room and is shown 9 4-min comedy videos. After watching a video, the participant must then evaluate what he/she saw by grading it 1–100, and if it is something he found to be particularly interesting, then he is free to give his thoughts on it. The videos are shown to the participant randomly, but in order to establish a base for evaluation across all participants, each participant begins by watching the same video (video A).

4.4   Evaluation Method

First, confirmed smile confidence results are extracted from the analyzed facial expression data gained through the test system. Figure 5 features an example result taken from participant 3. It displays the time scale of the expression "smile" which was analyzed during video A and video D viewing.

Next, the average smile confidence scores are defined as the smile volume, which is calculated into each video each participant watched. These volumes, big or small, are compared, and the videos the participants watched are placed into a ranked list. Furthermore, results from the survey data taken from each participant is applied and a ranked list is also created from this data. Spearman's rank correlation coefficient was used in order to confirm a relationship between the estimated evaluation based on a participant's facial features during a video and the explicitly evaluated data they
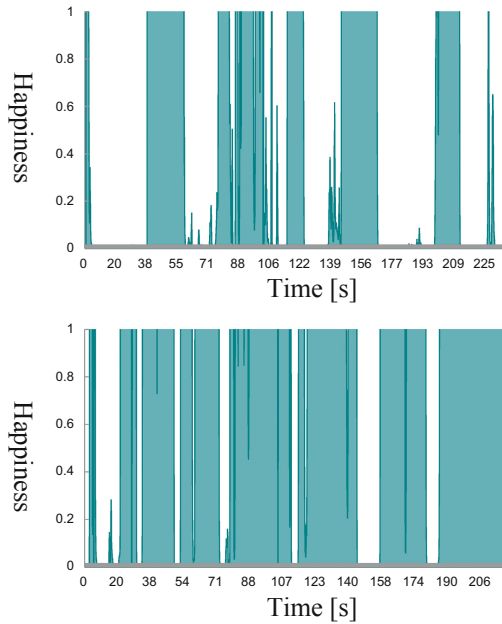
Fig. 5. Participant 3's time scale on smile confidence scores during video A (upper) and video D (lower) viewing

submitted when grading a video [9]. Spearman's rank correlation coefficient takes two variables converted through rank and finds their correlation by defining them through the following Eq. (1). A rank correlation coefficient of +1 or −1 occurs when each of the variables is a perfect monotone function of the other.

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{1}$$

n: number of videos, d_i: difference between the two ranks.

4.5 Evaluation Results

Table 3 features the calculated results after running both evaluation ranks (participants' graded video scores and the volume of smiles per video) through Spearman's rank correlation coefficient.

A strong rank correlation coefficient is considered to be $0.7 \leq |r_s|$, while a weak correlation coefficient would be $0.3 \leq |r_s| < 0.7$. Out of the 28 participants in this study, 13 expressed what is considered a strong positive rank correlation coefficient. Furthermore, 9 participants expressed a weak positive rank correlation coefficient. However, 1 participant expressed a strong negative rank correlation coefficient.

Table 3. Spearman's rank correlation coefficient of each participant

No.	Correlation	No.	Correlation	No.	Correlation
1	0.535	11	0.851	21	−0.824
2	−0.067	12	0.863	22	0.940
3	0.834	13	0.717	23	0.872
4	0.373	14	0.661	24	−0.017
5	0.895	15	0.638	25	0.274
6	0.370	16	0.262	26	0.740
7	0.770	17	0.559	27	0.655
8	0.899	18	0.832	28	0.891
9	0.538	19	0.586		
10	0.929	20	0.119		

4.6   Observations

In the case of participant 21, there is a possibility that the facial recognition algorithm misunderstood his/her smile and incorrectly identified it as a crying face. There is also a possibility that the explicit evaluation the participant gave did not match the implicit (facial expressions) evaluation while participating in the experiment. Therefore, there is reason to believe that either the written evaluation or the user's own smile fails to accurately reflect the user's true evaluation of the video.

Conversely, since the other 22 participants showed a positive correlation in regards to their smiles during the videos, there is reason to believe that implicit smile-based video evaluation is indeed possible. By digitizing the evaluation data, it seems further possible that it can be applied to currently available recommendation systems like collaborative filtering. Essentially, this means that recommendation feedback loops, like users being able to discover recommended videos from other users with the same interests, all without having to perform explicit evaluation tasks like pressing a thumbs up button, will be possible.

5   Video Recommendation via a Viewer's Smile

5.1   Outline

As the experiment in Sect. 4 has shown, smiling during a video and that video's evaluation seem to be related to each other. In this chapter, recommendations based on the data gained through Sect. 4's experiment will be used, and the possibility of the implementation of a recommendation system attained through smiling during a video will be explored.

5.2   Recommendation Method

The method in this video recommendation experiment is based off the ideas concerning collaborative filtering. Collaborative filtering is a method that seeks to find a degree of

similarity between users and then, based on that similarity, suggests recommended videos that a user has yet to see based on similar tastes in previously evaluated videos. This experiment will pre-process the obtained data and calculate a degree of similarity. A ranking of the recommended videos will then be determined after passing the results through the 3 processes of prediction value evaluation.

First, pre-processing of the previously obtained data in order to create a model is to be carried out. It is understood that users currently using such services watch videos that pertain to their interests. There, recommendations will be taken from the user's top 3 rated, previously watched, received recommendations. An outline of the pre-processing for participant 18 is explained in Fig. 6.

No.	A	B	C	D	E	F	G	H	I
18	35.046	74.030	58.349	70.632	44.583	46.659	46.227	2.486	30.498

No.	A	B	C	D	E	F	G	H	I
18	?	74.030	58.349	70.632	?	?	?	?	?

Fig. 6. Pre-processing for participant 18

Next, a calculation will be run from the 3 videos in order to find a degree of similarity between the former recipient and another participant. The method used to find the degree of similarity is the Pearson's correlation coefficient. The values are participants i and j, as well as the degree of similarity similarity(i, j). This is then put into the formula (2) and becomes $-1 \leq$ similarity(i, j) ≤ 1.

A sample of participant 18's results through this method is shown in Fig. 7. A degree of similarity was calculated from the extracted videos B, C, and D during pre-processing and the amount of smiling other participants did during said videos. For example, participant 1's degree of similarity was similarity(18,1) = 0.3666, and participant 14's degree of similarity was similarity(18,14) = 0.9999. By extracting the top 4 participants that shared a positive correlation, the prediction value evaluation for participant 18 resulted in the 4 participants being participant 14, 17, 23, and 27.

$$similarity(i,j) = \frac{\sum_{p \in P}\left(r_{ip} - \overline{r_i}\right)\left(r_{jp} - \overline{r_j}\right)}{\sqrt{\sum_{p \in P}\left(r_{ip} - \overline{r_i}\right)^2}\sqrt{\sum_{p \in P}\left(r_{jp} - \overline{r_j}\right)^2}} \tag{2}$$

i: Recommended participant, j: Participants excluding the recommended participant,

r_{ip}: Evaluation of video(p) by participant(i),
$\overline{r_i}$: Average evaluation of participant(i),
r_{jp}: Evaluation of video(p) by participant(j),
$\overline{r_j}$: Average evaluation of participant(j),
P: Top 3 videos evaluated by i, p: video.

Object participant

No.	B	C	D	correlation
18	74.030	58.349	70.632	-

& Top 4 participants

No.	B	C	D	correlation
1	96.914	96.740	96.645	0.3666
⋮	⋮			⋮
14	22.416	0.009	17.586	0.9999
⋮	⋮			⋮
17	69.886	9.379	61.997	0.9961
⋮	⋮			⋮
24	6.893	2.715	5.618	0.9955
⋮	⋮			⋮
27	53.332	13.175	48.506	0.9952
28	51.761	33.075	34.661	0.7226

Fig. 7. Degree of similarity calculated from participant 18

Next, an estimation of the number of smiles an unwatched video recommended through the prediction value evaluation based on the degree of similarity will get will be determined. This prediction formula is displayed in formula (3). The estimation takes the values from the top 4 participants based on their degree of similarity but excludes any participants who have a negative correlation coefficient. The recommendation rank of videos will be listed by rank from the highest predicted evaluation value.

$$predict(i,p) = \overline{r}_i + \frac{\sum_{j \in N} similarity(i,j)(r_{jp} - \overline{r}_j)}{\sum_{j \in N} |similarity(i,j)|} \tag{3}$$

i: Recommended participant, j: (i) and participants that have a high degree of similarity (Max 4),

p: Predicted videos,

$\overline{r}_i$: Evaluation average of participant(i),

$\overline{r}_j$: Evaluation average of participants(j),

r_{jp}: Participants'(j) evaluation of videos(p),

N: i and users with a high degree of similarity.

The process used for participant 18 is shown in Fig. 8. Since participant 18's top 4 similar users were participants 14, 17, 24, and 27, the number of smiles for participant 18 and the previously mentioned users for each video were used to find a degree of similarity. Then, based off this data, a prediction regarding the number of smiles for unseen videos was made. For example, the calculated estimation value for video A is 72.246 and the value for video D is 58.706.

No.	A	B	C	D	E	F	G	H	I
14	7.257	22.416	0.009	17.586	5.574	5.952	4.243	8.312	2.426
17	36.032	69.886	9.379	61.997	9.692	85.243	82.425	0.922	58.621
24	12.971	6.893	2.715	5.618	2.474	5.888	10.155	1.837	7.384
27	58.263	53.332	13.175	48.506	42.577	37.465	24.787	6.373	37.241

▼

No.	A	B	C	D	E	F	G	H	I
18	72.246	74.030	58.349	70.632	58.706	77.255	74.022	48.004	70.035

Fig. 8. Participant 18's prediction value evaluation

Extrapolating from the recommendation method explained above, we can conclude that recommendations through grading videos work in the same way even by replacing explicit participant grading with the volume of implicit smiles made during the video.

5.3 Evaluation Method

Spearman's rank correlation coefficient was used in order to evaluate the calculated recommendation rank taken from the previous paragraph's recommendation method [9]. Specifically, it seeks out the rank correlation coefficient between two variables. The first is the rank which is created from videos based on smile volume. The second is the evaluation rank based on smile volume the user receiving the recommendation made which was measured by the system in regards to the same videos. Further, it seeks out the rank correlation coefficient between the displayed rank of recommendations based on marked results and the evaluated rank based on actual markings.

Participant 18's example can be found in Fig. 9. Regarding pre-processing, the ranks for the 6 unwatched videos A, E, F, G, H, I were based on the measured smile volumes and the predicted smile volumes the system made. A strong positive correlation of 0.829 was found when seeking out Spearman's rank correlation coefficient based on the ranks.

5.4 Evaluation Results

Upon calculating the rank correlation coefficient of each participant based on each of the methods, we find that, in terms of recommendations based on smile volume, 5 participants showed a strong positive correlation, 3 participants showed a weak positive correlation, and 1 participant showed a strong negative correlation. However, for videos that were graded or marked, 2 participants showed a strong positive correlation, 13 participants showed a weak positive correlation, and 1 participant showed a weak negative correlation.
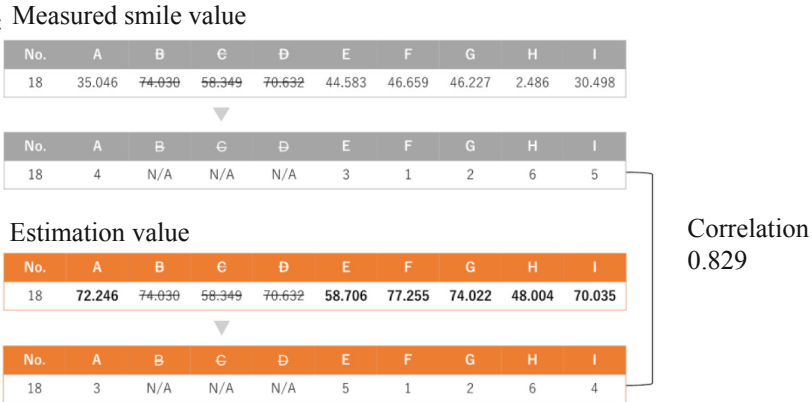
Measured smile value

No.	A	B	C	D	E	F	G	H	I
18	35.046	~~74.030~~	~~58.349~~	~~70.632~~	44.583	46.659	46.227	2.486	30.498

▼

No.	A	B	C	D	E	F	G	H	I
18	4	N/A	N/A	N/A	3	1	2	6	5

Estimation value

No.	A	B	C	D	E	F	G	H	I
18	72.246	~~74.030~~	~~58.349~~	~~70.632~~	58.706	77.255	74.022	48.004	70.035

▼

No.	A	B	C	D	E	F	G	H	I
18	3	N/A	N/A	N/A	5	1	2	6	4

Correlation
0.829

Fig. 9. Evaluation of participant 18's recommended video results

5.5   Observations

From the results, the data showed that participants had both rank correlation coefficients (the rank correlation coefficient between graded recommendation ranks and evaluated ranks and the rank correlation coefficient between recommendation ranks based on smiles and evaluated ranks) that showed strong positive correlations greater than 0.7. This data is shown in Table 4.

Table 4. Rank correlation coefficient of participants ranging 0.7+

No.	Correlation between ranking by evaluation and	
	Ranking based on score	Ranking based on smile
1	-0.293	1.000
3	N/A	0.714
5	-0.131	0.714
9	0.667	0.771
10	0.841	-0.029
16	0.736	0.143
18	0.493	0.829

From the table, we can see there were no participants who showed a strong positive correlation between recommendations created through both grading and smiles. Therefore, it is possible to assume that the recommendations based on the recommendation index varies from person to person.

6   Conclusion

To recap, the aim of this paper's research was to propose a method which could be used to obtain implicit video evaluation based on a user's facial features while they watched a video. Additionally, the relationship between facial features and video evaluation was explored by photographing the face of a user while they watch a video and developing a system which utilizes facial expression recognition APIs and SDKs. Development on the recommendation method was carried out in part due to the facial features data obtained through the experiment.

It was found, through the results from investigating the relationship between the facial features appearing on the viewer while watching a video and video evaluation, that there seemed to be a relationship between graded videos by users deemed by the system to have clear facial expressions and the volume at which facial features changed during the viewing of a video. Additionally, it was found that facial features changed often when viewing videos that made people smile. Therefore, if a smile is made during a video, and that smile is easily recognized by the system, then there is a possibility that a relationship between video evaluation and changes in facial features exists.

Next, concerning the results obtained after investigating the relationship between smiles during a video and video evaluation, it was found that of the 28 tested participants, 22 showed a positive correlation between how they evaluated the videos by grading and how they were evaluated by smile volume. Of those 22 participants, 13 showed a strong positive correlation. This showed that facial features could be used as an implicit index much in the same way grading works as an explicit index, which means that it is possible for videos to be evaluated by an index of smiles while the video is being watched.

Furthermore, the possibility of creating a video recommendation index through the volume of smiles expressed during a video was investigated, in which similar recommendations were based off the data set obtained through the prior experiment.

From the results of the experiment, it was found that no participants expressed a strong positive rank correlation between both recommendations made from an explicit graded index and recommendations made from an implicit smile volume index. This meant that it was possible for applicable video recommendation indexes to differ depending on the person.

However, though recommendation results based on smile volumes of specially selected participants seems to be effective, the rank correlation coefficient average was found to be lower than recommendation results based on grading. Throughout the experiment, a simple index for recommendations was made by taking the average of smile confidence scores during a video. However, there is a possibility that this index may not have been appropriate. There is still much to be investigated in terms of just how far a model created through smile confidence data will go when used in recommendations.

Additionally, the target of this report was restricted to smiles. However, there exist many videos which depict drama or horror which in turn can bring about a plethora of varying facial expressions. The evaluation index of these emotions is a worthwhile topic for further study.

References

1. Gomez-Uribe, C.A., Hunt, N.: The netflix recommender system: algorithms, business value, and innovation. ACM Trans. Manag. Inf. Syst. **6**(4), 1–19 (2016). 13
2. Covington, P., Adams, J., Sargin, E.: Deep neural networks for YouTube recommendations. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 191–198 (2016)
3. Suganuma, M., Kawamura, E., Kameyama, W.: Preliminary study on estimating degree of interest for audiovisual information using user's bio signals. In: The Institute of Image Information and Television Engineers Winter Annual Convention 2011, pp. 7–4 (2011)
4. Natsume, D., Iribe, Y., Katsurada, K., Nitta, T.: Proposal of video recommendation method using eye mark. In: Information Processing Society National Conference Proceedings, pp. 857–858 (2008)
5. Ekman, P., Friesen, W.V.: Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues. Malor Books, 202 p. (2003)
6. Microsoft Corp.: Emotion API–Emotion detection. https://azure.microsoft.com/services/cognitive-services/emotion. Accessed 26 Jan 2018
7. Senechal, T., Mcduff, D., Kaliouby, R.: Facial action unit detection using active learning and an efficient non-linear kernel approximation. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 10–18 (2015)
8. Ekman, P., Friesen, W.V.: Measuring facial movement. Environ. Psychol. Nonverbal Behav. **1**(1), 56–75 (1976)
9. Kendall, M.G.: Rank Correlation Methods. Hodder Arnold, 272 p. (1948)