A Pilot Study on Affective Classification of Facial Images for Emerging News Topics

Ligang Zhang^{#1}, Andy Cher Han Lau^{*2}, Dian Tjondronegoro^{*3}, Vinod Chandran^{*4}

[#] Faculty of Computer Science and Engineering, Xi'an University of Technology
5 South Jinhua Road, Xi'an, Shaanxi, 710048 China

¹ lg.zhang@xaut.edu.cn

^{*} Science and Engineering Faculty, Queensland University of Technology
2 George St, Brisbane, QLD 4000 Australia

²⁻⁴ {andy.lau, dian, v.chandran}@qut.edu.au

Abstract—The proliferation of news reports published in online websites and news information sharing among social media users necessitates effective techniques for analysing the image, text and video data related to news topics. This paper presents the first study to classify affective facial images on emerging news topics. The proposed system dynamically monitors and selects the current hot (of great interest) news topics with strong affective interestingness using textual keywords in news articles and social media discussions. Images from the selected hot topics are extracted and classified into three categorized emotions, positive, neutral and negative, based on facial expressions of subjects in the images. Performance evaluations on two facial image datasets collected from real-world resources demonstrate the applicability and effectiveness of the proposed system in affective classification of facial images in news reports. Facial expression shows high consistency with the affective textual content in news reports for positive emotion, while only low correlation has been observed for neutral and negative. The system can be directly used for applications, such as assisting editors in choosing photos with a proper affective semantic for a certain topic during news report preparation.

I. INTRODUCTION

The advancement of information technology has dramatically promoted the proliferation of news reports in online websites. Most international news corporations have their own websites with a huge volume of report archives represented in text, image, audio and video formats. The content of the website is also updated frequently according to the news topics that occur at the moment, international or domestic. At the same time, with the increasing popularity of social media such as Twitter and the growing popularization of mobile devices such as smart phone, a large amount of live discussions related to any particular interesting and hot news topic are recorded and shared between social media users every day. Many Internet users, especially youngsters, prefer to publish their ideas and share their feelings and emotions towards news topics via social media.

Given a large amount of news data, it would be very useful for both news editors and readers to be informed of the hot topics happening at the current time, along with accompanying affective content such as expressions of the

face in relevant reports. Automatic affective content analysis in images can provide valuable insights about the emotional attitudes of the writer, and the entities involved (e.g. politicians and celebrities) with respect to the topic portrayed in the electronic media, such as Barack Obama's attitude towards the Malaysia Airlines MH370 Incident and Vladimir Putin's feeling about the results of the 2014 Winter Olympics in Sochi. It can also help summarize the key topics and content that are most likely to evoke similar emotion responses from the readers and/or are associated with human emotions from different resources, including news website, films, and social media etc.

Existing approaches for affective content analysis focus on extracting low-level features (e.g. lighting, motion, and colour) derived from the theories of psychology, art theory, colour theory, aesthetics, cinematography, etc., or constructing mid-level representations (e.g. dialog, fight, and spatial distribution of edges and colour harmony) from low-level features. These features are then mapped into emotion categories or dimensional spaces. These approaches are often criticized for the incapability to bridge the affective gap [1] between low-level features and high-level human perception.

It is desirable to adopt high-level features in multimedia content that have closer link with user affective perception. The facial expression of subjects is a major way of expressing key ideas and primary moods in multimedia content, and thus directly impacts user affective perception. Intuitively, news editors might prefer to deliberately select facial images with a certain expression to reflect the affective content in the article. However, facial expressions have not been fully exploited for affective classification of practical data, such as news images.

This paper proposes a novel system that combines text and facial expression information for affective classification of facial images on emerging news topics. The text module dynamically monitors the current emerging news topics based on textual keywords in news articles and social media discussions, while the facial expression module classifies facial images relevant to these topics into three categorized emotion classes. To the best of our knowledge, it is the first system that adopts online live textual information from social media and news websites and high-level facial expressions for affective classification of news images.

The rest of the paper is organized as follows. Section II reviews related work. Section III introduces the proposed system. Section IV reports the experiments on two realistic facial image datasets. Conclusions are drawn in Section V.

II. RELATED WORK

The dominant existing approaches for affective content analysis on multimedia data are based on low-level features (e.g. lighting, motion, and colour) extracted from audio, visual or textual modalities. These features are further mapped into discrete emotion categories or continuous dimensional spaces, such as arousal-valence (A-V). The mapping is achieved by building up a group of decision rules based on the theoretical and empirical knowledge learnt from psychology, art theory, colour theory, aesthetics, and cinematography etc. [1], [2] or constructing machine learning models by employing a train-test process using algorithms, such as neural network [3], support vector machine (SVM), and hidden Markov models. One of the earliest affective image classification systems was K-DIME [3], which mapped low-level features in images to impression words using a neural network. Jana and Allan [2] classified artistic photos and abstract paintings into eight emotional categories based on a set of colour, texture, composition and face features. For affect analysis on video, one of the representative works was conducted by Hanjalic and Qun [4], who modelled arousal and valence intensities linearly and separately using two individual sets of audio-visual features extracted from video. Niu et al. [5] modelled arousal and valence values using two audio-visual feature sets and computed video similarities in the A-V space for personalized video recommendation. Canini et al. [6] extracted audio, visual and film grammar features and mapped them into a connotative space for affective movie recommendation. However, these approaches were generally criticized for their inability to bridge the affective gap [1], which is caused by the lack of coincidence between the measurable signal properties (i.e. features) and the expected affective perception of the user, which is often subjective and time-varying [7].

To reduce the affective gap, recent studies proposed to construct mid-level feature representations from low-level features that are expected to have a closer link with reviewer affective perception. Acar et al. [8] built mid-level representations from Mel-Frequency Cepstral Coefficients and colour values using convolutional neural networks, revealing an improved performance on affective classification of video clips. Liu et al. [9] used the spatial distribution of edges and colour harmony, together with a set of low-level features, for affective classification of images. Ionescu et al. [10] predicted mid-level concepts (blood, firearms, fights etc.) for violence detection in Hollywood films. Xu et al. [11] constructed mid-level features to indicate dialog, audio emotional events and textual concepts. Although these representations have shown promising performance to infer high-level affective content, but they still cannot fully reflect user affective understanding of the content.

It is advisable to adopt high-level features in multimedia content, which have direct impact on user affective perception, such as the face and facial expression of subjects. In real scenarios, the face in an image often strongly draws the attention of human observers and has shown as a key indication of the affect in the image [2]. The facial expression of the face is also a major way of expressing key ideas and primary moods in multimedia content, and thus it is most likely to evoke similar affective responses from the reviewer. A face with a fear emotion often imposes similar affect on the reviewer.

In the field of news report publication, by our intuition, news editors often deliberately select facial images with a certain expression to reflect the affective content in the report or express their personal affective preferences or ideas about the topic. It has been proved that compelling negative news images have strong effect on human memory for information in the news stories [12]. However, existing works [13], [14] using news data focus primarily on extracting audio-visual or textual features such as subtitles for annotating semantic concepts (e.g. event and subject), which are not directly linked with high-level human affect perception. There are also studies [15], [16] that investigated affective classification of news articles, but they are based on only textual information. No previous study has been found on using facial expressions for affective classification of facial images related to certain emerging news topics.

III. SYSTEM FRAMEWORK

Fig. 1. Framework of the proposed system

Fig.1 shows the proposed system for affective classification of facial images on emerging news topics comprising of a text

module and a facial expression module. The text module extracts and combines textual keywords about certain topics published in news websites and social media (i.e. Twitter), and then adopts natural language processing methods to rank the affective interestingness of topics. Based on the top ranked topics, the facial expression module focuses on analysing the affective states of people appearing in news images. It classifies each image with human faces into three emotions of positive, negative and neutral using a fusion of texture and geometry features extracted from the detected facial region. In the outputs of the system, all news topics are ranked in a list from the hottest to the least interesting. If there exist images with facial expressions appropriate for certain topics, they are also provided along with the corresponding topics as shown in the right part of Fig.1.

A. Text Module

The text module aims to identify the emerging news topics and it evaluates the interestingness of news topics by analysing text extracted and crawled from news websites and social media using natural language processing and sentiment analysis techniques. The news websites used for data collection include a large amount of main Australian news publishing channels, including smh.com.au, watoday.com.au, businessday.com.au, canberratimes.com.au etc., covering various international and national news topics ranging from business, technology, entertainment, sports, environment to travel and lifestyle etc. The information for social media mainly comes from Twitter.

If a topic contains many interesting elements, i.e. text documents with strong emotion keywords, that topic is expected to be of high interest. The focus on extracting emotion keywords is to provide an indication of the affective content relating to a certain topic, from the views of either the writers or the public. From a large collection of articles relating to a news topic, the interest level of each article can be revealed from the emotion words used within. For instance, in sad events, people would more likely use terms such as “sad”, “unfortunate”, or “devastating”, whereas in happy events, terms such as “happy”, “glory” and “success” will more likely be used. An affective score between -1 to 1 is calculated and the sign denotes the polarity of a topic and the number represents the strength. By aggregating the polarity values of all articles, the polarity of a news topic can be obtained and further used to identify the emerging topics.

In our implementation of the text module, the subjectivity lexicon generated by OpinionFinder¹ is employed to measure the affective states and their polarities. The OpinionFinder subjectivity lexicon is a large repository including 8,221 lexicons with weak or strong polarity specifications. Note that this module is domain-independent because no expert knowledge is applied to mark specific keywords for a certain topic. The text module has already been successfully applied in the Fairfax Scoop Digital Editorial Support System², to

dynamically monitor emerging new topics around the world and provide sentimental scores.

B. Facial Expression Module

Fig. 2. Processing steps of the facial expression module

Given the news reports relating to the top ranked topics in the text modal, the images in these reports are then collected, and further input into the facial expression modal, which classifies these images based on the facial expression state of subjects in them, if there exist subjects and faces. For an input image, the face is detected using the Viola-Jones detector and 68 facial points are detected using a self-trained active shape model (ASM). If multiple facial regions are detected in an image, only the one with the largest size is retained. The ASM is trained using a set of annotated facial images collected from the World Wide Web with variations in face pose, subject and illumination conditions, thus it is expected to be able to handle most faces extracted from real-world news images. The local binary patterns (LBP) descriptors are then extracted around 53 interior facial points and further concatenated into a single vector representing texture features. A subset of the most discriminative features is selected using the minimal redundancy maximal relevance criterion (mRMR) algorithm. Geometric features are composed of 43 distances defined using an ASM and face animation parameters defined in the ISO MPEG-4 standard. A feature-level fusion of the LBP subset and 43 FAPs is then employed and input into a support vector machine (SVM) with a radial basis function kernel for classifying three expressions, including neutral, positive, and negative. Details of the system can be found in [17].

IV. EXPERIMENTS

This section evaluates the performance of the proposed system on two facial image datasets: a politician dataset and a news dataset.

A. Data Collection

1) *Politician Facial Image Dataset*: A politician facial image dataset was collected from the World Wide Web to evaluate the classification accuracy of the facial expression module. The choice of “politician” was made due to the fact that a significant proportion of emerging news topics is directly or indirectly related to the government, especially political leaders. Facial images of 12 politicians who were previous (or current) government leaders from seven nations were obtained by inputting their full names as query keywords in the Google Images search engine. This process was facilitated using the tool - Google Image Downloader v2.0. To simulate practical situations as close as possible, no restriction was enforced on the search results, yielding 756 images with various challenging factors, such as variations in image size,

¹ <http://www.cs.pitt.edu/mpqqa>

² <http://www.scoop.smartservicescrc.com.au>

face pose, illumination, occlusion, subject age and gender etc. Table I describes the statistics of the dataset and Fig. 3 illustrates a subset of samples. After performing face detection using Viola-Jones and facial point detection using ASM, 685 out of 756 images were returned with successful results (checked manually by a researcher) and used for the experiment here.

Fig. 3. Samples from the politician facial image dataset

TABLE I
STATISTICS OF THE POLITICIAN FACIAL IMAGE DATASET

Item	Number (Content)
Politician	12 (Angela Merkel, Barack Obama, David Cameron, Francois Hollande, Hillary Clinton, Hu Jintao, John Howard, Julia Gillard, Kevin Rudd, Malcolm Turnbull, Tony Abbott, Vladimir Putin)
Image	756 (55, 55, 62, 54, 57, 56, 52, 106, 82, 61, 64, 50)
Gender	3 female and 9 male
Country	7 (America, Australia, China, France, Germany, Russia, United Kingdom)

2) *News Facial Image Dataset*: A news facial image dataset was also collected based on the top emerging news topics retrieved from the news websites and Twitter using the text module during the period from 1st to 17th February 2013. The top 15 topics identified are used here, and they include AI Gore, Federal Trade Commission, Current TV, KGB, US Federal Reserve, etc. A group of text articles associated with these topics was crawled from the news websites and the images in these articles were then extracted, resulting in a total of 112 articles and 163 images. The images collected have various types of content (e.g. building, landscape, and event) and only around half of them contain one or multiple human faces, as examples shown in Fig.4. By applying the Viola-Jones face detector and the ASM, 112 out of 163 images were identified as containing faces, within which one image has no human face, one images have larger than 80 degree pose rotations, one image is duplicate based on manual inspection, and they were excluded for the evaluation. Thus, the final dataset has 109 facial images from 97 news articles, which are used for the experiment here.

Fig. 4. Samples of the collected news images without (top) or with (bottom) human faces

B. Experimental Set-up

1) *Training Dataset*: An individual dataset is used for training the facial expression module to simulate the variations between training and test data in realistic conditions. The training dataset is the image subset of the QUT-FER dataset [17], which is composed of 4,390 images (1,502 for positive, 1,411 for neutral, and 1,477 for negative) collected from 219 subjects (102 females and 117 males) and three real-world video resources - news, dramas, and YouTube. The facial images contain a lot of realistic challenges such as different emotion intensities, varying face poses, occlusion, gender, and complicated illumination condition etc. Fig. 5 shows a subset of samples. As the training data is completely independent of the test data in the experiment here, they are expected to be different in a wide range of factors, such as subject, pose, age, gender, occlusion, background, illumination, etc., and this train-test strategy warrants the evaluation of the proposed system closer to practical conditions.

Fig. 5. Samples in the training dataset for the facial expression module

2) *Annotation of Database Ground Truths*: Five subjects - all PhD candidates majoring in Computer Science, are recruited to manually annotate the facial expression in the dataset images. They received a pre-training session of about 30 minutes using materials such as the Six Basic Facial Expression⁴. All annotators are then asked to classify all images into one of the three emotions based on their perception on the facial expressions. All results are rechecked and finalized by a researcher with more than seven years research experience on emotion recognition.

The same process is also applied to obtain the affective ground truths of text articles associated with the top news topics identified by the text module. It outputs an affective score ranging from [-1, 1] for each article, which indicates the level of negative and positive in the content respectively. To facilitate direct comparisons with the three categorized emotions produced by the facial expression module, the continuous affective score X for each article is converted into discrete values D by the following equation:

$$X = \begin{cases} \text{Positive} & \text{if } D > t \\ \text{Neutral} & \text{if } -t \leq D \leq t \\ \text{Negative} & \text{if } D < -t \end{cases} \quad (1)$$

where, t is set to 0.05 experimentally. The use of variable t is due to the fact that the text module often produces probability values around zero rather than exact zero for neutral emotion.

⁴ <http://www.cs.unc.edu/~andrei/expressions>

C. Affective Classification Performance

1) *Classification Results on the Politician Facial Image Dataset*: The performance of the facial expression module is obtained by comparing its classification results with manual annotation results. If an image is classified as having one emotion by the facial expression module, while it is perceived as other emotions by the majority of all subjects (i.e. more than three subjects), then a classification error is recorded for this emotion class. The average classification rate over all images is used as the final result.

Table II shows the matrix confusion of the classified results and the overall correct recognition accuracy (i.e. Overall Acc.) for positive, neutral and negative emotions. As can be seen, neutral and negative have similar classification performance with accuracies around 91%, while positive has the lowest accuracy of 80.9%. Negative is prone to be misclassified as positive, while positive is more easily misclassified as neutral. It was observed that a ‘speaking mouth’ impacts the most on the performance of positive emotion, as many politician images are captured during public presentation or in a debate. Faces with a ‘speaking mouth’ are often classified as positive, while the true emotions are negative or neutral, as examples shown in Fig. 6.

TABLE II
CLASSIFICATION RESULTS OF THE FACIAL EXPRESSIONS MODULE ON POLITICIAN FACIAL IMAGE DATASET

	Positive	Neutral	Negative	Overall Acc.
Positive	225	31	22	80.9%
Neutral	3	102	7	91.1%
Negative	22	5	268	90.8%

Fig. 6. Samples misclassified as positive due to a speaking mouth

2) *Classification Results on the News Facial Image Dataset*: Table III reports the results of the facial expression module in classifying images in the news facial image dataset. As can be seen, positive is the easiest emotion for correct classification with 90% accuracy, which is followed by neutral, while negative is the most difficult one with only 50% accuracy and it is misclassified with an approximately equal probability to positive and neutral. This result is contrary to the results on the politician image dataset in Table II, where positive is most difficult for correct recognition and often misclassified as neutral and negative. A closer looking at the facial images between the two datasets reveals that the general people in news topics is not as exaggerated and frequent as politicians in presenting their emotions using facial expressions. It seems that the Google Images already has the function to filter the images with no face, multiple faces, or small faces etc., but there is no such filtering when directly crawling images from news websites. Thus, affective

classification of news images seems to be more challenging than politician images retrieved from image search engines.

TABLE III
CLASSIFICATION RESULTS OF THE FACIAL EXPRESSIONS MODULE ON NEWS FACIAL IMAGE DATASET

	Positive	Neutral	Negative	Overall Acc.
Positive	36	2	2	90.0%
Neutral	4	28	1	84.8%
Negative	10	8	18	50.0%

3) *Comparison of Classified Expression in Facial Images with Affective Content in News Articles*: the news facial image dataset includes not only a set of facial images, but also a set of text articles associated with these images. Intuitively, news editors might prefer to include facial images with the expression that is closely associated with the affective content in news articles. For instance, a happy face is expected in articles reporting a player winning a gold medal or giving a charity donation as shown in Fig.7. Is there any correlation between facial expressions in images and the affect in text articles? To answer this question, we also compare the classified facial expression for each image with the affective ground truths in the associated articles as illustrated in Table IV. If there are multiple facial images in an article, a majority voting is used for obtaining the classified expression result.

Fig. 7. An example of a positive news topic accompanied with a happy facial expression from David Beckham

As can be seen from Table IV, positive emotion has much higher overall classification accuracy than neutral and negative. There appears to exhibit a strong correlation between facial expressions in images and textual content in news articles for positive emotion, with 85% accuracy. It seems that, in most cases, neutral and negative are easier to be misclassified as positive emotion, and this may imply that in real-world news reports, editors tend to adopt images with happy faces even the report has a neutral or negative content. In examples shown in Fig. 8, all the six images contain positive facial expressions, but they actually were used in

reporting negative topics. Thus, the results seem to imply that facial expression is very useful for improving the accuracy of classifying positive affective content in news articles, while it has much lower contributions to that for neutral and negative emotions.

TABLE IV
COMPARISON RESULTS OF CLASSIFIED EXPRESSION IN FACIAL IMAGES WITH AFFECTIVE CONTENT IN NEWS ARTICLES

	Positive	Neutral	Negative	Overall Acc.
Positive	34	3	3	85.0%
Neutral	14	13	6	39.4%
Negative	16	8	12	33.3%

Fig. 8. Samples of news images that have positive facial expressions, but are actually used in reporting negative news topics (e.g. crime)

V. CONCLUSION

This paper presents a machine learning system that adopts the textual information and facial expression information for affective classification of facial images on emerging news topics. It is one of the first systems in this field that is capable of monitoring the current hot news topics using emotional keywords and classifying the facial expressions in news images associated with these topics.

The system is evaluated on two facial expression datasets: a politician dataset collected using the Google Images and a news dataset crawled from news websites. The results demonstrate more than 80% accuracy for all three emotions, except for negative on the news dataset with only 50% accuracy. Neutral and negative emotions are easier for correct classification than positive in politician images, primarily due to a speaking mouth as many politician images are captured during public presentations or in a debate. News images extracted from websites contain more challenging factors in image size, content, and face etc. for affective classification than politician images retrieved from image search engines.

We also analyze the correlation between facial expression in images and affective content in text articles. Positive facial expression shows strong correlation with the affect in text articles while there is only low correlation for neutral and negative. The results also indicate that new editors tend to include images with happy faces in news reports, even these reports talk about neutral or negative topics.

Many open issues still need to be explored in our future work. It is expected that higher accuracy can be achieved by

adopting a suitable decision fusion strategy (e.g. based on the maximum emotion probability) to combine the results from the text module and the facial expression module. It is also interesting to explore whether readers prefer an alignment of negative news using textual and facial images. Automatically collecting facial images of a politician over a certain period can be used to track her (his) emotion changes towards a certain news topic spanning that period. The proposed system can be applied in fields such as assisting editors in choosing facial images with an expression appropriate for a certain news report.

ACKNOWLEDGMENT

This paper was funded by Australian Smart Services CRC and Xi'an University of Technology Grant No. 116-211406.

REFERENCES

- [1] A. Hanjalic, "Extracting moods from pictures and sounds: towards truly personalized TV," *Signal Processing Magazine, IEEE*, vol. 23, pp. 90-100, 2006.
- [2] M. Jana and H. Allan, "Affective image classification using features inspired by psychology and art theory," in *Proceedings of the international conference on Multimedia*, Firenze, Italy, 2010, pp. 83-92.
- [3] N. Bianchi-Berthouze, "K-DIME: an affective image filtering system," *Multimedia, IEEE*, vol. 10, pp. 103-106, 2003.
- [4] A. Hanjalic and X. Li-Qun, "Affective video content representation and modeling," *Multimedia, IEEE Transactions on*, vol. 7, pp. 143-154, 2005.
- [5] J. Niu, *et al.*, "Affivir: An affect-based Internet video recommendation system," *Neurocomputing*, vol. 120, pp. 422-433, 2013.
- [6] L. Canini, *et al.*, "Affective Recommendation of Movies Based on Selected Connotative Features," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 23, pp. 636-647, 2013.
- [7] L. Joonwhoan and P. EunJong, "Fuzzy Similarity-Based Emotional Classification of Color Images," *Multimedia, IEEE Transactions on*, vol. 13, pp. 1031-1039, 2011.
- [8] E. Acar, *et al.*, "Understanding Affective Content of Music Videos through Learned Representations," in *MultiMedia Modeling*, vol. 8325, C. Gurrin, *et al.*, 2014, pp. 303-314.
- [9] N. Liu, *et al.*, "Associating Textual Features with Visual Ones to Improve Affective Image Classification," in *International Conference on Affective Computing and Intelligent Interaction (ACII2011)*, vol. 6974, 2011, pp. 195-204.
- [10] B. Ionescu, *et al.*, "A naive mid-level concept-based fusion approach to violence detection in Hollywood movies," in *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, Dallas, Texas, USA, 2013, pp. 215-222.
- [11] M. Xu, *et al.*, "A three-level framework for affective content analysis and its case studies," *Multimedia Tools and Applications*, pp. 1-23, 2012.
- [12] J. E. Newhagen and B. Reeves, "The Evening's Bad News: Effects of Compelling Negative Television News Images on Memory," *Journal of Communication*, vol. 42, pp. 25-41, 1992.
- [13] R. Yan and A. Hauptmann, "A review of text and image retrieval approaches for broadcast news video," *Information Retrieval*, vol. 10, pp. 445-484, 2007.
- [14] Z. Wang, *et al.*, "Semantic context based refinement for news video annotation," *Multimedia Tools and Applications*, vol. 67, pp. 607-627, 2013.
- [15] S. M. Al Masum, *et al.*, "Emotion Sensitive News Agent: An Approach Towards User Centric Emotion Sensing from the News," in *Web Intelligence, IEEE/WIC/ACM International Conference on*, 2007, pp. 614-620.
- [16] D. Loureiro, *et al.*, "Sentiment Analysis of News Titles," in *Progress in Artificial Intelligence*, vol. 7026, 2011, pp. 1-14.
- [17] L. Zhang, *et al.*, "Facial expression recognition experiments with data from television broadcasts and the World Wide Web," *Image and Vision Computing*, vol. 32, pp. 107-119, 2 2014.