Facial Recognition and Emotion Detection System for Dynamic Advertisement Allocation

Frank Yeong-Sung Lin[1], Evana Szu-Han Fang[1], Chiu-Han Hsiao[2]
Department of Information Management, National Taiwan University[1]
Research Center for Information Technology Innovation, Academia Sinica[2]
Taipei, Taiwan
email: yeongsunglin@gmail.com, evanafang@gmail.com, chiuhanhsiao@citi.sinica.edu.tw

***Abstract*—Advertisements represent a persuasive method of communication for convincing people to change their thoughts or attitudes. Conventional advertisements do not always provide optimal marketing effectiveness because the advertisements are presented uniformly to viewers. To overcome the limitations of traditional methods in advertising research, a dynamic advertisement model is proposed in this paper, and facial expression detection is applied to real-time measurement during media exposure. This is a novel model to recognize viewers' facial expression for emotion regulation and then adjust the decision of content sequence according to their emotions. A decision tree algorithm is used, and each demographic measurement results from a few scenarios. The decision is determined through bottom-up branch searching. Based on the study results, personalized advertising and audience targeting with accurate facial expression analysis can allow marketing and advertising researchers to better understand viewers' emotional valence and behavior and to employ mathematical formulation for establishing the optimal advertising approach.**

Keywords-dynamic advertisement; facial recognition; emotion detection; audience targeting; decision tree.

I. Introduction

Advertising is a persuasive method of communication for convincing people to change their thoughts or attitudes [1]. This is a type of brand-related stimulus that conveys brand experiences, consisting of subjective and internal customer responses [2]. With time, enduring memories of brand experiences in customers' minds affect customer satisfaction and loyalty [3][4]. Therefore, advertising and marketing companies actively seek an optimal instrument that can recognize true feelings from customers, and they cannot hide their thoughts [5]. Enhancing understanding, evaluation, and advertising effectiveness is of great value both in theory and in practice.

People experience certain emotional responses when viewing an advertisement, which may be positive or negative. This greatly affects the sales of a product, reduces price sensitivity, and creates brand value [6]. Hence, viewers' emotions can be used to predict an advertisement's effectiveness [7]. As a strong connection exists between emotions and facial expression, researchers have been interested in developing methodologies to effectively measure the facial expression and emotions experienced [5][8]-[12]. The facial expression is the clearest method of establishing a person's affective state [13]. Research has indicated that variables correlated to advertising success such as advertisement likability [14], recall [15], and "zapping" [12] can be predicted by facial expressions.

Exposure to an advertising stimulus evokes emotions among people; their attention is subsequently affected, and zapping is also affected by emotion and attention; the extent of these effects varies during exposure to an advertisement [12]. Emotion regulation is a dynamic process. Thus, traditional uniform content advertised to audiences cannot achieve optimal marketing effectiveness, because of different preferences of individuals and their emotions. Brands may squander the opportunity to communicate when targeted customers zap, skip, and zip advertisements [12]. To retain viewers and maximize marketing effectiveness, a novel real-time content adjustment model based on emotion detection is proposed.

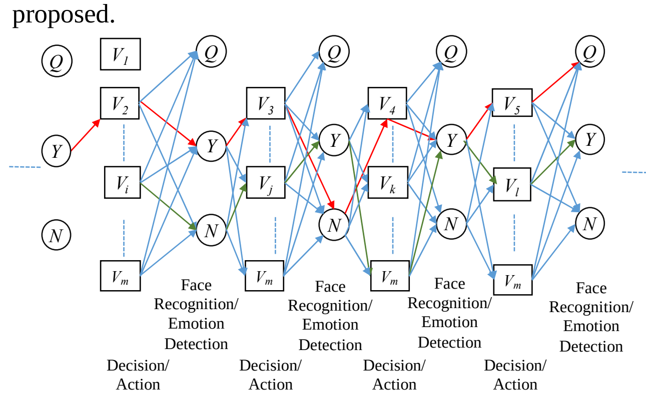

Figure 1. Decision tree for facial recognition and emotion detection.

Decision trees constitute a nonparametric supervised learning method used for classification and regression [16]. They predict the value of a target variable by learning simple decision rules inferred from data features. In this paper, the decision tree learns from data related to facial recognition and emotion detection, as seen in Figure 1, to approximate a set of if-then-else decision rules. If a given situation is observable in a model, the condition is easily explained by Boolean logic. Understanding video clip interpretations is simple, and the selected decision rules are determined completely if viewing promotes positive emotions. In practice, a preferred classification model can be verified through statistical tests and adjusted within a set time interval for real-world applications.

The remainder of this paper is structured as follows. Section II reviews relevant studies on the linkage between facial expression detection, emotion detection, and advertisement content. The bottom-up decision-making method is proposed in Section III. Section IV details the process flow and cases of identifying decisions and rules in practical applications. Finally, conclusions are drawn in Section VI.

II. RELATED WORK

Traditional research methods, such as self-report provide limited understanding of the linkage between emotion and advertisement content as well as of how advertisement effectiveness is measured [17]. Though self-report is cheap, fast, and valid, it cannot capture low-order emotion or the temporal nature of emotion regulation during an advertisement; it may also increase cognitive bias [5][16]. People communicate valence and emotional states powerfully through their faces [18]. Therefore, marketing and advertising researchers can better understand viewers' emotions and behaviors through facial expression analysis and can accordingly establish strategies to improve advertisement effectiveness. Moreover, they may design interactive advertisements to improve viewer experiences [19]. Research has used automatically measured facial expressions to predict emotional valence and advertisement preference during media exposure [14][20][21]. Thus, the use of automated tools augments the feasibility of the approach and exhibits higher predictive capability than self-report does [5][7]. Many studies have initially investigated feigned or acted facial behaviors [16]. Nevertheless, research has progressively emphasized naturalistic and spontaneous behavior [22]-[24] and subtle expressions [25].

Automated facial expression detection combines the fields of psychology, computer vision, and machine learning [18]. In [26], the authors proposed a new nonlinear tensor factorization based on deep variational autoencoders called Factorized Variational Auto-Encoders (FVAE) for modeling movie audiences' facial expressions. The effectiveness of FVAE was determined for a large facial expression dataset extracted from 3179 movie audience members. Even when using only 5% of data for initial observation, FVAE could reconstruct facial reactions more precisely using data from movie audiences than traditional baseline applying entire data. One study [19] focused on predicting user behavior and viewing experiences based on facial expressions during online advertisement viewing. A metric termed Moment-to-Moment Zapping Probability (MMZP) was used to predict user skipping; the preference information extracted from users may be used to enhance advertisement effectiveness. In addition, the authors categorized smiling as a primary facial expression during analysis. Because amusement is a desirable response that advertisers strive to elicit, the entertainment level of an advertisement directly relates to smiling. Sparse reconstruction coefficients were used as features for classifying smiling to make MMZP predictions. The authors in [18] collected spontaneous facial expressions from viewers during the 2012 US presidential debates to predict voter preferences and found an average precision of over 73%. The Facial Action Coding System [21][27][28] was implemented to measure and score facial activity reliably and to distinguish subtle differences in facial expressions [29]. In [30], online video advertisements viewed by Japanese people were analyzed for physiological responses involving facial expressions, heart rate signals, and gaze. The authors integrated each mode's features and evaluated advertisement likability and purchase intent. In [31], the authors proposed an interactive advertisement system with 3D tracking and facial recognition to produce audience profile surveys. They reviewed the conclusions of Lord and Burnkrant [32] regarding the increased attention levels involved when viewers were immersed in highly interactive programs. The psychological sensation of presence could explain this cognitive state. Moreover, [33] indicated that this experience of presence would affect product knowledge, brand attitude, and the purchasing intentions of consumers.

III. PROPOSED METHOD AND PROCESS FLOW

The model depicted in Figure 2 is proposed to maximize viewer experiences during advertisements. A decision tree algorithm of content personalization is employed to implement decisions into fields to achieve objectives. Reward measurement and observation uses facial recognition and emotion detection techniques.

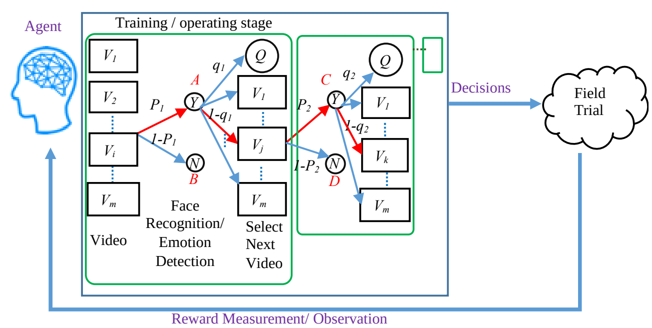

Figure 2. Abstract system architecture.

In Figure 3, the process flow of an advertisement involves several video clips. The decision tree algorithm can be applied to a dynamic scenario mechanism for advertisements based on recognition of viewers' facial expressions. Training and testing involves two stages: probability and value evaluation are determined by facial recognition and emotion detection processes. The bottom-up algorithm is then applied to select video clips forming a complete advertisement tree structure with advertisement video clips. In Figure 3, the constructed path selected in a tree structure represents the sequence of the advertisement video clips for a type of demographic classification. To optimize the effects of the advertisements, a clip is selected on the basis of the viewer's emotion detection results after the end of the preceding clip. The algorithm finally classifies viewers according to the probability and value evaluation determined by facial recognition and emotion detection in the bottom-up backward process.

 ISBN: 978-1-61208-728-3

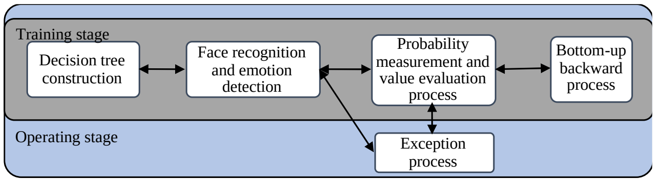

Figure 3. Process flow of proposed stages.

A. Decision Tree Construction

The distributions of viewers' demographic characteristics (e.g., age, sex, etc.) are applied to the construction of the optimal trees. Once a viewer is categorized through facial recognition analysis, the tree path of that category is extracted for implementation in the model. Probability and reward values are acquired through measurement.

B. Facial Recognition and Emotion Detection for Probability Measurement and Value Evaluation

The parameters P_i represent the probabilities of the next selected video clips. Before finishing the construct of the optimal trees, we have a training and experiment phases with a number of viewers to measure each probability shown in Figure 5. The experiment is implemented and all the viewers see clip V_1 firstly, then some of them may have positive emotion (the node *AY*), the others may have negative emotion (the node *BN*). P_1 and P_2 are measured in this step. After that, assuming the experiment goes to *AY*, half of *AY*'s viewers are aired V_2; half of them are aired V_3 at random. The measured P_3, P_4, P_5 and P_6 correspond to *CY*, *DN*, *EY* and *FN*. We implement the procedures given above to measure the probabilities if the experiment goes to other branches.

A, B, C, D, E,…, which represent marginal effects with facial recognition to expand all possibilities before the optimal decision tree is not coming out yet in the training stage. The model records every viewing experience and establishes the best advertisement editing and composition strategy for each demographic group of viewers in the operating stage. A 5- point or 7-point Likert scale can be used to measure viewers' perceptions and purchase intentions after they view an advertisement. The incentive is provided in accordance with the scale.

C. Bottom-up Decision-Making

Binning or discretization is the process of transforming numerical variables into categorical counterparts [34]. Numerical variables are usually discretized in modeling methods based on decision trees, such as *A*, *B*, *C*, *D*, *E*, … representing rewards for positive or negative emotions detected in Figure 5.

In bottom-up decision-making, the reverse approach is applied to top-down decision-making. To ensure that bottom-up decision-making is effective, emotion detection information is used in the predictive model, and outcomes are accordingly predicted. Descriptive modeling is the assignment of observations into decision trees. The rules employed permit associations among observations, and they are based on the entropy using the frequency table of two attributes, which are the expected rewards of emotion detection. The equation used is $E(K,N)=\sum_{k\in K}\sum_{n\in N}P(V_k)E(n)$. $E(n)$ is the emotion detection result. $P(V_k)$ is the probability of selecting the video V_k. For example, the value of (P_3C+P_4D) is the expected reward related to the decision V_2. The value of (P_5E+P_6F) is related to the decision V_3. Entropy and decisions are constructed bottom-up to form the decision tree depicted in Figure 6. The pseudocode is presented in Figure 4, as follows:

Training data input in experiment
Generation of Tree (Decisions *K*, Emotion Detection *N*)
If stopping_condition(*K*,*N*) = true then
leaf = createNode()
leaf.label= Classify(*K*)
return leaf
root = createNode()
root.test_condition = findBestSplit(*K*,*N*)
$E(K,N)=\sum_{k\in K}\sum_{n\in N}P(V_k)E(n)$: list possible outcome of root.test_condition
for each value E for branches
Select the maximum E related to decisions *K*;
Build child = TreeGrowth(*K*, *N*) ;
Add child as a descent of root and label the edge
return root

Figure 4. Pseudocode of the decision tree algorithm.

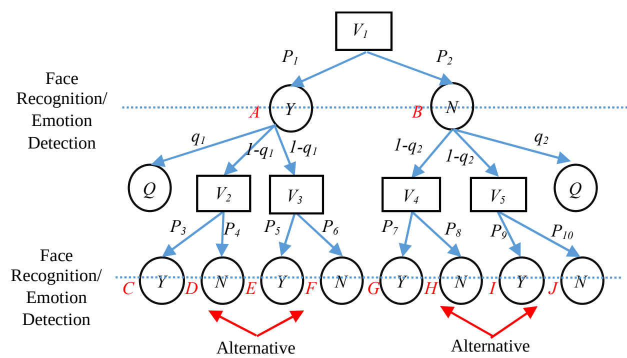

Figure 5. Bottom-up alternative selection.

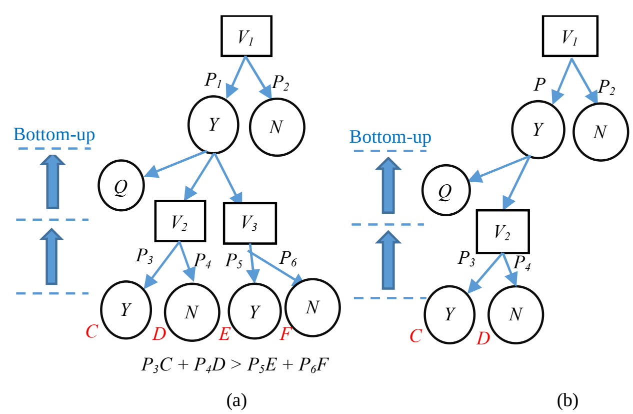

Figure 6. Bottom-up mechanism for the expected value calculation.

To apply the bottom-up mechanism for calculation, the proper video clip is assigned and inserted to continue the sequence of clips by detecting emotions at any level from facial recognition results. For example, after watching V_1, facial expression recognition indicates that the viewer experiences a positive emotion, so *AY* is subsequently selected, and the model may choose V_2 or V_3 to continue the advertisement: $P_3C + P_4D$ for V_2 or $P_5E + P_6F$ for V_3, depending on which one has the higher expected value for effect. If $P_3C + P_4D$ for V_2 is more favorable, then the branch $P_5E + P_6F$ is deleted.

D. Exception Process

In the case of a negative emotion for V_1, *BN* is subsequently selected, as depicted in Figure 7, and the aforementioned method can be applied to select V_4 or V_5. If more video clips are available, the bottom-up method can be used to select from the bottom to the top of the model.

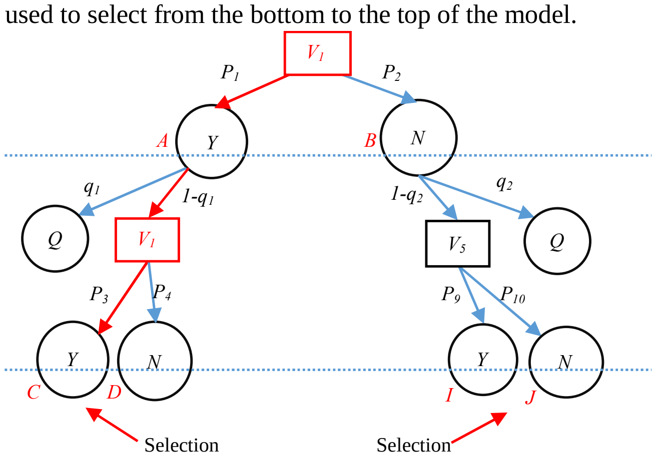

Figure 7. Branch selection for decision-making.

After running the model for a period, as in Figure 8, if the on-record $P_3C + P_4D$ for V_2's expected value is lower than the value of $P_5E + P_6F$, V_2 is switched to V_3. The top branch of V_3 is adjusted accordingly. Continuing to measure the bottom expected value under V_3, if it is lower than the bottom expected value under V_2, then the selection is again switched from V_3 to V_2, and the top branch of V_2 consistently.

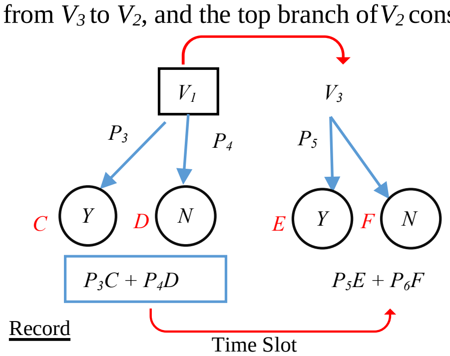

Figure 8. Recording emotion detection values in a period.

A scenario is dynamically created according to the viewers. The structure for selecting proper video clips to suit the advertisement scenario is as follows. As seen in Figure 9, the viewer watches the first clip (V_1), and when it ends, the next clip is selected based on the viewer's emotion detection results.

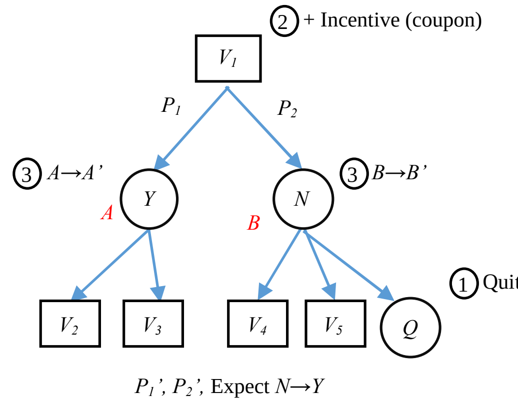

Figure 9. Incentivizing viewers to continue watching videos.

This model enables the timely identification of the user's emotion from the last clip. For example, in Figure 9, if the viewer wants to stop watching additional clips, an incentive can be provided to entice the viewer to continue watching.

IV. Conclusion & Future Work

Although Information Communication Technology is developing rapidly, it is applied in management relatively infrequently, and it has yet to be used to fully establish independent technology. Traditional methods, such as surveys, provide limited understanding of the linkage between emotion and advertisement content as well as of how to measure advertisement effectiveness. Therefore, this paper proposes a dynamic scenario mechanism for advertising based on a decision tree algorithm and on viewers' facial expressions recognized during viewing. Dynamic content changes result from viewer facial expression recognition. The content customization guides the viewer's concentration [35]. Subsequently, this work explores the research topics involved in technology and management issues, and to obtain the results for applying theory to practice, as a suitable reference for the video clip strategies used to the operator or related industry company in optimal advertisement display and well predictive analysis in the future. The future directions are summarized as follows:

- Level of emotion detection:

Emotional responses can be defined and distributed into several categories, such as positive versus negative emotions. They may even be classified according to three types of emotions: happiness or approval, neutral, and disapproval. However, happiness may be further subdivided into extreme delight, surprised excitement, or tears of joy. Due to the complexity of emotions, it can be posited that emotions can be distributed into two types: positive and negative.

- Multiple viewers:

When the camera detects more than one viewer (e.g., three people comprising two male and one female), these people can be distributed into two categories by demographic characteristic. The decisions are determined by the two trees in accordance with these two categories. Then, both trees are extracted, overlapped, and superposition or weighted sum of the probabilities for each branch to form a new tree.

- Level of emotion detection for multiple viewers:

When only one optimal tree exists for a group of viewers, people may leave or start watching halfway through an advertisement; therefore, future research can uncover solutions regarding how this optimal tree can be altered dynamically. Such a study will entail complications, but is definitely worthwhile for developing more customized advertisements for optimizing the viewing experience

REFERENCES

[1] J. Meyers-Levy and P. Malaviya, “Consumers’ processing of persuasive advertisements: An integrative framework of persuasion theories,” Journal of Marketing, vol. 63, no. 4, Oct. 1999, pp. 45–60.

[2] J. Brakus, B. H. Schmitt, and L. Zarantonello, “Brand experience: What is it? How is it measured? Does it affect loyalty?” Journal of marketing, vol. 73, no. 3, May, 2009, pp. 52–68.

[3] R. L. Oliver, Satisfaction: A Behavioral Perspective on the Consumer, Boston, NY: McGraw-Hill, 1997.

[4] F. F. Reichheld, The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value. Boston, NY: Harvard Business School Press, 1996.

[5] P. Lewinski, M. L. Fransen, and E. S. H. Tan, “Predicting advertising effectiveness by facial expressions in response to amusing persuasive stimuli,” Journal of Neuroscience Psychology and Economics, vol. 7, no. 1, Mar. 2014, pp. 1–14.

[6] P. Saraswat, H. Nagar, and S. Khandelwal, “Make it feel: Use of facial imaging technique to analyze the impact of each emotional spot on ad success,” Proc. The 9th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS 2015), IEEE, Jul. 2015, pp. 502–507.

[7] K. Poels and S. Dewitte, “How to capture the heart? Reviewing 20 years of emotion measurement in advertising,” Journal of Advertising Research, vol. 46, no. 1, Nov. 2006, pp. 18– 37.

[8] P. M. Cole, “Children’s spontaneous control of facial expression,” Child Development, vol. 57, no. 6, Dec. 1986, pp. 1309–1321.

[9] J. J. Gross and R. A. Thompson, “Emotion regulation: Conceptual foundations,” Handbook of Emotion Regulation, pp. 3–26. New York, NY: Guilford Press, 2007.

[10] C. E. Izard, “Facial expressions and the regulation of emotions,” Journal of Personality and Social Psychology, vol. 58, no. 3, Mar. 1990, pp. 487–498.

[11] Darwin and Charles, The Expression of the Emotions in Man and Animals. London: Murray, 1872.

[12] T. Teixeira, M. Wedel, and R. Pieters, “Emotion-induced engagement in Internet video advertisements,” Journal of Marketing Research, vol. 49, no. 2, Apr. 2012, pp. 144–159.

[13] H. P. Mal and P. Swarnalatha, “Facial expression detection using facial expression model,” Proc. International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS 2017), IEEE, Aug. 2017, pp. 1259–1262.

[14] D. McDuff, R. Kaliouby, T. Senechal, D. Demirdjian, and R. W. Picard, “Automatic measurement of ad preferences from facial responses gathered over the internet,” Image and Vision Computing, vol. 32, no. 10, Oct. 2014, pp. 630–640.

[15] R. L. Hazlett and S. Y. Hazlett, “Emotional response to television commercials: Facial EMG vs. Self-report,” Journal of Advertising Research, vol. 39, no. 2, Mar. 1999, pp. 7–24.

[16] A. Suarez and J. F. Lutsko, “Globally optimal fuzzy decision trees for classification and regression,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 12, Dec. 1999, pp. 1297–1311.

[17] D. McDuff, R. Kaliouby, J. F. Cohn, and R. W. Picard, “Predicting ad liking and purchase intent: Large-scale analysis of facial responses to ads,” IEEE Transactions on Affective Computing, vol. 6, no. 3. Jul.-Sep. 2015, pp. 223–235.

[18] D. McDuff, R. Kaliouby, E. Kodra, and R. W. Picard, “Measuring voter's candidate preference based on affective responses to election debates,” Proc. Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII 2013), IEEE, Sep. 2013, pp. 369–374.

[19] S. Yang and L. An, “Analyzing user behavior in online advertising with facial expressions,” Proc. 23rd International Conference on Pattern Recognition (ICPR 2016), IEEE, Dec. 2016, pp. 4238–4243.

[20] D. McDuff, R. Kaliouby, K. Kassam, and R. W. Picard, “Affect valence inference from facial action unit spectrograms,” Proc. Computer Society Conference on Computer Vision and Pattern Recognition – Workshops (CVPRW 2010), IEEE, Jun. 2010, pp. 17–24.

[21] E. Kodra, T. Senechal, D. McDuff, and R. el Kaliouby, “From dials to facial coding: Automated detection of spontaneous facial expressions for media research,” Proc. 10th International Conference and Workshops on Automatic Face and Gesture Recognition (FG 2013), IEEE, Apr. 2013, pp. 1–6.

[22] D. McDuff, R. Kaliouby, and R. W. Picard, “Crowdsourcing facial responses to online videos,” IEEE Transactions on Affective Computing, vol. 3, no. 4, Oct.-Dec. 2012, pp. 456–468.

[23] M. Pantic, “Machine analysis of facial behaviour: Naturalistic and dynamic behaviour,” Philosophical Transactions of the Royal Society of London B: Biological Sciences, vol. 364, no. 1535, Dec. 2009, pp. 3505–3513.

[24] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, “Toward practical smile detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 11, Nov. 2009, pp. 2106–2111.

[25] T. Senechal, J. Turcot, and R. Kaliouby, “Smile or smirk automatic detection of spontaneous asymmetric smiles to understand viewer experience,” Proc. 10th International Conference and Workshops on Automatic Face and Gesture Recognition (FG 2013), IEEE, Apr. 2013, pp. 1–8.

[26] Z. Deng et al. “Factorized variational autoencoders for modeling audience reactions to movies,” Proc. Conference on Computer Vision and Pattern Recognition (CVPR 2017), IEEE, Jul. 2017, pp. 6014–6023.

[27] P. Ekman and W. V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, CA: Consulting Psychologists Press, 1978.

[28] J. F. Cohn, Z. Ambadar, and P. Ekman, “Observer-based measurement of facial expression with the Facial Action Coding System,” The Handbook of Emotion Elicitation and Assessment. Oxford University Press Series in Affective Science, pp. 203–221. New York, NY: Oxford University Press, 2007.

[29] C. Hjortsjö, Man’s face and mimic language. Lund, Sweden : Studentlitteratur, 1969.

[30] G. Okada, K. Masui, and N. Tsumura, “Advertisement effectiveness estimation based on crowdsourced multimodal affective responses,” Proc. Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2018), IEEE/CVF, Jun. 2018, pp. 1344–1348.

[31] M. Taspinar, A. T. Naskali, G. Eren, and M. Kurt, “The importance of customized advertisement delivery using 3D tracking and facial recognition,” Proc. 2nd International Conference on Digital Information and Communication Technology and it's Applications (DICTAP 2012), IEEE, May, pp. 520–524.

[32] K. Lord and R. Burnkrant, “Attention versus distraction: the interactive effect of program involvement and attentional devices on commercial processing,” Journal of Advertising, vol. 22, no. 1, Mar. 1993, pp. 47–60.

[33] H. Li, T. Daugherty, and F. Biocca, “Impact of 3-D advertising on product knowledge, brand attitude, and purchase intention: The mediating role of presence,” Journal of Advertising, vol. 31, no. 3, May, 2002, pp. 43–57.

[34] D. Bacciu, A. Micheli, and A. Sperduti, “Compositional generative mapping for tree-structured data—Part I: Bottom-up probabilistic modeling of trees,” IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 12, Dec. 2012, pp. 1987–2002.

[35] K. Lee, S. Rho, and E. Hwang, “Scenario based dynamic content management system for e-Learning environment,” Proc, 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA 2016), IEEE, Mar. 2016, pp. 183–186.