# Identify computer generated characters by analysing facial expressions variation

Duc-Tien Dang-Nguyen, Giulia Boato, Francesco G.B. De Natale

*Department of Information and Computer Science - University of Trento*
*via Sommarive 5, 38123 Trento - Italy*
dangnguyen@disi.unitn.it
boato@disi.unitn.it
denatale@ing.unitn.it

***Abstract*—Significant improvements have been recently achieved in both quality and realism of computer generated characters, which are nowadays often very difficult to be distinguished from real ones. However, generating highly realistic facial expressions is still a challenging issue, since synthetic expressions usually follow a repetitive pattern, while in natural faces the same expression is usually produced in similar but not equal ways. In this paper, we propose a method to distinguish between computer generated and natural faces based on facial expressions analysis. In particular, small variations of the facial shape models corresponding to the same expression are used as evidence of synthetic characters.**

## I. Introduction

Digital graphics tools are nowadays widespread and exploited to create realistic computer media data both from professional and non-professional users. In particular, computer generated (CG) characters are increasingly used in many applications such as talking-faces, e-learning, virtual meeting and especially video games. Since the first virtual newsreader Ananova[1] introduced in 2000, significant improvements have been achieved in both quality and realism of CG characters, which are nowadays often very difficult to be distinguished from real ones.

At one hand, these results open a new area for advance human-computer interaction. On the other hand, non existing subjects or situations can be generated leading to the need of techniques assessing data trustability and authenticity with sufficient confidence. Therefore, the research community has recently focused on the development of tools supporting the discrimination between natural and CG multimedia content in an accurate and reliable way.

In multimedia forensics, approaches distinguishing between CG and natural data have been developed since 2005. Most of them focus on still images, by estimating statistical differences in wavelet-based decomposition [1][2]; by modelling physical differences like local patch statistics, fractal and quadratic geometry, and gradient on surface [3]; by evaluating the noise of the recording device [4]; by combining different informations like the hybrid approach in [5]. Recently, a geometric approach supporting the distinction of CG and real human faces has been presented in [6], which exploits face asymmetry as a discriminative feature. However, to the best of our knowledge, there is no multimedia forensics approach that aims at discriminating between CG versus natural objects or subjects in video sequences. Such a goal requires different techniques with respect to the state-of-the-art.

In this paper we propose a method to distinguish between CG and real characters by analysing facial expressions. Reproducing facial expressions is one of the most challenging issues in creating virtual characters [7] and there are studies back to 1971 that analyse this problem (see for instance [8]). Most of the algorithms generate synthetic facial expressions following the Facial Action Coding System (FACS) by Ekman [9][10] or MPEG-4 standard [11]. In FACS, the muscles on face are coded as Action Units (AUs), and an expression is then represented as a combination of AUs. In MPEG-4, explicit movements for each point on the face is defined by Facial Animation Parameters (FAPs). Based on these parameters (FACS or FAPs), a physically-based model is applied to make it more realistic. However, when CG contents become very realistic they often become also unfamiliar (so called *Uncanny valley* [12]) and recently some approaches attempt to overcome such a problem [7][13]. Here, we propose to exploit this gap to differentiate between computer generated and natural faces.

The underlying idea is that facial expressions in CG characters follow a repetitive pattern, while in natural faces the same expression is usually produced in similar but not equal ways (e.g., human beings do not always smile in the same way). Our forensic technique take as input various instances of the same character expression (extracting corresponding frames of the video sequences) and determine whether the character is CG or natural based on the analysis of the corresponding variations. We show that CG faces often replicate the same expression exactly in the same way, i.e., the variations is smaller than the natural ones, and can therefore be automatically detected.

The rest of this paper is organized as follows: the proposed method is described in section II, experimental results are reported in section III, while section IV draws some conclusions.

[1] http://news.bbc.co.uk/2/hi/entertainment/718327.stm

*WIFS'2012, December, 2-5, 2012, Tenerife, Spain.*
*978-1-4673-2287-4/12/$31.00 ©2012 IEEE.*

## II. Proposed Method

Our method contains five steps as detailed in Figure 1. From a given video sequence, frames that contain human faces are extracted in the first step A. Then, in step B, facial expression recognition is applied in order to recognize the expressions of the faces. Six types of facial expressions are used in this step, following the six universal expressions of Ekman (happiness, sadness, disgust, surprise, anger, and fear) [9] plus a 'neutral' one. Based on the recognition results, faces corresponding to a particular expression (e.g., happiness) are selected for the next steps. Notice that the 'neutral' expressions are not considered, i.e., faces showing no expression are not taken into account for further processing. In the next step C the Active Shape Model (ASM), which represents the shape of a face, is extracted from each face. In order to measure their variations, all shapes have to be comparable. Thus, in step D, each extracted ASM is then normalized to a standard shape. After this step, all ASM shapes are normalized and are comparable. Finally, in step E, differences between normalized shapes are analysed, and based on the variation analysis results, the given sequence is confirmed to be CG or natural.

The right part of Figure 1 shows an illustration of the analysis procedure on happiness expression. Seven frames that contain faces are extracted in step A. Then, facial expression recognition is applied in step B and three happy faces are kept. For each face, the corresponding ASM model, which is represented by a set of reference points, is extracted in step C. Then, each model is normalized to a standard shape, step D. All normalized shapes are then compared together in step E, and based on the analysis results the given character is confirm as computer generated since the differences between the normalized shapes are small (details about the variation analysis are given in the following Subsection II-E).

### A. Human faces extraction

Face detection problem has been solved with the Viola-Jones method [14], which can be applied in real-time applications with a high accuracy. In this step, we reuse this approach to detect faces from video frames, and frames that contain faces are extracted. More details about this well-known method can be found in [15] and [14]. It is worth mentioning that in this first work we do not face the problem of face recognition, thus assuming to have just a single person per video sequence (the analysed character).

### B. Facial expression recognition

Facial expression recognition is a nontrivial problem in facial analysis. In this study, we applied an EigenFaces-based application [16] developed by Rosa for facial expression recognition. The goal of this step is to filter out the outlier expressions and keep the recognized ones for further steps. Notice that this application associates an expression to a given face without requiring any detection of reference points. In Figure 1 an example of results of this application is shown with 7 faces (3 happy, 2 disgust, 1 surprise and 1 neutral).

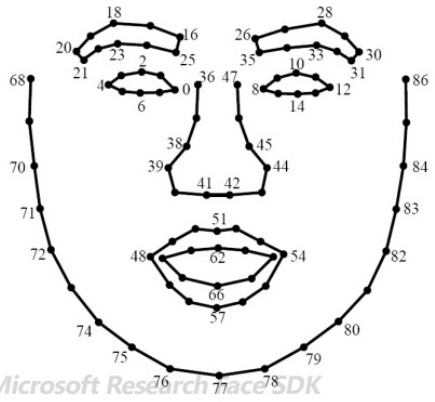



Fig. 2. The 87 points of Active Shape Model (ASM). Source: Microsoft Research Face SDK.

### C. Active Shape Model Extraction

Input images for this step are confirmed to have the same facial expression of the same person, thanks to the preprocessing in the first two steps. In order to extract face shapes, which are used in our analysis, an alignment method is applied. In this step, we follow the Component-based Discriminative Search approach [17], proposed by Liang et al. The general idea of this approach is to find the best matching from the mode candidates, where modes are important predefined points on face images (e.g., eyes, nose, mouth) and are detected from multiple component positions [17]. Given a face image, the result of this step is a set of reference points, representing the detected face. In Figure 3 (a) an example of this step is shown, where the right image shows reference points representing the face in the left image. In this method, the authors exploit the so called ASM, which contains 87 reference points as shown in Figure 2. Another example of this step on a CG face is also reported in Figure 3 (c), where the left image shows the synthetic facial image and the right one shows the corresponding ASM.

### D. Normalized Face Computation

ASM models precisely and suitably represent faces, but they are incomparable since faces could be different in sizes or orientations. They need to be normalized in order to be comparable. In this step, we apply the traditional approach from [18] to normalize a shape of a face in order to have a common coordinate system. This normalization is an affine transformation used to transform the reference points into fixed positions. Since eye inner corners and the philtrum are stable under different expressions, these points have been chosen as reference points. Shown in Figure 2, the reference points number 0 and 8 are two inner eye corners. The last reference point, the philtrum, can be computed via the top point of outer lip and two nostrils (point 51 and 41, 42 on the ASM model, respectively), as follows:

$$p_{philtrum} = \frac{\frac{p_{41}+p_{42}}{2} + p_{51}}{2} \quad (1)$$

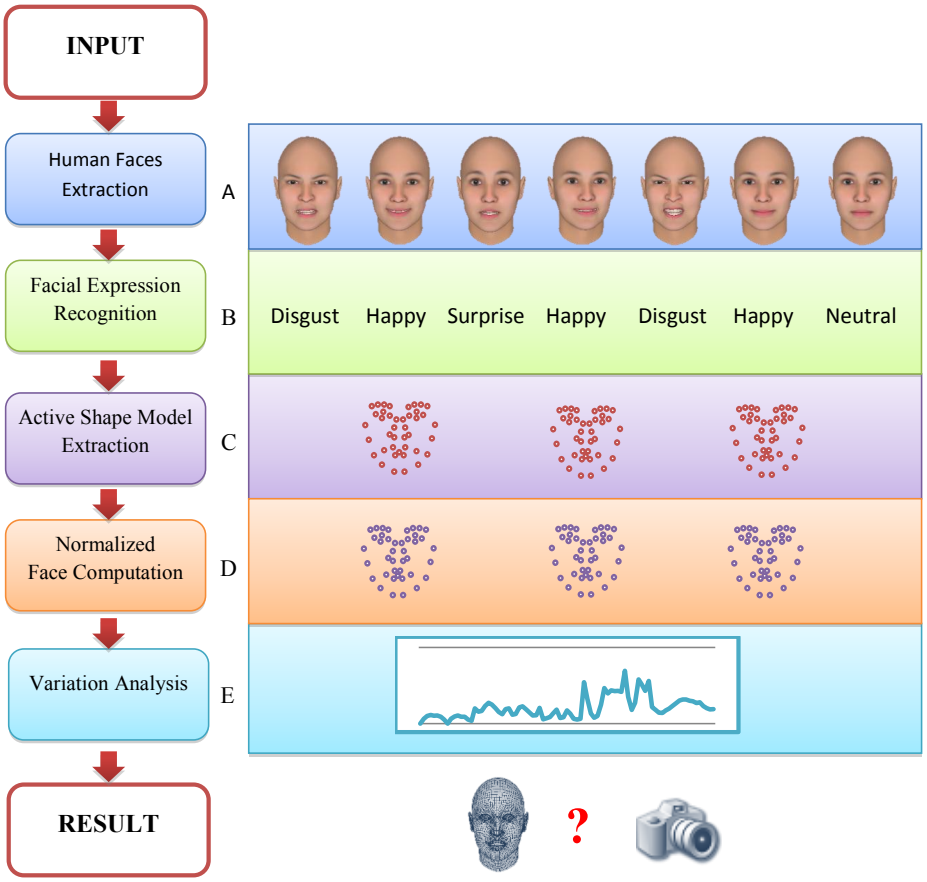



Fig. 1. Schema of the proposed method: A. Human faces are extracted from the video sequence(s). B. Facial expressions are recognized (in the example 3 happy, 2 disgust, 1 surprise and 1 neutral). C. Faces with the same expression are selected (in this example only happy faces) and their active shape models are extracted. D. The extracted models are normalized. E. Differences on the normalized models are analysed to determine whether the character is CG.

where $p_{41}$, $p_{42}$, and $p_{51}$ are the reference points on the extracted ASM.

After computing the three reference points, each ASM model is normalized by moving $\{p_{41}, p_{42}, p_{philtrum}\}$ into their normalized positions, as follows: (i) rotate the segment $[p_{41}, p_{42}]$ into an horizontal line segment; (ii) shear the philtrum to be on the perpendicular line through the middle point of $[p_{41}, p_{42}]$; and finally (iii) scale the image so that the length of segment $[p_{41}, p_{42}]$ and the distance from $p_{philtrum}$ to $[p_{41}, p_{42}]$ have predefined fixed values (see [18] for more details).

Shown in Figure 3 (b) and (d) are examples of the normalized faces after Face Normalization step. The left images show the normalized faces and the right ones show the normalized reference points.

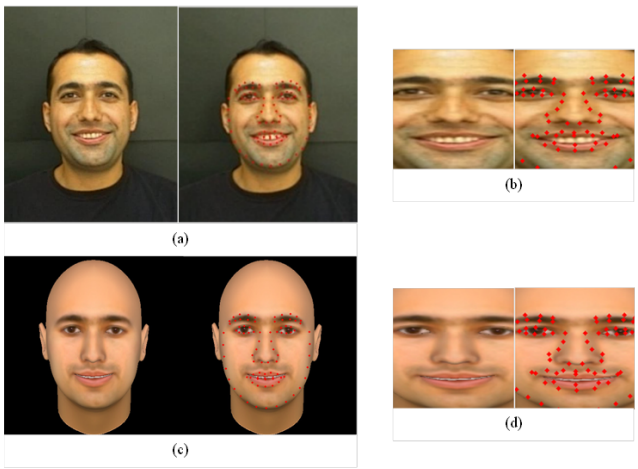



Fig. 3. ASM and normalized ASM: (a) and (c) show a photographic and a computer generated happy face, respectively, and their corresponding ASM points; (b) and (d) show the normalized images of (a) and (c), respectively, and their corresponding normalized points.

### *E. Variation Analysis*

In this step, differences among normalized ASM models are analysed in order to determine if a given character (and therefore the corresponding set of faces) is CG or real. We analyse the differences as described in the following paragraphs.

First, the distance $d_{i,p}$ of each reference point $p$ on a model

$i$ to the average of all points $p$ of all models is calculated as:

$$d_{i,p} = \|(x,y)_{i,p} - (\overline{x},\overline{y})_p\| \quad (2)$$

where $(x,y)_{i,p}$ is the position of the reference point $p$ on the model $i$; $(\overline{x},\overline{y})_p = \frac{1}{N}\sum_{i=1}^{N}(x,y)_{i,p}$, where $N$ is the number of normalized ASM models; and $\|\cdot\|$ is Euclidean distance.

Depending on the facial expression $\xi$ (among six universal expressions), a subset $S_\xi$ of reference points (not all 87 points) are selected for the analysis. For example, with the happy facial expression ($\xi = 1$) only reference points from 0 to 15 and from 48 to 67, which represent the eyes and the mouth, are considered, i.e., $S_1 = \{0, 1, 2..15, 48, 49, .., 67\}$. The subsets are selected based on our experiments and suggestions from EMFACS [9], in which a facial expression is represented by a combination of AUs codes. Shown in Table I are the reference points selected in our method and the correspondent AUs codes from EMFACS. Some explanations of the AUs codes are also listed in Table II. Full codes in EMFACS could be seen in [9].

TABLE I
EXPRESSIONS WITH ACTION UNITS AND CORRESPONDENT ASM POINTS

| $\xi$ | Expression | Action Units (AUs) | Reference Points ($S_\xi$) |
|---|---|---|---|
| 1 | Happiness | 6+12 | $S_1 = \{0-15, 48-67\}$ |
| 2 | Sadness | 1+4+15 | $S_2 = \{0-35, 48-57\}$ |
| 3 | Surprise | 1+2+5B+26 | $S_3 = \{16-35, 48-67\}$ |
| 4 | Fear | 1+2+4+5+20+26 | $S_4 = \{16-35, 48-57\}$ |
| 5 | Anger | 4+5+7+23 | $S_5 = \{0-64\}$ |
| 6 | Disgust | 9+15+16 | $S_6 = \{0-15, 48-67\}$ |

TABLE II
EXAMPLE OF SOME FACIAL ACTIONS [9]

| AU Number | FACS name |
|---|---|
| 1 | Inner Brow Raiser |
| 4 | Brow Lowerer |
| 6 | Cheek Raiser |
| 12 | Lip Corner Puller |
| 15 | Lip Corner Depressor |
| .. | .. |

Two main properties are taken into account in this analysis: mean and variance, calculated as their traditional definitions:

$$\mu_p = \frac{1}{N}\sum_{i=1}^{N} d_{i,p}, \text{ and } \sigma_p = \frac{1}{N}\sum_{i=1}^{N} ||d_{i,p} - \mu_p||^2 \quad (3)$$

where $\mu_p$ and $\sigma_p$ are the mean and variance of all distances $d_{i,p}$ at reference point $p$ over all models.

The given set of models on expression $\xi$ is confirmed to be CG or natural by comparing the *Expression Variation Value* $EVV_\xi$ to the threshold $\tau_\xi$. The value of $EVV_\xi$ is computed as follows:

$$EVV_\xi = \alpha_\xi \frac{\frac{1}{|S_\xi|}\sum_p \mu_p}{\lambda_{1,\xi}} + (1-\alpha_\xi)\frac{max_p\{\sigma_p\}}{\lambda_{2,\xi}} \quad (4)$$

where $\alpha_\xi$ is a weighted constant, $\alpha_\xi \in [0;1]$; $\lambda_{1,\xi}$ and $\lambda_{2,\xi}$ are the normalization values used to normalize the numerators into $[0;1]$. In our experiments $\alpha_\xi$ are set to 0.7 for $\xi = 1, ..., 6$.

$EVV_\xi$ is then compared with $\tau_\xi$, recognizing the character corresponding to the set of faces as CG if $EVV_\xi < \tau_\xi$, natural otherwise.

Shown in Figure 4 are the mean values, corresponding to all 87 ASM points, for the sadness expression ($\xi = 2$) analysed on the two set of images shown in Figure 5 (a). The horizontal axis represents $p$, from 1 to 87, while the vertical axis shows the value of $\mu_p$. Since the facial expression is sadness ($\xi = 2$), only the values from $\mu_0$ to $\mu_{35}$ and from $\mu_{48}$ to $\mu_{57}$ are considered (see the selected reference points in Table I). In this example, the *Expression Variation Value* $EVV_2$ of the CG face is 0.35 comparing to 0.74 of the natural one ($\tau_2 = 0.6$).

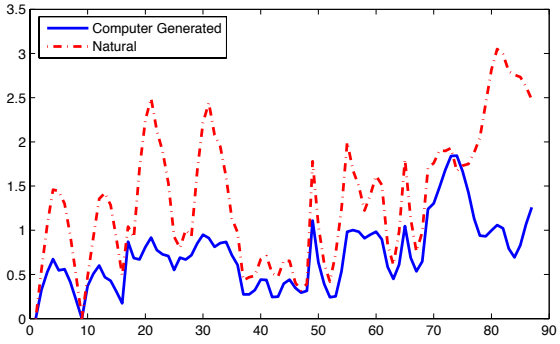



Fig. 4. Example of differences on the mean of ASM points between CG and photographic sad faces of Figure 5 (b).

Values of the thresholds $\tau_{\xi(\xi=1..6)}$ are manually set based on experiments, with the goal of keeping the miss classification as small as possible.

## III. EXPERIMENTAL RESULTS

In our experiments, we use two public datasets:

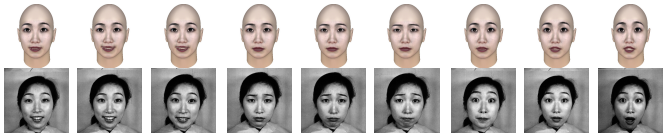
- Boğaziçi University Head Motion Analysis Project Database (BUHMAP-DB) [19], which contains 440 videos of 11 people (6 female, 5 male) performing 5 repetitions on 8 different gestures. We selected the happiness and sadness from this database, since the other six gestures are not related to our topic. Finally, we have 110 videos from this dataset. Each video lasts about 1 - 2 seconds.
- The Japanese Female Facial Expression (JAFFE) Database [20], which contains 213 images of 7 facial expressions posed by 10 Japanese female models.

The first experiment is performed on happiness and sadness expressions from BUHMAP-DB videos. Starting from the 11 people of BUHMAP-DB, we created 11 CG characters by using FaceGen [21] and morphed all of them into both happy and sad faces. FaceGen is a powerful tool which can be used in building complex face structures from one to three images. In our case, we pass a 'neutral' image to FaceGen in order to

build the face structure, then we use Morph options to generate happiness and sadness expressions on the new generated face. Thus, we obtained 110 sets of happy and sad faces, where each model has 5 sets corresponding to happiness and 5 sets corresponding to sadness. Shown in Figure 5 are two examples of the CG versions and the original faces from BUHMAP-DB.

(a) (b)

Fig. 5. Examples of (a) happy, and (b) sad faces from BUHMAP-DB and the corresponding CG faces generated via FaceGen.



(a) (b) (c)

Fig. 6. Examples of (a) happy, (b) sad, and (c) surprised faces from JAFFE and the corresponding CG faces generated via FaceGen.

The goal of this experiment is to analyse the differences from CG models with the natural faces in order to confirm the idea of the proposed method. The analysis is performed as follows: for each video sequence 10 frames are uniformly extracted and similarly for each CG model 10 images are selected. Then, the sets of images are analysed and the corresponding *Expression Variation Values* computed as described in Section II-E. In this case since the expressions are already known, we implement the method from step C. In this step, we use Microsoft Face SDK [22] to extract the ASM models. Finally, we apply step D and E to get the results.

Shown in Figure 7 are $EVV_1$ values computed on the 55 sets of CG and the 55 sets of natural happy faces. These values are well separated between CG and natural. There is only one miss classification using the threshold $\tau_1 = 0.45$. The accuracy, therefore, is 99% (equals 109/110).

On sadness expression, the result is even better, with 100% of accuracy using the threshold $\tau_2 = 0.6$. The $EVV_2$ values for CG and natural characters are perfectly separated, as shown in Figure 8.

Our second experiment is performed on the JAFFE database, which contains all six expressions. Also in this case we used FaceGen [21] to create the CG models reproducing the JAFFE models (see Figure 6 for some examples). For each model in this database, we reproduced all 6 expressions. Therefore, we perform the second test on 120 sets of images, 60 sets of CG and 60 sets of JAFEE real faces. The complete proposed approach described in Section II is applied as a classification approach on these sets.

Shown in Figure 9 is the average $EVV_\xi$ for each expression ($\xi = 1, ..., 6$). The inner blue boundary represents the $EVV_\xi$ computed from CG sets of images, and the outer red boundary represents the natural $EVV_\xi$. Results show that CG and natural *Expression Variation Values* can be differentiated by using and comparing with a set of thresholds $\tau_\xi$, visualized by the green boundary. The classification performance of this experiment is in average 96.67%. Details for each expression are reported in the confusion matrices, Table III.

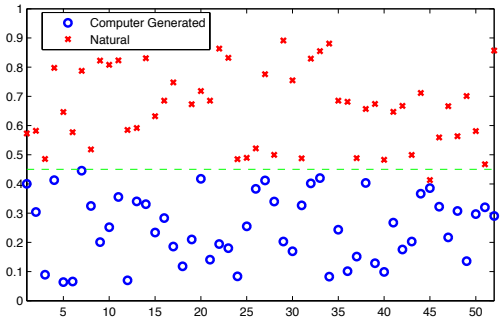



Fig. 7. Facial Expression Values computed on happiness expression. The threshold value $\tau_1$ is 0.45. The separation between CG and natural $EVV_1$ is clearly visualized with only one miss classification.

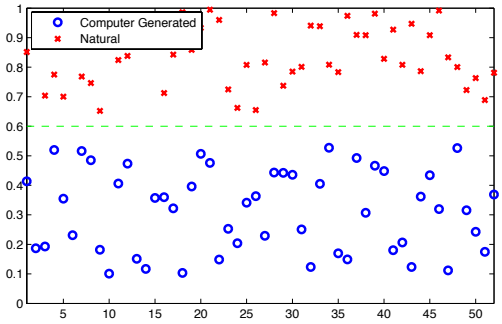



Fig. 8. Facial Expression Values computed on sadness expression. The threshold value $\tau_2$ is 0.6. CG and natural $EVV_2$ are clearly separated.

TABLE III
CONFUSION MATRICES ON CG AND NATURAL FACES, COMPUTED ON JAFFE DATABASE.

| $\xi$ | Expression | | CG | Natural |
|---|---|---|---|---|
| 1 | Happiness | CG | **100%** | 0% |
| | | Natural | 0% | **100%** |
| 2 | Sadness | CG | **100%** | 0% |
| | | Natural | 0% | **100%** |
| 3 | Surprise | CG | **100%** | 0% |
| | | Natural | 0% | **100%** |
| 4 | Fear | CG | **90%** | 10% |
| | | Natural | 0% | **100%** |
| 5 | Anger | CG | **100%** | 0% |
| | | Natural | 0% | **100%** |
| 6 | Disgust | CG | **80%** | 20% |
| | | Natural | 10% | **90%** |

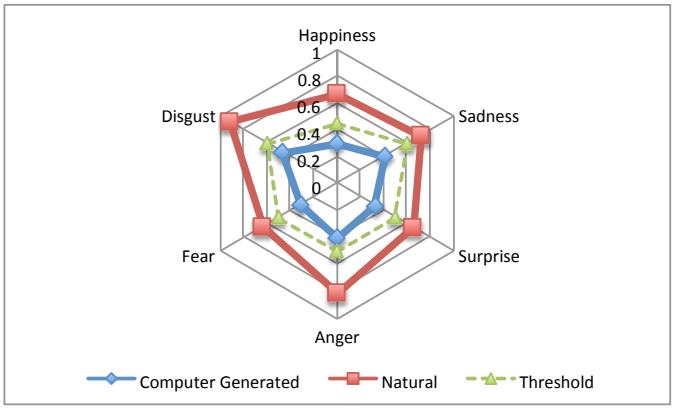



Fig. 9. Average of *Expression Variation Values* analysed for all expressions. CG and natural $EVV_\xi$ are separated for all $\xi = 1, ..., 6$.

The last experiment is performed by comparing *Star Trek Aurora*[2] movie, a fully-animated product, against *Star Trek Odyssey*, a live action movie from *Star Trek: Hidden Frontier* series[3]. In Star Trek Aurora, two graphics applications, namely Poser and Cinema 4D, are used to create the entire 3D world and characters. We extracted 4 female characters in each movie and selected frames that contain happy expression of those characters. Happy faces are then confirmed by using Rosa application [16]. Some examples of two characters in happy emotion are shown in Figure 10. Finally, $EVV$s are computed and compared. Using the same threshold as in the first experiment ($\tau_1 = 0.45$), all $EVV_1$ calculated for the 4 characters of Star Trek Aurora are smaller than $\tau_1$ while all of the $EVV_1$ from Star Trek Odyssey are over $\tau_1$, i.e., the CG characters can be recognized and separated from the natural ones.

(a)

(b)

Fig. 10. Examples of happy faces extracted from (a) Star Trek Aurora, (b) Star Trek Odyssey.

## IV. Conclusions

In this study, we introduced a novel problem about differentiating between CG and natural human faces in video sequences and we presented a method that allows distinguishing CG characters based on facial expression analysis. Indeed, results show that CG persons usually present smaller differences in face shape changing among the same expression, in comparison with real persons. Although experimental results are performed just on small datasets, we proved that the method can be effective. Further work will be devoted to automatic selection of thresholds and exploitation of transitional parameters of faces.

[2]http://auroratrek.com

[3]http://www.hiddenfrontier.com

## References


[1] S. Lyu and H. Farid, "How realistic is photorealistic?" *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 845–850, 2005.

[2] Y. Wang and P. Moulin, "On discrimination between photorealistic and photographic images," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2006, pp. II.161–II.164.

[3] T. T. Ng, S. F. Chang, J. Hsu, L. Xie, and M. P. Tsui, "Physics-motivated features for distinguishing photographic images and computer graphics," in *ACM Multimedia*, 2005, pp. 239–248.

[4] N. Khanna, G. T. C. Chiu, J. P. Allebach, and E. J. Delp, "Forensic techniques for classifying scanner, computer generated and digital camera images," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 1653–1656.

[5] V. Conotter and L. Cordin, "Detecting photographic and computer generated composites," in *SPIE Symposium on Electronic Imaging*, 2011.

[6] D.-T. Dang-Nguyen, G. Boato, and F. G. B. DeNatale, "Discrimination between computer generated and natural human faces based on asymmetry information," in *European Signal Processing Conference (EUSIPCO)*, Bucharest, 2012.

[7] A. Tinwell, M. Grimshaw, D. A. Nabi, and A. Williams, "Facial expression of emotion and perception of the Uncanny Valley in virtual characters," *Computers in Human Behavior*, vol. 27, no. 2, pp. 741–749, Mar. 2011.

[8] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

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

[10] P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System (FACS): Manual*. Salt Lake City (USA): A Human Face, 2002.

[11] ISO, *ISO/IEC 14496-2:1999: Information technology — Coding of audio-visual objects — Part 2:Visual*, 1999.

[12] M. Mori, "Bukimi no tani (the uncanny valley)," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.

[13] K. F. MacDorman, R. D. Green, C.-C. Ho, and C. T. Koch, "Too real for comfort? Uncanny responses to computer generated faces," *Computers in Human Behavior*, vol. 25, no. 3, pp. 695–710, May 2009.

[14] P. A. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[15] ——, "Rapid object detection using a boosted cascade of simple features," in *CVPR*, 2001, pp. 511–518.

[16] L. Rosa, "EigenExpressions for Facial Expression Recognition," http://www.advancedsourcecode.com/facialexpression.asp, 2007.

[17] L. Liang, R. Xiao, F. Wen, and J. Sun, "Face alignment via component-based discriminative search," in *Proceedings of the 10th European Conference on Computer Vision: Part II*, ser. ECCV '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 72–85.

[18] Y. Liu, K. L. Schidt, J. F. Cohn, and S. Mitra, "Facial asymmetry quantification for expression invariant human identification," *Computer Vision and Image Understanding Journal*, vol. 91, no. 1/2, pp. 138–159, 2003.

[19] O. Aran, I. Ari, M. A. Güvensan, H. Haberdar, Z. Kurt, H. . Türkmen, A. Uyar, and L. Akarun, "A database of non-manual signs in Turkish sign language," in *Signal Processing and Communications Applications (SIU2007)*, Eskişehir, 2007.

[20] M. K. J. G. Michael J. Lyons, Shigeru Akamatsu, "Coding facial expressions with gabor wavelets," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 200–205.

[21] "FaceGen Modeller from Singular Inversions," http://www.facegen.com, 2004.

[22] "Microsoft Research Face SDK," http://research.microsoft.com/en-us/projects/facesdk/, May, 2012.