

Personality, Culture, and System Factors - Impact on Affective Response to Multimedia

Sharath Chandra Guntuku, Michael James Scott,
Gheorghita Ghinea *Member, IEEE*, Weisi Lin *Fellow, IEEE*

Abstract—Whilst affective responses to various forms and genres of multimedia content have been well researched, precious few studies have investigated the combined impact that multimedia system parameters and human factors have on affect. Consequently, in this paper we explore the role that two primordial dimensions of human factors - personality and culture - in conjunction with system factors - frame rate, resolution, and bit rate - have on user affect and enjoyment of multimedia presentations. To this end, a two-site, cross-cultural study was undertaken, the results of which produced three predictive models. Personality and Culture traits were shown statistically to represent 5.6% of the variance in positive affect, 13.6% in negative affect and 9.3% in enjoyment. The correlation between affect and enjoyment, was significant. Predictive modeling incorporating human factors showed about 8%, 7% and 9% improvement in predicting positive affect, negative affect and enjoyment respectively when compared to models trained only on system factors. Results and analysis indicate the significant role played by human factors in influencing affect that users experience while watching multimedia.

I. INTRODUCTION

Multimedia content produces diverse affective (emotional) responses in humans. When warmth and competence shape our judgements of people and organizations, and when perceived together they cause active behavioral responses from the viewers [27]. Daily we witness several organizations put forward their missions in the form of ad campaigns. While most of these ads fail to attract our attention, some of them leave a lasting impression in our minds. Take the example of the campaign by Volvo, which was listed as one of most unforgettable ad campaigns of 2013 [72] or Singapore's Ministry of Education 'Teach' campaign. The huge success of such ad campaigns is attributed to how story-telling components are shaped into emotion-evoking communication, structured to stimulate action.

Ad campaigns are but one specific scenario which illustrate the importance and challenge of modeling multimedia-evoked emotions. Publicity campaigns, movies, sports, educational material, games, to name a few, all require research into investigating a user's Quality of Experience (QoE) [42], [87], of which affect is an important dimension. Experience of affect

is defined as the positivity of emotions that viewers feel while watching videos. The problem is not just limited to content- or genre-based analysis of multimedia. This is because a video which arouses a positive emotion in one person might arouse a negative emotion in the other (depending on the nature of content and users' cultural and psychophysical frameworks which influence their perception) [86]. Whilst this is understood, how system parameters impact on the affective experience of those viewing multimedia content remains largely unexplored. What is also relatively unexplored is whether, and if so, to what degree, human factors also impact upon affective responses. These are the two main issues which we address in this paper - does multimedia content and system quality parameters with which it is presented evoke different affective responses depending on an individual's personality and culture?

Answering this question involves understanding the subjective nature of emotions and how crucial a role human factors play in modeling experience of affect (emotion), thereby addressing users' needs for emotion-sensitive video retrieval [19]. In this work, we attempt to understand how personality [52] and culture [39] influence users' experience of affect and enjoyment in multimedia. Specifically, the following research questions are posed:

- RQ 1. Can a model based on multimedia system characteristics (Bit-Rate, Frame-Rate and framesize) and human factors (i.e., personality and culture) predict the intensity of affect (both positive and negative) and enjoyment?
- RQ 2. Which system characteristics and human factors influence the experience of affect and enjoyment the most?
- RQ 3. What is the relationship between experience of affect (both positive and negative) and enjoyment across stimuli?
- RQ 4. How do predictive models perform on the task of automatic assessment of experience of affect and enjoyment of videos?

By investigating how different dimensions of these human factors modulate users' experience of affect and enjoyment, and specifically by understanding the correlation between enjoyment and perception of affect, we intend to provide initial findings for multimedia content creators to achieve maximal user satisfaction with respect to the content they create and deliver to diverse users.

Sharath Chandra Guntuku and Weisi Lin are with School of Computer Engineering, Nanyang Technological University, Singapore, 639798 E-mail: sharathc001@e.ntu.edu.sg, wslin@ntu.edu.sg .

Michael James Scott is with the Games Academy, Falmouth University, Cornwall, United Kingdom E-mail: michael.scott@falmouth.ac.uk.

Gheorghita Ghinea is with Department of Computer Science, Brunel University, London, United Kingdom E-mail: george.ghinea@brunel.ac.uk.

Manuscript received .

TABLE I: Datasets for Affective Modeling of Videos: Most of them implicitly assume that, given a video, the affect experienced by different users will be the same.

Dataset	Category	# Videos	# Annotators	Type of Annotation	Users' Profile	Enjoyment
Mutual Information-Based Emotion Recognition [22]	2 Mainstream Movies	655	-	One value each (0,1) for valence and arousal	No	No
Predicting Emotions in User-Generated Videos [44]	User Generated Videos	1101	-	Plutchik's emotions used as search keywords on Youtube and Flickr	No	No
A Connotative Space for Supporting Movie Affective Recommendation [7]	25 Mainstream Movies	25	240	Warmth of the scene atmosphere, Dynamic pace of the scene, Energetic impact	No	No
Music Video Affective Understanding Using Feature Importance Analysis [21]	Multilingual Music Videos	250	11	4 point valence, arousal	No	No
Utilizing Affective Analysis for Efficient Movie Browsing [92]	13 Mainstream Movies	4000	-	One value each (0,1) for valence and arousal	No	No
Affective visualization and retrieval for music video [91]	Music Videos	552	27	One value each (0,1) for valence and arousal	Individual user's ratings	No
Affective level video segmentation by utilizing the pleasure-arousal-dominance information [3]	13 Mainstream Movies	43	14	Ekman's 6 Emotions (1-10)	No	No
Emotional identity of movies [15]	87 Mainstream Movies	87	-	First 2 Genres from IMDB	No	No
Affective audio-visual words and latent topic driving model for realizing movie affective scene classification [40]	24 Mainstream Movies	206	16	Plutchik's 8 emotions (1-7)	No	No
Determination of emotional content of video clips by low-level audiovisual features [78]	24 Mainstream Movies	346	16	Pleasure Arousal Dominance values (1-7)	No	No
LIRIS-ACCEDE [6]	160 Creative Commons Movies	9800	1517	Rating-by-Comparison on Valence-Arousal	No	No
FilmStim [67]	70 Mainstream Movies	70	364	24 emotional classification criteria	No	No
MediaEval [65]	Travelogue series	126	-	Popularity/Boredom	No	No
Content-based prediction of movie style, aesthetics and affect: Data set and baseline experiment [77]	14 Mainstream Movies	14	73	ValenceArousal	No	No
DEAP [46]	Music Videos	40	32	PhysiologicalSignals, FaceVideos, Valence, Arousal	No	Yes
MAHNOB HCI [74]	Mainstream Movies	20	27	ValenceArousal, 6 emotion categories	No	No
CP-QAE-I (Our Dataset)	14 Mainstream Movies	144	114	16 sets of emotion-related adjectives from DES	Yes	Yes

II. RELATED WORK

There are several studies which aim to predict affective responses to multimedia (see [13], [83], [90] for a thorough review). Some focus on distilling the influence of specific cinematographic theories [14], types of segment and shot [8], the use of colour [85] and connotative space [7]. Apart from the works mentioned above, there has been research focused on modeling the different audio-visual features to predict emotions [49], [54], [68], [73], [76], [78]. The features used in this work are inspired by those used in the literature, along with certain content-based descriptors which have been shown to perform well in several content understanding tasks [11], [93].

Research on modeling emotional response in videos also often takes into account the facial expressions of viewers [18],

[60], [69], [84], [70] and a range of complementary sensors (e.g., heart rate, EEG) to help measure the evoked emotions [36], [46], [74]. However, the extent to which physiological responses capture the subjective intensity of affect (which varies as a consequence of users' innate psychology) is unclear.

A consequence of this is that such studies implicitly assume that, given a video, the affect experienced by different users will be more or less the same. This is equally the case with affective video datasets (as seen in Table I). However prior research shows that individual differences can lead to varied experiences [86]. To illustrate this, evidence reveals a complex relationship between affective video tagging and physiological signals [1], [19], [82]. As such, it is important to consider the subjective nature of affective perception. However, it is to be noted that we do not aim at creating a large-scale video

dataset for affective modeling, rather our aim is to understand the influence of users' individual traits on their perception of affect and consequently develop a dataset towards this goal.

Personality can be a good tool to explore the systematic differences in users' individual traits [52]. One popular model is the Five Factor Model (FFM) [31].

Certain traits are considerably influenced by the cultural background to which an individual belongs. Shared conceptions and collective norms characterize a local environment, and thereby shape the perception and cognition of those who associate with it. Differences in culture have been studied by Hofstede et al [39]. Six cultural traits constitute the model – masculinity, individualism, uncertainty avoidance index, pragmatism, power distance, and indulgence.

Both human factors targeted in our study, namely personality and culture, are shown to reliably capture individual differences in multiple domains like language [2], intonation of voice while speaking [57], [58], kind of photos one likes [34], type of people one befriends [30], etc. (see [80] for a thorough review). Other examples include preference of genre for language learning in different cultures [5] and the respective cultural acceptance of some movie content [20] etc. Due to the consistency shown between these human factors and user behaviors, we use them to study how they influence users' experience of affect and enjoyment in multimedia.

III. DATA COLLECTION

To address the concern of modeling the influence of individual differences on the experience of affect and enjoyment, we build a dataset using videos which are annotated by users with diverse personality and cultural traits. This section describes the videos, the procedure to used to collect annotations and the descriptive statistics.

A. Video Dataset

This study uses the CP-QAE-I dataset [35] (<http://1drv.ms/1M1bnwU>), which consists of 12 purposely selected short clips from popular movies to cover different affective categories [67]. Clips from a wide range of valence but low variance on arousal [67] have been adopted to reduce content-based biases, and the clips, their description and the means of positive and negative affect are given in Table II for the reader's reference. The content parameter also varies in cinematographic schemes utilized in the original movie production. There are three video coding parameters, namely bitrate – with the settings 384kbps and 768kbps, resolution (frame-size) – with the settings 480p and 720p, and framerate – with the settings 5fps, 15fps and 25fps, and these result in twelve coding-quality combinations. As a result, the dataset contains 144 (12*12) video sequences.

From the analysis with G*Power 3 [26] using the F -statistic as well as repeated measures, the minimum required sample size is 64, utilizing the conventional error probabilities ($\alpha = 0.05$, $\beta = 0.2$) and with the assumption of existing medium effects ($f = 0.39$) with $r = 0.8$ correlation.

TABLE III: BFI-10 questionnaire and associated Personality Traits: Each question is associated to a Likert scale. Questions are taken from [31].

Question	Trait
I have few artistic interests	O
I have an active imagination	O
I tend to find fault with others	A
I am generally trusting	A
I get nervous easily	N
I am relaxed, handle stress well	N
I am reserved	E
I am outgoing, sociable	E
I tend to be lazy	C
I do a thorough job	C

B. Procedure

The participants for data collection were 57 college students from each of the two universities with which the authors are affiliated (so totally 114 participants): 43 from Britain, 22 from India, 16 from China, 15 from Singapore and 18 from other nationalities. 28.9% of the participants were female and 23.9 years was the mean age, $\sigma = 3.68$. The corresponding cultural and personality traits are given in Table IV.

We applied a lab-based subjective testing approach. A set of videos were located locally on servers at the authors' universities. Users answered an online questionnaire from the local server (to avoid any latency issues over the Internet). Each user saw all 12 clips (in a random order), with different system characteristics and rated the experience of affect and their enjoyment of each sequence by completing questions immediately after viewing each. Informed consent and anonymity were assured at every stage of the study.

Since human factors are studied, we aim to maximize ecological validity in recording users' viewing behavior, so there was no limit on time to finish the survey. However, owing to the nature of such studies, some participants dropped out.

Participants started the survey by answering the BFI-10 [32] and the VSM-2013 [39] for assessment of personality and cultural traits. Afterwards, they were shown 12 videos under test, and were expected to give their ratings on all sequences. 73.7% of the 114 participants did so. However, all participants rated a minimum of 3 videos, with the average being 10.8 ($\sigma = 2.56$). Over all, 1,232 ratings were collected (90% of the maximum possible).

C. Measures

1) *Positive and Negative Affect*: was measured using Differential Emotions Scale [53]. This includes 16 sets of emotion-related adjectives. We refer the reader to [53] for the list of all sets. Each such set is linked to one of the 5 Likert scales and each participant rated the intensity of felt emotion. The emotions joy, warmth, love, calm, and so on were clubbed as positive affect, and anger, fear, anxiety, sadness, etc. were grouped as negative affect and their aggregate scores were calculated. Descriptive statistics on ratings are shown in Figures 1 and 2.

2) *Enjoyment*: This was measured using one of the 5-point Guttman-type scales, and each participant indicated how much

TABLE II: Marginal means of perceived responses (affect and enjoyment) on clips, after fixing the co-variates

MovieClip (Duration in Mins:Secs)	Description from [67]	+ve Affect	-ve Affect	Enjoyment
A_FISH_CALLED_WANDA (2:56)	One of the characters is found naked by the owners of the house	0.184	-0.536	-0.037
AMERICAN_HISTORY_X (1:06)	A neo-Nazi kills an African-American man, smashing his head on the curb	-0.397	0.756	-0.607
CHILDS_PLAY_II (1:07)	Chucky beats Andys teacher with a ruler	-0.231	0.698	-0.158
COPYCAT (1:04)	One of the characters gets caught by a murderer in a toilet	-0.33	0.418	-0.315
DEAD_POETS_SOCIETY_1 (2:34)	A schoolboy commits suicide	-0.331	0.341	-0.504
DEAD_POETS_SOCIETY_2 (2:23)	All the students climb on their desks to express their solidarity with Mr Keating, who has just been fired	1.053	-0.553	0.725
FOREST_GUMP (1:47)	Father and son are reunited	0.992	-0.523	0.656
SE7EN_1 (1:39)	By the end of the movie, Kevin Spacey tells Brad Pitt that he beheaded his pregnant wife	-0.346	0.248	0.42
SE7EN_3 (0:24)	Policemen find the body of a man tied to a table	-0.431	0.03	-0.306
SOMETHING_ABOUT_MARY (2:00)	Mary takes sperm from Teds hair mistaking it for hair gel	0.468	-0.72	0.471
THE_PROFESSIONAL (2:44)	The two main characters are separated forever	-0.194	0.216	0.254
TRANSPOTTING (0:40)	The main character dives into a filthy toilet	-0.477	-0.389	-0.654

Based on estimated marginal means of a mixed-effects regression model. Covariates in the model are evaluated at the following values: AGREEABLENESS = 7.45; EXTRAVERSION = 5.42; CONSCIENTIOUSNESS = 6.59; OPENNESS = 6.77; NEUROTICISM = 5.67; POWER DISTANCE = -34.29; MASCULINITY = -6.73; INDIVIDUALISM = 22.44; UNCERTAINTY AVOIDANCE = 40.83; INDULGENCE = -11.60 PRAGMATISM = 22.82;.

TABLE IV: Sample Descriptives.

Human Factors	Min	Max	$\bar{x}(NTU)$	$\bar{x}(BUL)$	\bar{x}	σ
Openness	4	10	6.60	6.91	6.75	1.424
Conscientiousness	2	10	6.40	6.70	6.55	1.523
Extroversion	2	9	5.61	5.46	5.54	1.689
Neuroticism	2	10	5.56	5.68	5.62	1.716
Agreeableness	3	10	7.33	7.31	7.22	1.533
Individualism	-140	140	25.79	11.67	18.73	50.619
Power Distance	-155	140	-35.61	-36.32	-35.96	53.219
Masculinity	-140	105	3.68	-6.14	-1.23	53.483
Pragmatism	-130	155	16.14	17.54	16.84	58.090
Uncertainty Avoidance	-120	130	52.54	36.67	44.61	47.182
Indulgence	-220	185	-22.63	-11.32	-16.97	65.522

he/she enjoyed a video sequence. A value of 1 represents “no” enjoyment and a value of 5 denotes “high” enjoyment. Descriptive statistics on ratings are shown in Figure 3.

3) *Culture*: It was measured using the VSM-2013 questionnaire [37] according to the following aspects: individualism (IDV), power distance (PDI), uncertainty avoidance (UAI), pragmatism (PRG), masculinity (MAS), and indulgence (IVR). We refer the reader to [38] for the list of all questions and how they relate to different cultural traits. Most of these questions are on a scale of 1-5.

4) *Personality*: This was measured using the BFI-10 [32] questionnaire, according to the FFM [31], measuring conscientiousness (Con), openness (Ope), Extroversion (Ext), Neuroticism (Neu), and Agreeableness (Agr). The questions of BFI-10 along with the corresponding traits is shown in Table III. Most of these questions are on a scale of 1-5.

IV. METHODOLOGY

In this section, we introduce different statistical methods used, the features extracted to build the predictive models and the evaluations.

A. Statistical Analysis

The analysis has been conducted in PASW 20.0. Linear mixed-effects modeling has been adopted for repeated measures, with model parameters determined with the restricted maximum-likelihood method.

We build three computational models (namely baseline, extended and optimistic) to investigate the influence of system factors – framerate, bitrate and resolution, human factors – five personality factors and six culture factors, on the experience of affect and enjoyment. Each of them along with corresponding findings will be described in this section. Afterwards a comparison between the three models will be presented to address the four questions we pose in this paper.

1) *Baseline Model*: This model considers only system factors. For the CP-QAE-I video dataset, there are 12 variations of the system factors – framerate, bitrate, and resolution/framesize. Factors such as format of the files and network protocol were held constant. Due to the expected interactions between these conditions (e.g., an attempt to minimise Bit-Rate while maximising Frame-Rate and framesize would likely create artefacts), they have been modelled as factorial interactions. In addition, the movie clip itself is included as a parameter to reflect differences in cinematographic schemes used to create the movies, along with the nature of the content. This was modelled as a main effect.

2) *Extended Model*: The extended model adds additional fixed parameters to the baseline model. These were cultural traits: individualism, power distance, masculinity, pragmatism, uncertainty avoidance, and indulgence. Additionally, personality traits were also added: extroversion, agreeableness, conscientiousness, neuroticism, and openness. These were incorporated into the model as covariates with direct effects.

3) *Optimistic Model*: While a model aims at predicting a dependent variable as precisely as possible, not all of the residual variance can be solely attributed to human factors. A non-trivial proportion of the residual variance can also, to

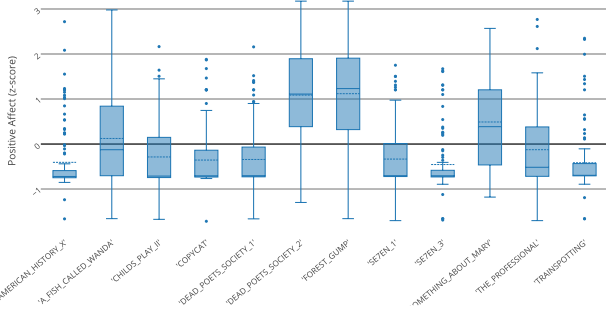


Fig. 1: Distribution of positive affect in the dataset. Mean is represented by the dotted line.

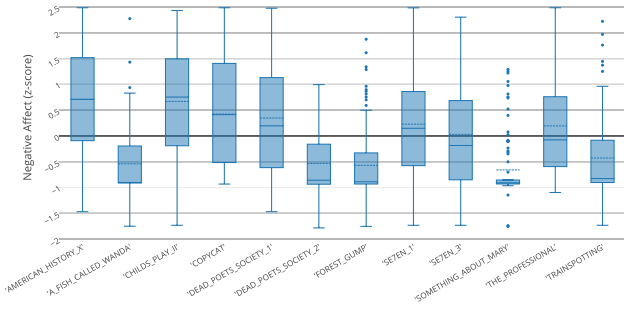


Fig. 2: Distribution of negative affect in the dataset. Mean is represented by the dotted line.

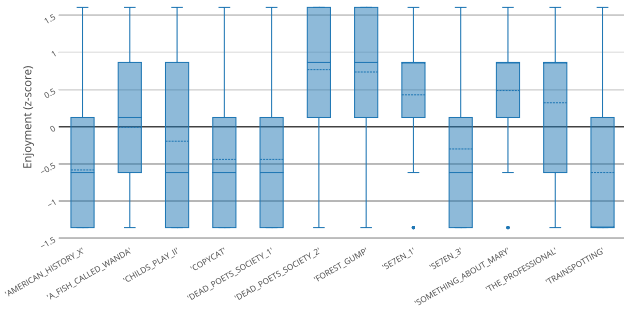


Fig. 3: Distribution of enjoyment in the dataset. Mean is represented by the dotted line.

name but a few, be attributed to random error, measurement error, and the limitations of the modelling technique (in this case, generalised linear regression). As such, an optimistic model can be used to estimate the part of the residual variance that might possibly be attributed to human factors in general and, to a small extent, because of the factors such as context and limitations in experimental control. This is achieved by modelling every participant as a random effect, i.e., measurements have been repeated to obtain a different intercept during the regression for every participant.

B. Predictive Modeling

While statistical analysis provides an understanding of the relationships of different dependent and independent variables in the data, predictive models help to forecast users' responses on new data samples.

We propose a prediction framework which takes input features in video content, system characteristics, personality and cultural traits, predicting experience of both positive and negative affect, and enjoyment on the video clips using L_1 regularised L_2 -loss sparse SVM with a linear kernel. We chose a linear kernel to avoid the problem of overfitting (as seen in literature [11], [24]). We use the libsvm framework [17] to carry out these experiments.

1) *Features*: We use 4 categories of features to represent content-based, system-based, affective and human factors, in an attempt to describe various features which might influence the experience of affect and enjoyment. They are described as follows.

Content-factors: To represent the different concepts in the video, we used HybridCNN [93], Adjective-Noun Pairs [11], Color histograms, Bag-of-Visual-Words [88], Aesthetic features [8], [50] and LBP features [62]. Representation for every video is gotten by mean-pooling on features obtained from all frames.

Color Histogram: Users' perception is greatly influenced by the color hues in the videos. Therefore, color histograms are chosen to represent users' color inclination.

Aesthetic Features: We used 2 sets of features to represent the aesthetic characteristics in videos: a.) art and psychology based features [50] to describe photographic styles (rule-of-thirds, vanishing points, etc.). b.) psycho-visual characteristics at cell, frame and shot levels proposed by [8].

LBP: LBP was used to encode users' perception of texture in the videos. As LBP represents facial information well and many of the videos have people, we use LBP features.

Bag-of-Visual-Words [88]: A 1500 dimension feature based on vector quantization of keypoint descriptors is generated for every frame. A bag of visual words representation is used by mapping the keypoints to visual words.

HybridPlacesCNN: CNN features from fc7-ReLu layer from ImageNet [47] and Places dataset [93] are used to represent objects and places in videos.

Adjective-Noun Pairs (ANP): Emotion based attributes are detected (2089 ANPs) [11] using the sentibank classifiers. 8 emotion categories [64] are used to define the adjectives, and objects and scenes are used to define the nouns.

Audio Affect-factors: Affective characteristics associated with the audio content in videos are extracted (OpenSmile [25]) as musical chroma features [59], prosodic features [16] and low-level descriptors such as MFCC features, intensity, loudness, pitch, pitch envelope, probability of voicing, line spectral frequencies, and zero-crossing rate. The visual affective content in the videos was expected to be represented by the aesthetic and ANP features described above.

System-factors: Bitrate, Framerate, resolution and perceptual characteristics [56] are used to represent quality characteristics of videos. Perceptual characteristics describe the

TABLE V: List of Features

Category	Feature	Description
Visual Content	Color Histogram	based on RGB values of each frame
	Visual Aesthetics	art and psychology based features to describe photographic styles (rule-of-thirds, vanishing points, etc.); psycho-visual characteristics at cell, frame and shot levels
	LBP	users' perception of texture in the videos
	BoVW	1500-dimension vector based on quantization of keypoint descriptors
	HybridPlaces CNN	features from the ReLu layer following fc7 layer of a CNN trained on 1.3 million images from ImageNet and 2.5 million images on the Places dataset
	Sentibank	detection of 2089 Adjective Noun Pairs based on emotion related concepts
Audio Content	Musical chroma	Opensmile was used to extract the affective features from audio signal
	Prosody	
	Low-level descriptors	intensity, MFCC, loudness, pitch, pitch envelope, probability of voicing, zero-crossing rate and line spectral frequencies
System parameters	Bit Rate, Frame Rate, Resolution	Quality of the distorted image is expressed as a simple distance metric between the model statistics and those of the distorted image
	Perceptual Characteristics	
Human factors	Personality	openness, conscientiousness, extraversion, agreeableness, neuroticism
	Cultural traits	power distance, masculinity, individualism, uncertainty avoidance, indulgence, pragmatism
	Demography	gender, age and nationality

no-reference quality metric [55]. Perceptual characteristics were represented by temporal distortions in the video, spatial domain natural scene statistics, statistical DCT features motion coherence feature describing the coherence in strength and direction of local motion due to temporal distortions, and reflecting the perceptual difference between pristine and distorted videos.

Human-factors: The five personality factors, six cultural traits, gender, age and nationality of the users are used to represent human factors.

V. RESULTS AND DISCUSSION

Results target at answering the four research questions raised at the outset of the paper. Sections V-A and V-A4 deal with RQ1 and RQ2, Section V-A5 deals with RQ3, and Section V-B answers RQ4.

A. Statistical Analysis

1) *Baseline Model:* Table VI shows the results. The movie clip itself had the largest impact on experience of affect and enjoyment. However, an interesting observation is that only Frame Rate had a statistically significant effect on enjoyment. This shows that system factors alone do not make a huge impact on how the content is perceived. That is, given two videos of different natures at different bitrate, resolution/framesize and framerate, the nature of the content alone is more likely to influence how it is perceived than the system settings at which it is delivered. Our findings can be corroborated by similar observations in QoE [28], [89].

2) *Extended Model:* Table VII shows the results of the extended model. Many personality and cultural traits seem to be significant predictors of experience of affect and enjoyment. Among personality traits, extraversion and conscientiousness are significant predictors for positive affect, and agreeableness, neuroticism and conscientiousness are significant for negative

affect. Conscientiousness and openness are significant predictors for enjoyment [41], [63]. Among cultural traits, masculinity and indulgence are significant predictor for positive affect, indulgence alone for negative affect and uncertainty avoidance for enjoyment. None of the system factors (except Frame-Rate for enjoyment) and their interactions are significant predictors.

This suggests that multimedia system characteristics have little to no influence on the intensity of the affect that viewers experience. Additionally, there appears to be a different set of predictors for affect compared to overall enjoyment. F -statistic is generally quite small for most of the predictors. However, the predictors of agreeableness and neuroticism, for negative affect, are notably much larger. This suggests that a considerable amount of the variance in negative affect can be explained by these parameters.

3) *Optimistic Model:* Table VIII shows the results of the optimistic models. The model is quite similar to the baseline model with the exception of larger F -statistics, indicating that larger proportion of variance is explained because of considering random effects. Additionally, interaction between framesize and experience of affect is now significant.

4) *Model Comparison:* Paired t-tests on Mean Squared Residuals (MSR) are used to compare the models as shown in Table IX. Proportional reduction in overall MSE is shown (see [12]). The results show that human factors, namely personality and culture, play a crucial role in modeling the experience of affect and enjoyment, indicating that content production and delivery mechanisms should not just take into account multimedia system factors but also human factors to achieve maximal user satisfaction.

Models for Positive Emotion: From the baseline to optimistic model, the MSR reduced from 0.6304 ($\sigma = 1.050$) to 0.4051 ($\sigma = 0.886$; $p < 0.005$), representing a predicted variance of 55.3%. A part of this is contributed by culture and personality. 5.6% of variance attributable to human factors is predicted by the extended model, reducing the baseline MSR to 0.6177 ($\sigma = 1.005$; $p = 0.021$).

TABLE VI: Baseline fixed-effect multilevel linear regression model

Parameter	df_{num}	Positive Affect			Negative Affect			Enjoyment		
		df_{den}	F	p	df_{den}	F	p	df_{den}	F	p
Movie Clip	11	156.009	25.315	0.00	144.643	33.932	0.00	177.09	40.14	0.00
Frame Rate (FR)	2	803.739	0.32	0.73	710.192	0.056	0.95	1131.23	5.173	0.006
Frame Size (FS)	1	809.889	0.006	0.94	729.398	3.298	0.07	1146.39	2.846	0.092
Bit-Rate (BR)	1	816.675	1.724	0.19	714.909	0.30	0.58	1139.69	0.474	0.491

Interactions of System Factors namely FR \times FS, FS \times BR, FR \times BR, FR \times FS \times BR were found to be insignificant predictors and hence not included in the above table.

TABLE VII: Extended fixed-effect multilevel linear regression model

Parameter	df_{num}	Positive Affect			Negative Affect			Enjoyment		
		df_{den}	F	p	df_{den}	F	p	df_{den}	F	p
Movie Clip	11	193.163	35.925	0.00	206.260	39.739	0.00	171.956	39.733	0
Frame Rate (FR)	2	1071.695	0.18	0.84	1045.660	0.48	0.62	1136.577	4.695	0.009
Frame Size (FS)	1	1074.152	0.54	0.46	1061.874	2.10	0.15	1151.402	3.336	0.068
Bit-Rate (BR)	1	1083.535	2.334	0.13	1044.851	0.06	0.807	1145.171	0.257	0.612
Extraversion	1	1074.324	4.559	0.033	1059.767	0.08	0.78	1150.401	0.024	0.877
Agreeableness	1	1072.223	1.876	0.17	1059.481	24.314	0.00	1152.475	2.001	0.157
Conscientiousness	1	1077.950	9.474	0.002	1041.655	3.964	0.047	1141.249	5.271	0.022
Neuroticism	1	1084.026	0.02	0.888	1050.845	25.227	0.00	1146.479	0.05	0.823
Openness	1	1074.213	2.670	0.103	1058.628	2.110	0.147	1145.365	4.344	0.037
Power Distance	1	1073.888	4.676	0.031	1055.500	0.00	0.985	1152.465	9.138	0.003
Individualism	1	1070.708	2.148	0.143	1052.462	2.486	0.115	1150.026	0.674	0.412
Masculinity	1	1074.304	4.874	0.027	1043.258	1.061	0.303	1141.312	3.312	0.069
Uncertainty Avoidance	1	1077.284	0.534	0.465	1044.360	0.306	0.580	1144.106	5.751	0.017
Pragmatism	1	1069.661	0.886	0.347	1064.578	0.175	0.676	1160.7	0.604	0.437
Indulgence	1	1070.162	5.863	0.016	1051.545	4.863	0.028	1149.178	2.206	0.138

Interactions of System Factors namely FR \times FS, FS \times BR, FR \times BR, FR \times FS \times BR were found to be insignificant predictors and hence not included in the above table.

TABLE VIII: Optimistic mixed-effect multilevel linear regression model

Parameter	df_{num}	Positive Affect			Negative Affect			Enjoyment		
		df_{den}	F	p	df_{den}	F	p	df_{den}	F	p
Movie Clip	11	178.713	42.312	0.00	152.624	55.782	0.00	179.877	46.99	0.00
Frame Rate (FR)	2	701.036	1.788	0.168	945.140	1.392	0.249	1116.89	8.025	0.00
Frame Size (FS)	1	695.825	0.002	0.965	969.366	5.764	0.017	1120.818	3.13	0.077
Bit-Rate (BR)	1	715.664	1.159	0.282	972.050	1.457	0.228	1121.96	0.054	0.816

Interactions of System Factors namely FR \times FS, FS \times BR, FR \times BR, FR \times FS \times BR were found to be insignificant predictors and hence not included in the above table.

Models for Negative Emotion: From the baseline to optimistic model, the MSR reduced from 0.6514 ($\sigma = 0.889$) to 0.3615 ($\sigma = 0.536$; $p < 0.00$), representing a predicted variance of 58.1%. 13.6% of variance attributable to human factors, reducing the baseline MSR to 0.6118 ($\sigma = 0.8278$; $p < 0.00$).

Models for Enjoyment: From the baseline to optimistic model, the MSR reduced from 1.3684 ($\sigma = 1.63$) to 0.9481 ($\sigma = 1.22$; $p < 0.00$), which makes up 23.0% of the overall variance predicted. 9.3% of variance due to human factors is predicted by the extended model, which decreases the baseline MSR to 1.3290 ($\sigma = 1.58$; $p < 0.001$).

5) *Correlation between Affect and Enjoyment:* As introduced at the beginning of the article, there is a very close and significant relationship between what users enjoy and the emotion it evokes (results from correlation analysis are shown in Table X). In all clips, enjoyment is significantly correlated with interest, joy, satisfaction and the latent factor, positive emotion. This means that for a user to enjoy a video the content has to definitely draw his/her interest, but must also have moments of happiness and deliver something which satisfies the viewer [75].

There are also very few instances of negative emotions (sad, fearful, guilty, and ashamed) giving enjoyment to users. For instance, enjoyment was seen to have a significant positive correlation with emotions Ashamed and Guilty for both the clips

TABLE IX: Paired t-test showing the comparison of models for all three responses (w.r.t MSR)

Models	Positive Affect				Negative Affect				Enjoyment			
	$\Delta\bar{x}$	σ	t	p	$\Delta\bar{x}$	σ	t	p	$\Delta\bar{x}$	σ	t	p
<i>Baseline</i> \rightarrow <i>Extended</i>	0.013	0.193	2.311	0.021	0.039	0.277	5.008	0.00	0.039	0.430	3.219	0.001
<i>Baseline</i> \rightarrow <i>Optimistic</i>	0.2253	0.924	8.552	0.00	0.2898	0.726	14.014	0.00	0.4199	1.129	13.069	0.00

TABLE X: Significant correlations between enjoyment and experience of affect ($p < 0.05$)

Clip	Interest	Joy	Sad	Fearful	Disgust	Surprise	Warm	Loving	Guilty	Moved	Satisfied	Calm	Ashamed	+ve Affect
C-I	0.579	0.699	-	-	-0.332	0.430	0.626	0.419	-	-	0.456	0.451	-	0.610
C-II	0.505	0.304	-	-	-0.332	0.380	0.350	0.383	0.258	-	0.526	0.404	0.291	0.483
C-III	0.596	0.485	-	-	-0.314	0.250	0.298	0.247	-	-	0.432	0.286	-	0.426
C-IV	0.444	0.424	-	-	-	-	0.287	0.243	-	-	0.255	-	-	0.304
C-V	0.469	0.263	-	-	-	-	0.368	0.271	0.244	0.219	0.288	0.232	0.250	0.298
C-VI	0.514	0.385	-	-0.239	-	-	0.353	0.325	-	-	0.493	-	-0.215	0.456
C-VII	0.549	0.643	0.248	-	-	0.293	0.479	0.513	-	0.508	0.510	0.407	-	0.560
C-VIII	0.550	0.408	-	-	-	-	-	-	-	0.346	0.445	0.319	-	0.421
C-IX	0.340	0.394	-	-	-	-	0.323	0.251	-	-	0.219	0.244	-	0.292
C-X	0.512	0.658	-	-	-	0.258	0.301	-	-	-	0.267	0.354	0.290	0.419
C-XI	0.541	0.266	-	-	-	0.232	-	-	-	0.277	0.358	0.325	-	0.347
C-XII	0.590	0.688	-	-	-0.401	0.317	0.434	0.419	0.244	0.295	0.488	0.353	-	0.542

Movie Clips:- C-I: A_FISH_CALLED_WANDA; C-II: AMERICAN_HISTORY_X; C-III: CHILDS_PLAY_II; C-IV: COPYCAT; C-V: DEAD_POETS_SOCIETY_1; C-VI: DEAD_POETS_SOCIETY_2; C-VII: FOREST_GUMP; C-VIII: SE7EN_1; C-IX: SE7EN_3; C-X: SOMETHING_ABOUT_MARY; C-XI: THE_PROFESSIONAL; C-XII: TRAINSPOTTING. No significant correlations were observed between Enjoyment and Anxious, Angry, Disdain & -ve Affect and thus are not shown in the table. Entry $E_{i,j}$ represents the correlation between enjoyment and affect for movie clip i and emotion category j

AMERICAN_HISTORY_X (in which the protagonist is seen brutally torturing someone) and DEAD_POETS_SOCIETY (in which one of the main characters commits suicide by shooting himself) and with emotion Ashamed for SOMETHING_ABOUT_MARY (in which there is an obscenity involved and yet sounds joyful/funny) and Guilty for TRAINSPOTTING (in which a person is seen to get inside a dirty toilet bowl, and yet the music is of a totally different contrast). The predominant emotion in these clips are not widely enjoyable. However, these might be associated with how certain users (possibly with high scores on neuroticism) perceive certain contents [23], [48]

Apart from that, even the nature of the content itself can arouse contradictory emotions. For example, enjoyment is observed to be positively correlated with sadness in the movie clip FOREST_GUMP which has a defining excerpt in which the leading character encounters his son for the very first time. This is a scene with bitter-sweet connotations for viewers due to the fact that the protagonist, quite belatedly in his life, is faced with the news that not only has he fathered a son, but also that the son is doing well in school and is a fine student. So, such occurrences are due to the interaction of human factors and nature of the content both.

It is interesting to note that while most of the users might associate enjoyment with positivity, there are certain users who need to experience negative emotions to connect to the content's message. This insight gives content creators a better understanding of how to influence users with different personality and cultural traits to establish an emotional connection with them, which is very important to drive behavioral action

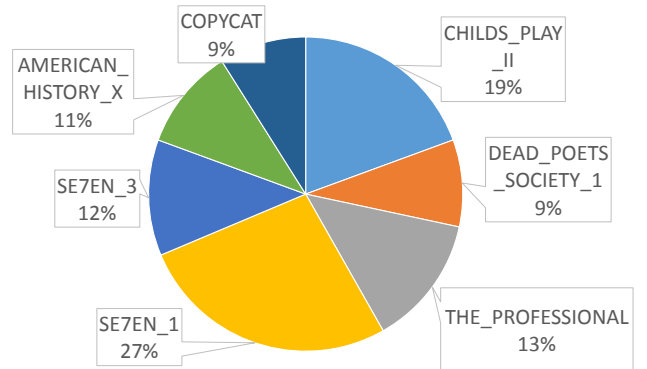


Fig. 4: Distribution of ratings on movie clips with high enjoyment and high -ve affect.

(especially in scenarios involving ad campaign design etc.).

To investigate this further, we selected the records with high enjoyment (i.e., 4 and 5) and with high negative affect (above mean) and low positive affect (below mean). The filtered subset (about 5.4% of the total ratings) was investigated to understand the nature of content which likely influences such rating behavior. Figure 4 shows distribution of clips in the filtered subset.

B. Predictive Models

Experiments are run on leave-one-video-out setting by binarising perceptual quality and enjoyment scores. Figure 5 depicts the accuracy with which the trained models predict user

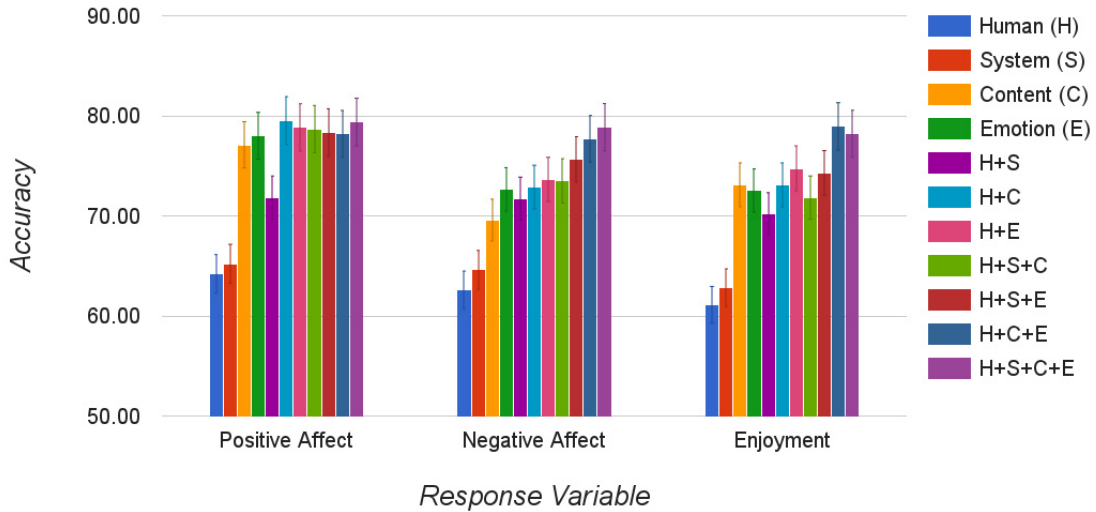


Fig. 5: Predictive models' accuracy when employing different features.

affect and enjoyment. Human, system, content and emotion factors (the latter based on audio affect) were employed, and features extracted to represent the same.

As far as positive and negative affect are concerned, models which gave the best performance were those trained on emotion and content factors (in this order). This is intuitive, as the nature of content plays the most important role in influencing viewers' experience of affect. It was also seen in our statistical analysis (in the previous section).

It was then explored if adding human factors to those pertaining to system, content and emotion would improve the predictive modeling performance.

In terms of positive affect, the combination of human and content, followed by that of human and emotion factors yielded the best results. It must be noted that the features representing content factors also include ANPs, which are especially designed for visual sentiment prediction [11]. This explains their almost equivalent performance when compared to emotion factors. Similar performance was seen when three factors (namely content, human and emotion) are combined to train the model. Similar observations are made for negative affect, however the performance was lower than that of predicting positive affect. This is similar to the observations made in other works [33], [45], [66], [71], [79], possibly due to the intrinsic challenge of modeling the nature of negative emotions.

As far as enjoyment is concerned, models trained on human, emotion, and content factors performed better than others. Our statistical analysis showed evidence of this, as content was found to be significantly correlated with enjoyment . The

impact of factors associated with emotion has been explored in other studies[4], [9], [81].

VI. CONCLUSION

Experience of affect and enjoyment in multimedia is influenced by an intricate interplay between characteristics of stimuli, individuals, and systems of perception. Returning to the research questions posed at the outset of the paper, we can now state that:

- RQ1 For positive affect, negative affect and enjoyment, personality and culture respectively represented 5.6%, 13.6% and 9.3% of variance. Notwithstanding the fact that these constitute sizeable proportions, follow up studies need to explore other potential contributing factors, such as sensory impairments/acuity, user cognitive style, and domain expertise [29].
- RQ2 Traits of extraversion, conscientiousness, masculinity and indulgence are significant predictors for positive affect, and agreeableness, neuroticism, conscientiousness and indulgence were important predictors for negative affect. Conscientiousness, openness and uncertainty avoidance were significant predictors for enjoyment.
- RQ3 The majority of the movie clips which were enjoyed were also rated high on positive affect, with a small exception of clips having high correlation between negative affect and enjoyment. Such behavior is possibly due to the interchange that potentially takes

place between human factors (e.g. neuroticism) and media content.

RQ4 Predictive models trained with a mixture of human factors and content, emotion and emotion factors yielded the highest achievement for positive affect, negative affect and enjoyment respectively with an accuracy of 79%, 77% and 76% respectively.

It is important to know the impact of human factors on user enjoyment, as this allows one to optimise this latter parameter especially in conditions in which other more traditional forms of adaptation (such as layered adaptation) are difficult/impractical to perform. As highlighted above, results obtained in our study showcase the important part that human factors have on two important aspects of user QoE, namely affect and enjoyment. Thus, integration of human factors in the optimistic model was shown to have significantly improved modeling performance; however, extended models based on personality and culture did not have impacts of the same magnitude. This means that several other human factors, apart from those taken into consideration in this study such as user behavior in a particular domain of study (e.g. movies rated or images liked) or other psychological constructs like mood etc. [10], [51], [61] can be explored. Nonetheless, results show that human factors, namely personality and culture, exert an important influence in modeling the experience of affect and enjoyment, indicating that content production and delivery mechanisms should not just take into account multimedia system factors but also human factors, in order to achieve maximal user satisfaction.

REFERENCES

- [1] Mojtaba Khomami Abadi, Seyed Mostafa Kia, Ramanathan Subramanian, Paolo Avesani, and Nicu Sebe. User-centric affective video tagging from meg and peripheral physiological responses. In *Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2013.
- [2] Shlomo Argamon, Sushant Dhawle, Moshe Koppel, and James Pennebaker. Lexical predictors of personality type. 2005.
- [3] Sutjipto Arifin and Peter YK Cheung. Affective level video segmentation by utilizing the pleasure-arousal-dominance information. *Multimedia, IEEE Transactions on*, 10(7):1325–1341, 2008.
- [4] Shaowen Bardzell, Jeffrey Bardzell, and Tyler Pace. Understanding affective interaction: Emotion, engagement, and internet videos. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–8. IEEE, 2009.
- [5] Saham Barza and Mehran Memari. Movie genre preference and culture. *Procedia-Social and Behavioral Sciences*, 2014.
- [6] Yoann Baveye, J-N Bettinelli, Emmanuel Dellandréa, Liming Chen, and Christel Chamaret. A large video database for computational models of induced emotion. In *Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2013.
- [7] Sergio Benini, Luca Canini, and Riccardo Leonardi. A connotative space for supporting movie affective recommendation. *Multimedia, IEEE Transactions on*, 13(6):1356–1370, 2011.
- [8] Subhabrata Bhattacharya, Behnaz Nojavanasghari, Tao Chen, Dong Liu, Shih-Fu Chang, and Mubarak Shah. Towards a comprehensive computational model for aesthetic assessment of videos. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 361–364. ACM, 2013.
- [9] Helena Bilandzic and Rick W Busselle. Enjoyment of films as a function of narrative experience, perceived realism and transportability. *Communications*, 36(1):29–50, 2011.
- [10] Susanne Boll. Multitube—where web 2.0 and multimedia could meet. *IEEE MultiMedia*, (1):9–13, 2007.
- [11] Damian Borth, Tao Chen, Rongrong Ji, and Shih-Fu Chang. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 459–460. ACM, 2013.
- [12] Roel Bosker and Tom Snijders. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*, 2nd ed. New York, 2012.
- [13] Rafael Calvo, Sidney D’Mello, et al. Affect detection: An interdisciplinary review of models, methods, and their applications. *Affective Computing, IEEE Transactions on*, 2010.
- [14] Luca Canini, Sergio Benini, and Riccardo Leonardi. Affective recommendation of movies based on selected connotative features. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(4):636–647, 2013.
- [15] Luca Canini, Sergio Benini, Pierangelo Migliorati, and Riccardo Leonardi. Emotional identity of movies. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1821–1824. IEEE, 2009.
- [16] Michael J Carey, Eluned S Parris, Harvey Lloyd-Thomas, and Stephen Bennett. Robust prosodic features for speaker identification. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pages 1800–1803. IEEE, 1996.
- [17] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.
- [18] Pojala Chiranjeevi, Viswanath Gopalakrishnan, and Pratibha Moogi. Neutral face classification using personalized appearance models for fast and robust emotion detection. *Image Processing, IEEE Transactions on*, 24(9):2701–2711, 2015.
- [19] Wonhee Choe, Hyo-Sun Chun, Junhyug Noh, Seong-Deok Lee, and Byoung-Tak Zhang. Estimating multiple evoked emotions from videos. In *Annual Meeting of the Cognitive Science Society*, 2013.
- [20] Samuel Craig, William H Greene, and Susan P Douglas. Culture matters: consumer acceptance of us films in foreign markets. *Journal of International Marketing*, 13(4):80–103, 2005.
- [21] Yue Cui, Jesse S Jin, Shiliang Zhang, Suhuai Luo, and Qi Tian. Music video affective understanding using feature importance analysis. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, pages 213–219. ACM, 2010.
- [22] Yue Cui, Suhuai Luo, Qi Tian, Shiliang Zhang, Yu Peng, Lei Jiang, and Jesse S Jin. Mutual information-based emotion recognition. In *The Era of Interactive Media*, pages 471–479. Springer, 2013.
- [23] Ed Diener, Shigehiro Oishi, and Richard E Lucas. Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual review of psychology*, 2003.
- [24] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. *arXiv preprint arXiv:1310.1531*, 2013.
- [25] Florian Eyben, Martin Woellmer, and Bjoern Schuller. The munich open speech and music interpretation by large space extraction toolkit. 2010.
- [26] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191, 2007.
- [27] Susan T Fiske, Amy JC Cuddy, and Peter Glick. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 2007.
- [28] George Ghinea and Johnson P Thomas. Qos impact on user perception and understanding of multimedia video clips. In *Proceedings of the sixth ACM international conference on Multimedia*, pages 49–54. ACM, 1998.
- [29] Gheorghita Ghinea and Sherry Y Chen. Perceived quality of multimedia educational content: A cognitive style approach. *Multimedia systems*, 11(3):271–279, 2006.
- [30] Jennifer Golbeck, Cristina Robles, and Karen Turner. Predicting personality with social media. In *CHI extended abstracts*. ACM, 2011.
- [31] Lewis R Goldberg. An alternative” description of personality”: the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216, 1990.
- [32] Samuel D Gosling, Peter J Rentfrow, and William B Swann. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.
- [33] Hatice Gunes, Björn Schuller, Maja Pantic, and Roddy Cowie. Emotion representation, analysis and synthesis in continuous space: A survey. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 827–834. IEEE, 2011.
- [34] Sharath Chandra Guntuku, Sujoy Roy, and Weisi Lin. Personality modeling based image recommendation. In *MultiMedia Modeling*, pages 171–182. Springer, 2015.

- [35] Sharath Chandra Guntuku, Michael James Scott, Huan Yang, Gheorghita Ghinea, and Weisi Lin. The CP-QAE-I: A video dataset for exploring the effect of personality and culture on perceived quality and affect in multimedia. In *QoMEX*. IEEE, 2015.
- [36] Junwei Han, Xiang Ji, Xintao Hu, Dajiang Zhu, Kaiming Li, Xi Jiang, Guangbin Cui, Lei Guo, and Tianming Liu. Representing and retrieving video shots in human-centric brain imaging space. *Image Processing, IEEE Transactions on*, 22(7):2723–2736, 2013.
- [37] Geert Hofstede, Gert Jan Hofstede, Michael Minkov, and Henk Vincken. Values survey module 2013. URL: <http://www.geerthofstede.nl/vsm2013>, 2013.
- [38] Geert Hofstede and Michael Minkov. Vsm 2013. *Values survey module*, 2013.
- [39] Geert Hofstede, Gert Jan Hofstede, and Michael Minkov. *Cultures and organizations: software of the mind: intercultural cooperation and its importance for survival*. McGraw-Hill, 2010.
- [40] Go Irie, Takashi Satou, Akira Kojima, Toshihiko Yamasaki, and Kiyoharu Aizawa. Affective audio-visual words and latent topic driving model for realizing movie affective scene classification. *Multimedia, IEEE Transactions on*, 12(6):523–535, 2010.
- [41] Carroll E Izard, Deborah Z Libero, Priscilla Putnam, and O Maurice Haynes. Stability of emotion experiences and their relations to traits of personality. *Journal of personality and social psychology*, 1993.
- [42] Ramesh Jain. Quality of experience. *IEEE MultiMedia*, 11(1):96–95, 2004.
- [43] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.
- [44] Yu-Gang Jiang, Baohan Xu, and Xiangyang Xue. Predicting emotions in user-generated videos. In *AAAI*, pages 73–79, 2014.
- [45] Brendan Jou, Subhabrata Bhattacharya, and Shih-Fu Chang. Predicting viewer perceived emotions in animated gifs. In *Proceedings of the ACM International Conference on Multimedia*, pages 213–216. ACM, 2014.
- [46] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *Affective Computing, IEEE Transactions on*, 2012.
- [47] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [48] Randy J Larsen and Timothy Ketelaar. Personality and susceptibility to positive and negative emotional states. *Journal of personality and social psychology*, 1991.
- [49] Bing Li, Weihua Xiong, Ou Wu, Weiming Hu, Stephen Maybank, and Shuicheng Yan. Horror image recognition based on context-aware multi-instance learning. *Image Processing, IEEE Transactions on*, 24(12):5193–5205, 2015.
- [50] Jana Machajdik and Allan Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, pages 83–92. ACM, 2010.
- [51] Pamela AF Madden, Andrew C Heath, Norman E Rosenthal, and Nicholas G Martin. Seasonal changes in mood and behavior: the role of genetic factors. *Archives of General Psychiatry*, 53(1):47–55, 1996.
- [52] Gerald Matthews, Ian J Deary, and Martha C Whiteman. *Personality traits*. Cambridge University Press, 2003.
- [53] Gregory J McHugo, Craig A Smith, and John T Lanzetta. The structure of self-reports of emotional responses to film segments. *Motivation and Emotion*, 6(4):365–385, 1982.
- [54] Angeliki Metallinou, Martin Wöllmer, Athanasios Katsamanis, Florian Eyben, Björn Schuller, and Shrikanth Narayanan. Context-sensitive learning for enhanced audiovisual emotion classification. *Affective Computing, IEEE Transactions on*, 3(2):184–198, 2012.
- [55] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *Image Processing, IEEE Transactions on*, 21(12):4695–4708, 2012.
- [56] Anish Mittal, Ravi Soundararajan, and Alan C Bovik. Making a completely blind image quality analyzer. *Signal Processing Letters, IEEE*, 20(3):209–212, 2013.
- [57] Gelareh Mohammadi and Alessandro Vinciarelli. Humans as feature extractors: combining prosody and personality perception for improved speaking style recognition. In *SMC*. IEEE, 2011.
- [58] Gelareh Mohammadi and Alessandro Vinciarelli. Automatic personality perception: Prediction of trait attribution based on prosodic features. *Affective Computing, IEEE Transactions on*, 3(3):273–284, 2012.
- [59] Meinard Müller and Sebastian Ewert. Chroma toolbox: Matlab implementations for extracting variants of chroma-based audio features. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR), 2011*. hal-00727791, version 2-22 Oct 2012. Citeseer, 2011.
- [60] Rajitha Navarathna, Patrick Lucey, Peter Carr, Elizabeth Carter, Sridha Sridharan, and Iain Matthews. Predicting movie ratings from audience behaviors. In *WACV*. IEEE, 2014.
- [61] Francis T O’Donovan, Connie Fournelle, Steve Gaffigan, Oliver Brdiczka, Jianqiang Shen, Juan Liu, and Kendra E Moore. Characterizing user behavior and information propagation on a social multimedia network. In *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, pages 1–6. IEEE, 2013.
- [62] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.
- [63] Luiz Pessoa. On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, 2008.
- [64] Robert Plutchik. *Emotion: A psychoevolutionary synthesis*. Harper & Row New York, 1980.
- [65] Timo Reuter, Symeon Papadopoulos, Giorgos Petkos, Vasileios Mezaris, Yiannis Kompatsiaris, Philipp Cimiano, Christopher de Vries, and Shlomo Geva. Social event detection at mediaeval 2013: Challenges, datasets, and evaluation. In *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop Barcelona, Spain, October 18-19, 2013*, 2013.
- [66] Dairazalia Sanchez-Cortes, Joan-Isaac Biel, Shiro Kumano, Junji Yamato, Kazuhiro Otsuka, and Daniel Gatica-Perez. Inferring mood in ubiquitous conversational video. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, page 22. ACM, 2013.
- [67] Alexandre Schaefer, Frédéric Nils, Xavier Sanchez, and Pierre Philippon. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, 24(7):1153–1172, 2010.
- [68] Björn Schuller, Michel Valstar, Florian Eyben, Gary McKeown, Roddy Cowie, and Maja Pantic. Avec 2011—the first international audio/visual emotion challenge. In *Affective Computing and Intelligent Interaction*. Springer, 2011.
- [69] Seyedehsamaneh Shojaeilangari, Wei-Yun Yau, Karthik Nandakumar, Jun Li, and Eam Khwang Teoh. Robust representation and recognition of facial emotions using extreme sparse learning. *Image Processing, IEEE Transactions on*, 24(7):2140–2152, 2015.
- [70] Muhammad Hameed Siddiqi, Rahman Ali, Adil Mehmood Khan, Young-Tack Park, and Sungyoung Lee. Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. *Image Processing, IEEE Transactions on*, 24(4):1386–1398, 2015.
- [71] Behjat Siddiquie, Dave Chisholm, and Ajay Divakaran. Exploiting multimodal affect and semantics to identify politically persuasive web videos. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 203–210. ACM, 2015.
- [72] Jacquelyn Smith. The most unforgettable ad campaigns of 2013, December 2013.
- [73] Mohammad Soleymani, Joep JM Kierkels, Guillaume Chanel, and Thierry Pun. A bayesian framework for video affective representation. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–7. IEEE, 2009.
- [74] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *Affective Computing, IEEE Transactions on*, 3(1):42–55, 2012.
- [75] Maria Teresa Soto Sanfiel, Laura Aymerich Franch, F Xavier Ribes Guardia, and J Reinaldo Martínez Fernandez. Influence of interactivity on emotions and enjoyment during consumption of audiovisual fictions. *International Journal of Arts and Technology*, 2011.
- [76] Kai Sun and Junqing Yu. Video affective content representation and recognition using video affective tree and hidden markov models. In *Affective Computing and Intelligent Interaction*, pages 594–605. Springer, 2007.
- [77] Jussi Tarvainen, Mats Sjöberg, Stina Westman, Jorma Laaksonen, and Pirkko Oittinen. Content-based prediction of movie style, aesthetics and affect: Data set and baseline experiments. 2014.
- [78] René Marcelino A Britta Teixeira, Toshihiko Yamasaki, and Kiyoharu Aizawa. Determination of emotional content of video clips by low-level audiovisual features. *Multimedia Tools and Applications*, 61(1):21–49, 2012.

- [79] Michel Valstar, Björn Schuller, Kirsty Smith, Florian Eyben, Bihan Jiang, Sanjay Bilakhia, Sebastian Schnieder, Roddy Cowie, and Maja Pantic. Avec 2013: the continuous audio/visual emotion and depression recognition challenge. In *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, pages 3–10. ACM, 2013.
- [80] Alessandro Vinciarelli and Gelareh Mohammadi. A survey of personality computing. 2014.
- [81] Valentijn T Visch, Ed S Tan, and Dylan Molenaar. The emotional and cognitive effect of immersion in film viewing. *Cognition and Emotion*, 24(8):1439–1445, 2010.
- [82] Julia Wache. The secret language of our body: Affect and personality recognition using physiological signals. In *ICMI*. ACM, 2014.
- [83] Shangfei Wang and Qiang Ji. Video affective content analysis: a survey of state of the art methods. *Affective Computing, IEEE Transactions on*.
- [84] Su-Jing Wang, Wen-Jing Yan, Xiaobai Li, Guoying Zhao, Chun-Guang Zhou, Xiaolan Fu, Minghao Yang, and Jianhua Tao. Micro-expression recognition using color spaces. *Image Processing, IEEE Transactions on*, 24(12):6034–6047, 2015.
- [85] Cheng-Yu Wei, Nevenka Dimitrova, and Shih-Fu Chang. Color-mood analysis of films based on syntactic and psychological models. In *ICME*. IEEE, 2004.
- [86] Kathy Winter and Nicholas Kuiper. Individual differences in the experience of emotions. *Clinical psychology review*, 1997.
- [87] Wanmin Wu, Ahsan Arefin, Raoul Rivas, Klara Nahrstedt, Renata Sheppard, and Zhenyu Yang. Quality of experience in distributed interactive multimedia environments: toward a theoretical framework. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 481–490. ACM, 2009.
- [88] Jun Yang, Yu-Gang Jiang, Alexander G Hauptmann, and Chong-Wah Ngo. Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pages 197–206. ACM, 2007.
- [89] Nick Yeung and Alan G Sanfey. Independent coding of reward magnitude and valence in the human brain. *The Journal of Neuroscience*, 2004.
- [90] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE TPAMI*, 2009.
- [91] Shiliang Zhang, Qingming Huang, Shuqiang Jiang, Wen Gao, and Qi Tian. Affective visualization and retrieval for music video. *IEEE Transactions on Multimedia*, 12(6):510–522, 2010.
- [92] Shiliang Zhang, Qi Tian, Qingming Huang, Wen Gao, and Shipeng Li. Utilizing affective analysis for efficient movie browsing. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1853–1856. IEEE, 2009.
- [93] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. In *Advances in Neural Information Processing Systems*, pages 487–495, 2014.