Dynamic Analysis of Automatic Facial Expressions Recognition 'in the Wild' Using Generalized Additive Mixed Models and Significant Zero Crossing of the Derivatives

Damien Dupré
Queen's University Belfast
Belfast, United Kindgom
damien.dupre@qub.ac.uk

Nicole Andelic
Sensum
Belfast, United Kindgom
n.andelic@sensum.co

Gawain Morrison
Sensum
Belfast, United Kindgom
gawain@sensum.co

Gary McKeown
Queen's University Belfast
Belfast, United Kindgom
g.mckeown@qub.ac.uk

The analysis of facial expressions is currently a favoured method of inferring experienced emotion, and consequently significant efforts are currently being made to develop improved facial expression recognition techniques. Among these new techniques, those which allow the automatic recognition of facial expression appear to be most promising. This paper presents a new method of facial expression analysis with a focus on the continuous evolution of emotions using Generalized Additive Mixed Models (GAMM) and Significant Zero Crossing of the Derivatives (SiZer). The time-series analysis of the emotions experienced by participants watching a series of three different online videos suggests that analysis of facial expressions at the overall level may lead to misinterpretation of the emotional experience whereas non-linear analysis allows the significant expressive sequences to be identified.

Emotion, Facial Expression, Automatic Recognition, Generalized Additive Mixed Model, Significant Zero Crossing of the Derivatives

1. INTRODUCTION

Understanding emotions represents a major issue for analysing not only Human-Computer Interactions (Picard, Vyzas & Healey 2001) but also Human-Virtual agents Interactions (Moridis & Economides 2012) and Human-Robot Interactions (Schacter, Wang, Nejat, *et al.* 2013). The analysis of facial expressions is currently a preferred method when inferring experienced emotion, and consequently significant efforts are currently being made to develop improved facial expression recognition techniques.

By labelling facial expressions into particular and meaningful categories/dimensions, automatic recognition systems provide insights about the evolution of individuals' emotional states. However, interpreting their results over time remains an important problem when dealing with large sample size and time series. Using statistical methods called Generalized Additive Mixed Models (GAMM) (Wood 2006) and Significant Zero Crossing of the Derivatives (SiZer) (Chaudhuri & Marron 1999), this paper presents a new way to identify the significant evolution of the emotion time series provided by the automatic recognition systems in order to take into account the idiosyncrasy of facial expressions 'in the wild'.

1.1 Facial expressions and emotions

Smiling, frowning, clenching the jaws or widening eyes are examples of movements that are perceived in others' faces most often without paying attention to them. Yet their importance is crucial to the smooth running of our social lives. The coordination of these facial movements is known as facial expression and it regulates the interactions that one maintains daily. Indeed, not understanding the expressions of others makes it difficult to initiate communication, or to prolong it (Chovil 1997). As evidence, patients with Mobius syndrome (Ekman 1992) or with Botulinum toxin (Botox) injections (Havas, Glenberg, Gutowski, *et al.* 2010) suffer more social exclusion than non-affected individuals. Nevertheless, the meaning of facial expressions is still questioned. By decomposing them into particular and meaningful categories/dimensions, researchers try to infer the meaning of these facial movements (Smith, Cottrell, Gosselin, *et al.* 2005). However, the way of making these inferences remains the subject of a

Published by BCS Learning and Development Ltd. Proceedings of British HCI 2018. Belfast, UK.

lively debate between the proponents of each theory.

As one of the most visible components related to emotions, facial expressions have often been given a privileged position in emotion related research. Although there is debate about their function and about their social implication, the connection between facial expressions and emotions is rarely questioned (Scherer 2005). Nevertheless it may be that while expressions are largely social and have socio-communicative motives–on occasions we are socially motivated to display our emotional state.

Two main methods can be used to investigate facial expressions of emotions. As Darwin's preferred method (Darwin 1872), observation of others' facial expressions provide useful information about behavioural displays. However, this method is laborious, time consuming and requires a large pool of annotators (expert or novice) to overcome their subjectivity. With the development of machine learning algorithms, a second method allows the automatic recognition of facial expressions. Automatic recognition systems are constantly improving their ability to assess the underlying muscle movements associated with facial expressions (Zeng, Pantic, Roisman, *et al.* 2009). The modern prevalence of recording cameras or web cams allows large amounts of this kind of data to be captured and facial expressions can be analysed in real-time as people watch television and web content on monitors in their normal household environment.

1.2 Automatic recognition of facial expressions

In order to analyze facial expressions, automatic recognition systems evaluate individual and compound movements most often using a Facial Action Coding System (FACS) (Ekman, Friesen & Hager 2002; Ekman & Friesen 1978) based approach. In the FACS, facial movements are categorized according to sets of muscle actions or Action Units (AU). The FACS method is agnostic to any relationship with felt emotions and only seeks to provide an objective classification of the observed facial muscle movement. The interpretation of emotional configurations for the facial movements is given by the EmFACS (Friesen & Ekman 1983). In this complementary work, six prototypic emotions—happiness, surprise, disgust, sadness, fear and anger—have been described with their corresponding AUs.

Based on systems that can automatically recognize EmFACS configurations or similar sets of emotional labels, significant effort has been put into the attempt to create accurate facial expression analyses of natural emotions. For example, challenges such as Facial Expression Recognition and Analysis challenge (FERA) (Valstar, Almaev, Girard, *et al.* 2015) or Audio/Visual Emotion Challenge (AVEC) (Schuller, Valstar, Eyben, *et al.* 2011) aim to develop the most efficient algorithms to recognize emotions both in laboratory settings and "in the wild" (Sandbach, Zafeiriou, Pantic, *et al.* 2012). Among the existing systems, FacioMetrics LLC[1] has developed a system that is easy to implement with web cam recordings (Xiong & De la Torre 2013). The FacioMetrics system can extract facial expression information from online sources and classify it in terms of the related emotion labels such as disgust, happy, sadness, and surprise and also the more functional labels of focus and attention as well as providing a baseline classification of facial expressions when a face is deemed to be in a neutral state.

Even though some systems are based on a dimensional perspective of emotions (e.g. Nicolaou, Gunes & Pantic 2011), most of automatic systems for facial expression recognition are based on a conception of discrete and basic emotions described by a short number of labels. However, other theoretical conceptions do not view emotions as taking a discrete or basic form (Russell 2016). For these conceptions, the use of specific communicative labels allows the persistent alignment of more complex and context dependent emotional states. Emotional labels serve a very useful communicative function and their prevalence in language and popular culture make them an important part of interaction in commerce related affective computing endeavours. Therefore, while there are important caveats and debates on the value of discrete emotions as abstract representations of emotional state, they remain culturally useful as communicative labels for the alignment of the component processes associated with emotional states.

1.3 Question asked by the remote recording of facial expressions

Implementing automatic recognition such as FacioMetrics system in online survey platforms allows researchers to reach a very high number of participants. The benefits of such panels are enormous for fast and efficient conduct of research but they raise questions concerning the measurement of facial expressions 'in the wild'. However, measuring emotions expressed in household environment limits the control over experimental conditions. It can be difficult to ascertain exactly how an experiment was conducted and there is a reliance on the honesty of the participant to be actively engaging in any experiment. When testing the power of commercial video experiences, they are an appealing option as the experimental setup is likely to be very similar to any web-browsing or television watching experience.

[1] FacioMetrics was acquired by Facebook inc. in 2017.

The ability to evaluate emotional reactions remains a difficult issue. One option that permits the evaluation of emotional reactions is enabled by the fact that most modern computers have webcam abilities either embedded in a computer monitor or easily attached as an inexpensive peripheral device. This means that the screen capable of presenting the material can also capture facial recordings of the person watching the associated material; these facial recordings can then be subjected to analysis of facial expressions for signs of emotional reactions. Therefore, in order to evaluate the facial expressions of emotions recorded by webcam, emotional videos were presented to participants using a web platform called *Sensum Insights*. The goal of the study was to assess the dynamic evolution that would be required to detect emotional reactions and to discriminate between them.

2. EXPERIMENT

In the current study the FacioMetrics system was incorporated into an online research platform created using Sensum Insights to provide real-time facial coding in combination with eye tracking and implicit testing. A recruitment provider was used to access a large panel of participants. Online participants viewed the content on their PC or Laptop. They were instructed to watch the content with correct lighting and viewing conditions.

2.1 Participants

For this study, 190 participants (93 males, 97 females, age M = 44.9, SD = 14.7) were recruited via online survey platforms and gave their consent to be recorded in advance. They were rewarded £4.20 for their participation.

2.2 Material

In order to compare emotions induced in different ways, three different TV commercials were chosen according to their content:

- Video 1 is a TV commercial (30s) in which a lawyer presents a bankruptcy solution. The tonality of the message is monotonic and no noticeable changes happen throughout the video except the final screen which summarises the message given.
- Video 2 is a TV commercial (20s) for a discount supermarket chain. In this a male model in underwear presents champagne with an unexpected voice.
- Video 3 is a TV commercial (18s) for a coffee drink, showing a couple on the beach watching the sunset followed by the sudden appearance of a monster screaming.

2.3 Measures

The FacioMetrics recognition system measures four basic emotions based on EmFACS (i.e. Happiness, Surprise, Sadness and Disgust) as well as two cognitive states (i.e. Attention and Focus) frame by frame (Table 1).

Table 1: *Facial feature associated to the recognition of the labels provided by FacioMetrics System*

Label recognized	Facial feature associated
Happiness	*Cheek Raiser, Lip Corner Puller*
Surprise	*Inner Brow Raiser, Outer Brow Raiser, Upper Lid Raiser, Jaw Drop*
Disgust	*Nose Wrinkler, Lip Corner Depressor, Lower Lip Depressor*
Sadness	*Inner Brow Raiser, Brow Lowerer, Lip Corner Depressor*
Attention	*Front face looking at the media displayed*
Focus	*Distance to the screen*

FacioMetrics's system also provides a confidence level of the recognition based on the quality of the measurement performed.

2.4 Protocol

Respondents were required to setup their viewing space with correct lighting before passing a webcam calibration test. This was to ensure a good quality data set. Each participant was asked to watch the video 1, 2 and 3 in a fixed order.

2.5 Data analysis

Because of the dynamic changes in emotion recognition time-series they are not suited to fitting with linear regression models, a non-linear linear analysis is required–although, on rare occasions, a linear increase or decrease may occur.

By estimating the degree of smoothness of a Bayesian spline smoothing using restricted maximum likelihood estimation (REML) (Wood 2011; Lin & Zhang 1999), Generalized Additive Mixed Models (GAMM) allow the identification of dynamic patterns underlying time-series while taking into account participants' idiosyncratic response (see McKeown & Sneddon 2014 for application with self-report analysis). The following model was tested:

$$y = X\beta + Zu + \varepsilon$$

$$u \sim N(0, \Psi\theta)$$

$$\varepsilon \sim N(0, \Lambda\sigma^2)$$

where y is the emotion recognition vector, X is the model design matrix according to the time (or Frames in our case), β is the β coefficient vector. Moreover, u contains a random effects vector, and Z is a model matrix for these random effects (for each participant), Ψ is the covariance matrix, and θ the unknown parameters within that covariance matrix. Λ is a matrix that is part of the error term and which can be used to model the residual autocorrelation. Finally, ε is the error term.

The output of the GAMM analysis is both an evaluation of the smoothness of the time series provides by effective degrees of freedom (edf) and the predicted trend resulting of the GAMM analysis. The edf refers to the optimal number of knots in the time series, the higher the edf, the more non-linear is the smoothing spline (Zuur, Ieno, Walker, *et al.* 2009).

Using GAMMs to analyse emotion recognition time series, it is possible to identify subtle changes nested in individual's idiosyncratic response to the video stimuli (Dupré, Booth, Bolster, *et al.* 2017). Even though GAMMs allow the assessment of time-series changes, it does not provide a statistical analysis of where these changes happen. Therefore, a Significant Zero Crossing of the Derivatives (SiZer) approach offers a method to identify the significant changes in the GAMM predicted values. SiZer methods enable meaningful statistical inference, while doing exploratory data analysis using statistical smoothing methods (Chaudhuri & Marron 1999). The SiZer analysis uses the kernel density estimation provided by the GAMM analysis. Kernel density estimator is defined in the following way (Rosenblatt 1956; Parzen 1962):

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$

where $X_1, X_2, \ldots, X_n$ is the random n-element sample, h is the smoothing parameter, $K(.)$ is the kernel function.

With the statistical analysis of the changes in the GAMM, SiZer methods extract the significant increasing and decreasing periods in the emotion recognition time series.

3. RESULTS

First of all, we analysed the descriptive statistic of emotion recognition for each video (Table 2).

Table 2: *Mean and Standard Deviation of emotion recognition for the three videos tested according the labels provided by FacioMetrics' system. Note. M and SD represent mean and standard deviation, respectively.*

	Video 1		Video 2		Video 3	
Label	***M***	***SD***	***M***	***SD***	***M***	***SD***
Happy	*0.04*	*0.15*	*0.07*	*0.22*	*0.05*	*0.18*
Surprise	*0.04*	*0.16*	*0.04*	*0.16*	*0.03*	*0.15*
Sadness	*0.25*	*0.40*	*0.24*	*0.38*	*0.21*	*0.37*
Disgust	*0.09*	*0.23*	*0.10*	*0.24*	*0.10*	*0.24*
Attention	*0.80*	*0.37*	*0.81*	*0.35*	*0.79*	*0.37*
Focus	*0.06*	*0.06*	*0.05*	*0.05*	*0.06*	*0.07*

Regardless the video type, average recognition rates for Attention reach 80 per cent which indicate that the participants have watched the videos as requested. However average recognition rates for Focus are fewer than 10 per cent which is indicates a posture away from the front camera that the participants were using. This result can be explained by the fact that the participants were at their own home in a comfortable position and watched the videos in a relaxed state.

Regarding average recognition of emotion labels, even if these overall descriptive statistics show a low recognition of the emotion for each video, the Fixed-Effects ANOVA (Table 3) shows differences between the labels is significant ($F_{(5,\ 2293258)} = 498136.26$, $p < .001$) as well as between the videos ($F_{(2,\ 2293258)} = 267.03$, $p < .001$).

This significant difference is due not only to the important number of participants, and thus data gathered, but also to the idiosyncrasy of emotional facial expressions that increased the residual spreading out. A second interesting result is the similar pattern across the different videos. The FacioMetrics system seems to measure a facial resting state made of negative emotional expressions with sadness and disgust recognition. It suggests that the emotion elicitation triggered by the video was not effective. This assumption is also supported by the evolution of the recognition rate according the time (Figure 1). At a first sight no

Table 3: *Fixed-Effects ANOVA results using Emotion as the criterion. Note. LL and UL represent the lower-limit and upper-limit of the partial η^2 confidence interval, respectively.*

Predictor	**Sum of Squares**	***df***	**Mean Square**	***F***	***p***	**partial η^2**	**partial η^2 90% CI [LL,UL]**
(Intercept)	*856.97*	*1*	*856.97*	*12715.68*	*<.001*		
Label	*167858.88*	*5*	*33571.78*	*498136.26*	*<.001*	*.52*	*[.52,.52]*
Video	*35.99*	*2*	*18.00*	*267.03*	*<.001*	*.00*	*[.00,.00]*
Error	*154553.58*	*2293258*	*0.07*				

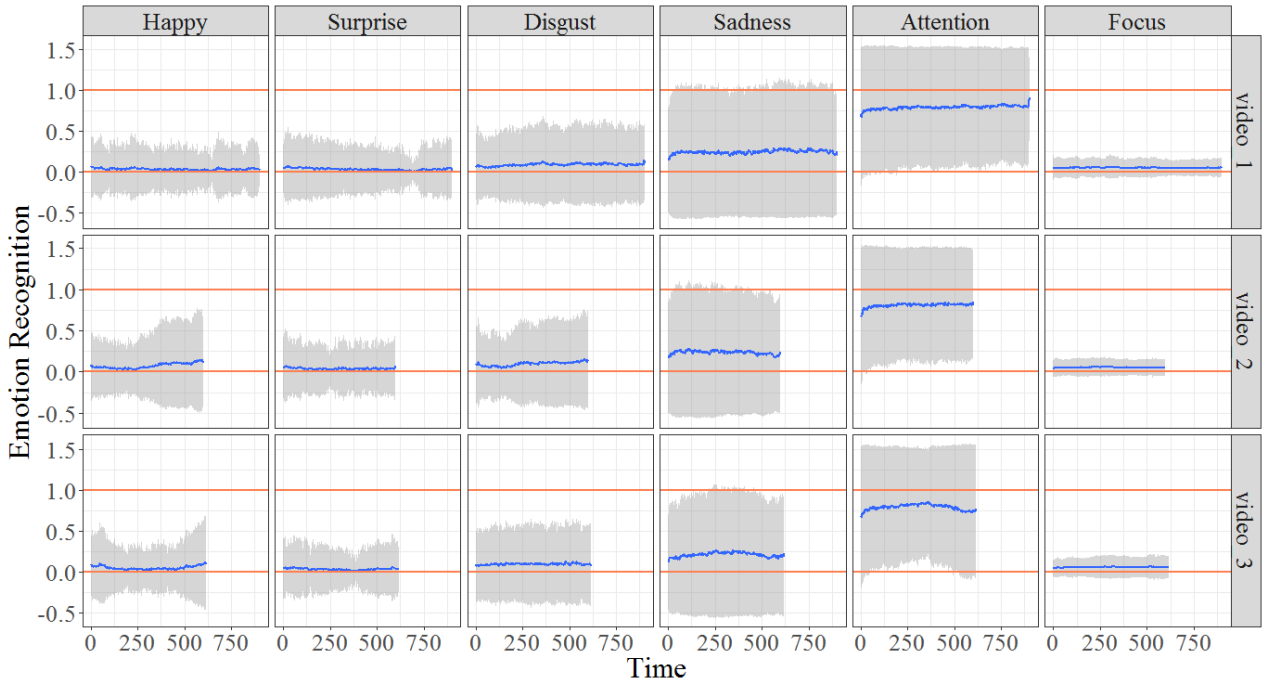

Figure 1: *Mean (blue) and standard deviance (grey) of label recognitions according the time for the three videos.*

specific pattern seems to appear from the raw data.

However, in the previous analysis a determinant variable was not taken into account: the individual emotion expression dynamic. Indeed, emotions are defined as subtle and fast phenomenon (Scherer 2005). Therefore, not taking this dynamic perspective into account is a mistake that must be avoided in emotion recognition analysis. Thus, we analysed the filtered data using a Generalized Additive Mixed Model (Wood, Pya & Säfken 2016) which include in the model a smooth pattern to analyse the evolution of emotion recognition according the time.

As expected, there is a significant time dependent pattern in the emotion recognition that explains the descriptive results (Table 4).

Indeed, the analysis of emotion recognition with overall indicators has important limits. The mean analysis hides the dynamic evolutions of emotion recognition which are not revealed with standard deviation. The use of time dependent model is thus a requirement to understand the automatic emotion recognition. The plots of the Generalized Additive Mixed Model for each video reveal the temporal dynamic of facial expressions changes (Figure 2).

Contrary to a descriptive observation of the evolution of the mean values over time, the GAMM analysis shows different variations of the emotion expression according to time and the videos by taking into account time-series autocorrelation as well as the participants as a random variable. Even if the variation intensity is subtle, it is now possible to infer the influence of the video on the emotion recognition. Moreover a second analysis of the significant changes in the GAMM predicted values will accurately identify the increase and decrease in the facial expressions. Thus a SiZer analysis was performed on the first derivatives of the GAMM with a 95% point-wise confidence interval (Figure 3).

Table 4: *Approximate significance of smooth terms for the Generalized Additive Mixed Model for each emotion in the Video 1, 2 and 3. Signif. codes: * p < .05, ** p < .01, *** p < .001.*

	Video 1			Video 2			Video 3		
Predictor	**edf**	**Ref.df**	**F**	**edf**	**Ref.df**	**F**	**edf**	**Ref.df**	**F**
Happy~Frame	*2.984*	*2.984*	*10.9****	*7.599*	*7.599*	*29.17****	*5.385*	*5.385*	*27.44****
Surprise~Frame	*3.495*	*3.495*	*16.76****	*7.869*	*7.869*	*8.018****	*7.118*	*7.118*	*13.69****
Disgust~Frame	*2.685*	*2.685*	*8.19****	*6.351*	*6.351*	*15.45****	*1*	*1*	*2.332*
Sadness~Frame	*4.987*	*4.987*	*3.39***	*2.43*	*2.43*	*8.201****	*3.407*	*3.407*	*8.591****
Attention~Frame	*7.668*	*7.668*	*22.84****	*7.807*	*7.807*	*33.19****	*7.939*	*7.939*	*47.15****
Focus~Frame	*8.759*	*8.759*	*27.9****	*8.786*	*8.786*	*50.71****	*8.765*	*8.765*	*34.08****

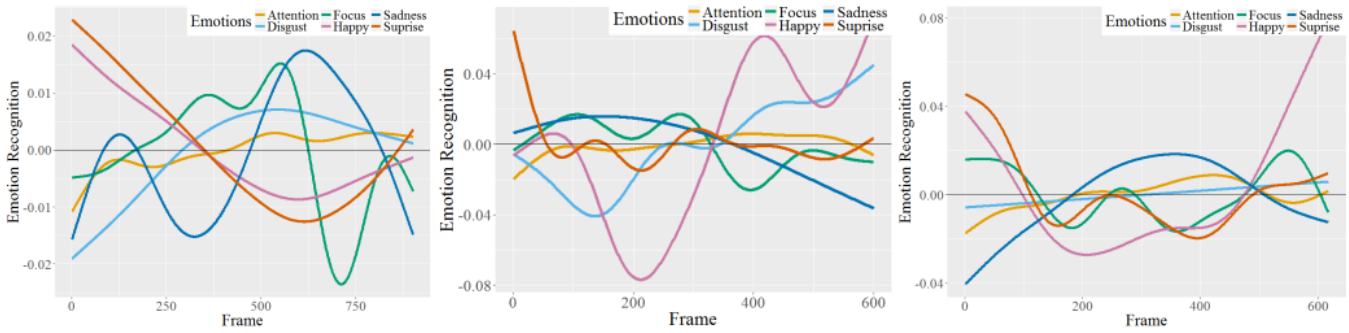

Figure 2: *Temporal evolution of spline smooth models provided by the fitted values of the GAMM analysis for the Video 1, 2 and 3. GAMM fitted values are standardized compered to predicted values.*

Video 1 was monotonic which explained a subtle increase in negative facial expression captured by the *Disgust* and *Sadness* recognition during the first part of the video. Inversely, positive facial expressions (i.e. *Happy* and *Surprise*) are decreasing over the time.

Video 2 first presents a male model in underwear and second his unexpected voice. After the second event, the GAMM evolution reveals an increase in *Happy* recognition which corresponds to the humorous impact of the unexpected voice on participants' facial expressions.

Video 3 showed a sudden appearance of a zombie with a huge scream towards the end of the video. This negative surprise appears to have been captured by the *Sadness* classification. However, the GAMM showed an important increase of the *Happy* and *Surprise* expressions at the end of the video. This positive expression may have resulted from relief after the previously felt disturbance.

4. DISCUSSION

Due to the questions being asked within this paper concerning the display facial expressions of emotion it is important to realize that while we do recognize the communicative value of these labels from a cultural perspective, we remain agnostic to the function of the expressions in terms of serving as readouts and indices of felt emotions or as socio-communicatively motivated signals of emotional state.

In this paper we asked questions of the data that we collected. We hypothesized that due to the dynamic nature of facial expressions classic analyses using overall descriptive results can lead

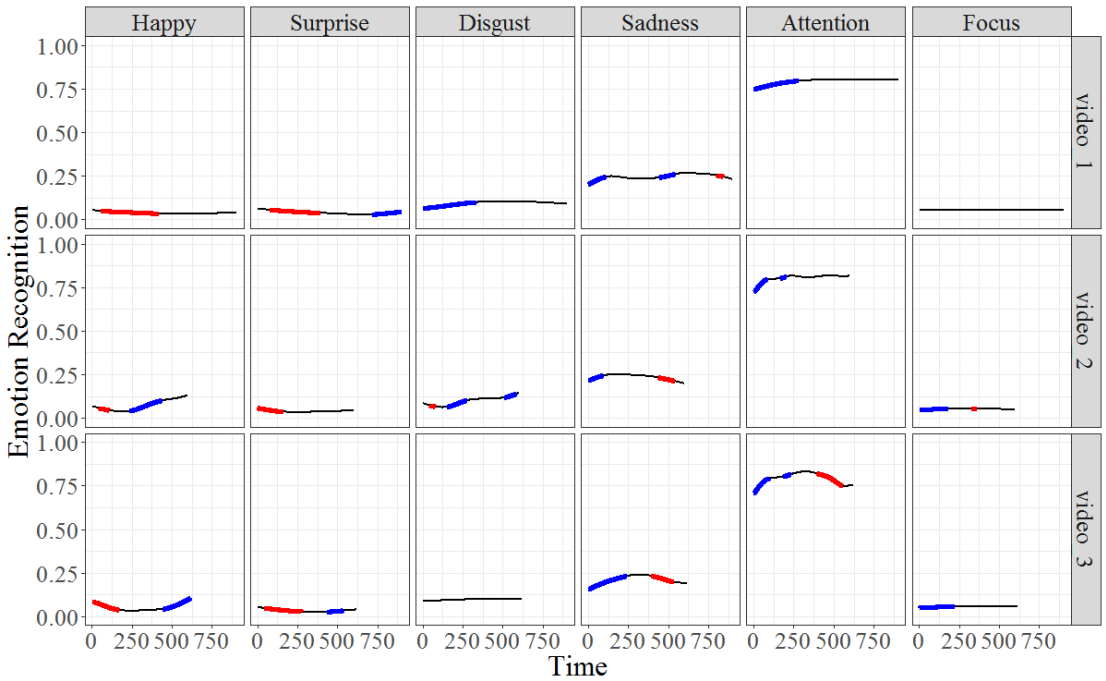

Figure 3: *SiZer analysis of the GAMM predicted values. A 95% point-wise confidence interval is shown. Significant periods are extracted from the first derivatives and reported on the actual GAMM predicted values where red periods indicate a significant decrease and blue periods a significant increase.*

to misinterpretations and mistakes. As a sub-hypothesis to this, we wondered how easy it would be to differentiate the different expressions in terms of their emotional labels if they were shown to be low in terms of intensity. Using Generalized Additive Mixed Models and Significant Zero Crossing of the Derivatives methods with automatic recognition systems is an opportunity to evaluate both dynamic pattern and intensity levels.

A first analysis of the emotions detected by the system suggests that taking a non-dynamic approach may lead to misinterpretations. Although statistics overall show a low emotion recognition for each video, the difference in accuracy between the emotion labels is significant. This significant difference is not only due to the sample size but also to the idiosyncrasy of emotional facial expressions that lead to an increase in the spread of the residuals. Moreover, the automatic recognition system seems to measure a facial resting state made of negative emotional expressions with sadness.

Because emotions are defined as subtle and rapidly changing phenomenon (Scherer 2005), we used a Generalized Additive Mixed Model with a smooth pattern to analyse the evolution of emotion recognition over time. Then we used the fitted data as an input of the Significant Zero Crossing of the Derivatives. As expected, there is a significant time dependent pattern in the emotion recognition that explains the descriptive results. The first video was monotonous which explained a subtle increase in negative facial expression captured by the *Sadness* recognition during the last part of the video. Inversely, positive facial expressions (i.e. *Happy* and S*urprise*) decrease over time. The second video paired a male model in underwear with an unexpected comical voice. After the voice was introduced, GAMM and SiZer evolution reveal an increase in *Happy* labelling of the participants' facial expression due to the humorous impact of the unexpected voice. The third video showed the sudden appearance of a zombie with a loud scream. This negative surprise was again captured by the *Sadness* classification. However, the GAMM and SiZer methodology showed an important increase of the *Happy* expression at the end of the video. This positive expression was probably due to the relief after the previously felt disturbance.

Using a paradigm based on EmFACS, the current automatic recognitions systems struggle to manage the complexity and subtleness of emotional facial expressions, resulting in high levels of noise in their data. However, Generalized Additive Mixed Models and Significant Crossing of the Derivatives are promising statistical methods that can be used to remove this noise and to identify significant patterns in the emotion recognition signal. Nevertheless, using a paradigm based on EmFACS, automatic recognitions cannot manage the complexity and the subtleness of emotion related facial expressions.

5. CONCLUSIONS

Measuring facial expressions with a webcam is a very fast and easy way to assess individual emotions. However, it runs up against the important psychological issue of emotion related expressions complex and dynamic characteristics. Using Generalized Additive Mixed Models and Significant Zero Crossing of the Derivatives offers an opportunity to assess these characteristics. Therefore, this methodology should be considered for in future analyses of facial expressions given by automatic recognition systems.

6. REFERENCES

Chaudhuri, P. and Marron, J.S. (1999) SiZer for exploration of structures in curves. *Journal of the American Statistical Association*. 94 (447), 807–823.

Chovil, N. (1997) Facing others: A social communicative perspective on facial displays. In: J. A. Russell and J. M. Fernández-Dols (eds.). *The psychology of facial expression*. Cambridge university press. Cambridge, UK. pp. 321–333.

Darwin, C. (1872) *The expression of the emotions in man and animals*. John Murray. London, UK.

Dupré, D., Booth, A., Bolster, A., Morrison, G., et al. (2017) Dynamic Analysis of Automatic Emotion Recognition Using Generalized Additive Mixed Models. In: *Symposium on Computational Modelling of Emotion:Theory and Applications*. Bath, UK. pp. 158–163.

Ekman, P. (1992) An argument for basic emotions. *Cognition & Emotion*. 6 (3–4), 169–200.

Ekman, P., Friesen, W. and Hager, J. (2002) *New Version of the Facial Action Coding System*. A Human Face Publication. Salt Lake City, USA.

Ekman, P. and Friesen, W.V. (1978) *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press. Palo Alto, USA.

Friesen, W.V. and Ekman, P. (1983) *EMFACS-7: Emotional facial action coding system*

Havas, D.A., Glenberg, A.M., Gutowski, K.A., Lucarelli, M.J., et al. (2010) Cosmetic use of botulinum toxin-A affects processing of emotional language. *Psychological Science*. 21 (7), 895–900.

Lin, X. and Zhang, D. (1999) Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 61 (2), 381–400.

McKeown, G.J. and Sneddon, I. (2014) Modelling continuous self-report measures of perceived emotion using generalized additive mixed models. *Psychological methods*. 19 (1), 155–174.

Moridis, C.N. and Economides, A.A. (2012) Affective learning: Empathetic agents with emotional facial and tone of voice expressions. *IEEE Transactions on Affective Computing*. 3 (3), 260–272.

Nicolaou, M.A., Gunes, H. and Pantic, M. (2011) Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *IEEE Transactions on Affective Computing*. 2 (2), 92–105.

Parzen, E. (1962) On estimation of a probability density function and mode. *The annals of mathematical statistics*. 33 (3), 1065–1076.

Picard, R.W., Vyzas, E. and Healey, J. (2001) Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 23 (10), 1175–1191.

Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*. 832–837.

Russell, J.A. (2016) A Sceptical Look at Faces as Emotion Signals. In: C. Abell and J. Smith (eds.). *The Expression of Emotion: Philosophical, Psychological and Legal Perspectives*. Cambridge, UK, Cambridge University Press. pp. 157–172.

Sandbach, G., Zafeiriou, S., Pantic, M. and Yin, L. (2012) Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image and Vision Computing*. 30 (10), 683–697.

Schacter, D., Wang, C., Nejat, G. and Benhabib, B. (2013) A two-dimensional facial-affect estimation system for human–robot interaction using facial expression parameters. *Advanced Robotics*. 27 (4), 259–273.

Scherer, K.R. (2005) What are emotions? And how can they be measured? *Social Science Information*. 44 (4), 695–729.

Schuller, B., Valstar, M., Eyben, F., McKeown, G., et al. (2011) *Avec 2011–the first international audio/visual emotion challenge*. In: International Conference on Affective Computing and Intelligent Interaction. pp. 415–424.

Smith, M.L., Cottrell, G.W., Gosselin, F. and Schyns, P.G. (2005) Transmitting and decoding facial expressions. *Psychological science*. 16 (3), 184–189.

Valstar, M.F., Almaev, T., Girard, J.M., McKeown, G., et al. (2015) Fera 2015-second facial expression recognition and analysis challenge. In: *International Conference and Workshops on Automatic Face and Gesture Recognition*. 2015 Ljubljana, Slovenia. pp. 1–8.

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 73 (1), 3–36.

Wood, S.N. (2006) Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models. *Biometrics*. 62 (4), 1025–1036.

Wood, S.N., Pya, N. and Säfken, B. (2016) Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*. 111 (516), 1548–1563.

Xiong, X. and De la Torre, F. (2013) Supervised descent method and its applications to face alignment. In: *IEEE Conference on Conference on Computer Vision and Pattern Recognition*. 2013 Portland, USA. pp. 532–539.

Zeng, Z., Pantic, M., Roisman, G.I. and Huang, T.S. (2009) A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 31 (1), 39–58.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., et al. (2009) *Mixed effects models and extensions in ecology*. Spring Science and Business Media. New York, USA.