DISCRIMINATION BETWEEN COMPUTER GENERATED AND NATURAL HUMAN FACES BASED ON ASYMMETRY INFORMATION

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

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

ABSTRACT

The recent development of information and communication technology has made computer software able to create highly realistic multimedia contents that can be, for human, impossible to distinguish from the natural ones. This fact leads to the need for tools and techniques that can reliably discriminate between natural and computer generated multimedia data in forensics applications. In this paper, we focus on the specific class of images containing faces, since we consider critical to be able to discriminate between photographic faces and the photorealistic ones. To this aim, we present a new geometric-based approach relying on face asymmetry information. Experimental results show that asymmetry information could be used as a hint to tackle this problem without requiring classification tools and training or combined with state-of-the-art approaches to improve their performances.

Index Terms— Digital Image Forensics, Computer Generated Multimedia Content.

1. INTRODUCTION

Nowadays, multimedia tools are able to create and manipulate multimedia data making them resemble the characteristics of the real world. Furthermore, with the advent of computer graphics technologies, the generation of realistic computer media data has become feasible also for non-expert users. Using these techniques, non-existent objects or scenes can be generated and usually it is very difficult for a human being and for existing methodologies assessing their provenance and authenticity with sufficient confidence. For example, the quality of graphics in the video game Pro Evolution Soccer 2012[1] is highly realistic, thanks to modern computer graphics. Furthermore, in the United States Supreme Court ruling, pornographic photographs depicting an actual child is prohibited, while computer generated child pornography is protected speech. Therefore, it is crucial to develop tools and techniques that can differentiate between natural and computer generated multimedia content in an accurate and reliable way.

The first approach to this problem was introduced in 2005 by Lyu and Farid [1]. In this study, the authors use a statistical model on 216-dimensional feature vectors calculated from the first four order statistics of the wavelet decomposition. In a similar way, Wang and Moulin [2] use a statistical model with only 144-dimensional feature vectors achieving slightly better results with respect to [1]. Based on the estimation of the noise pattern of the devices, in 2008, Khanna et al. presented in [3] a method for discriminating between scanned, non-scanned, and computer generated images. In this study, the basic idea is analyzing noises of the scanner from row to row and column to column, and then combining them with the noise of the camera, calculated as difference between the de-noised image and the input one. In 2011, Conotter and Cordin in [4] developed an hybrid method, which not only exploits the higher-order statistics of [1] but also uses the information from the image noise pattern (36-dimensional feature vectors calculated from the PRNU [5] and used also for source identification [6]). A geometric-based approach was proposed by Ng et al. in [7] to model physical differences between computer generated and photographic images. Using 192-dimensional feature vectors built from local patch statistic, fractal geometry, gradient on surface, quadratic geometry, and Beltrami flow, they uncovered such physical differences.

In this paper, we present a novel approach on a specific target: human faces. People are, in many cases, a crucial target for forgeries, both in doctored images and computer generated ones. Therefore, we consider critical to be able to distinguish between computer generated and photographic faces. Our main idea is to exploit face asymmetry information to develop a geometric-based method which can be used without requiring classification tools and training (see [8] for a similar idea on text manipulation) or combined with existing approaches to improve their performances, as described in the following sections. It is worth mentioning, that very recently also Farid and Bravo focused on a connected problem. Indeed, they presented in [9] results of a perceptual test done with 436 people about discrimination between computer generated and photographic faces. Some interesting results are given in their work, and some of them could be used as sug-

[1]http://www.konami.com/games/pes2012

gestions for novel ways to approach this issue.

2. PROPOSED APPROACH

To the best of our knowledge, when creating synthetic human faces, designers, in most cases, just make a haft of a face and then duplicate it to create the other one. Then, they often apply post processing to achieve photorealistic results but usually not modifying the geometry of the model. Hence, if a given face present a high symmetric structure, this could be considered as a hint that it is generated via computer. On the other hand, although human faces are symmetric, there does not exist a perfectly symmetrical face, as confirmed by Penton-Voak et al. in [10]. The combination of such two hints allow us to make the following assumption: the more asymmetric a human face, the lower its probability to be computer generated. Based on this assumption, we have developed a method to compute asymmetry information and thus discriminate between computer generated and photographic human faces.

Our method contains three main steps as detailed in Figure 1: shape normalization, illumination normalization and asymmetry estimation. First, in shape normalization step, the input image is transformed into the 'standard' shape, i.e. is normalized into the same coordinate system for every face, in order to make the measurements comparable. Then, in illumination normalization step, unexpected shadows, which could affect the accuracy of the measurements, are removed from the normalized face. Asymmetry measurements which are stable under different face expressions are then calculated in asymmetry estimation step. Finally, based on these measurements, we assign to the given face a probability whether it is computer generated or not.

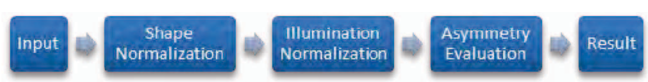

Fig. 1. Schema of the proposed method

An example of the process is shown in Figure 2, where a) represents the input image, b) the normalized face, and e) the result after illumination normalization. It is worth noticing, that the proposed method achieves an accurate discrimination result, and is stable under different expressions, different lighting conditions, and rotation variations.

2.1. Shape normalization

We apply the traditional approach from [11] to normalize a shape of a face in order to have a common coordinate system. This normalization is not only making the measurements easier, but allows to combine them with other facial features (e.g., EigenFace or Fisher Face). Figure 2 b) shows an example of this step applied to Figure 2 a).

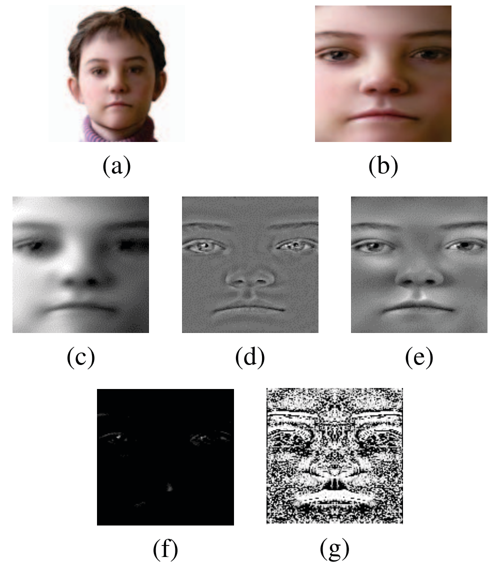

Fig. 2. Face asymmetry estimation: a) input photo; b) normalized photo; c, d) components of illumination normalization step; e) result after illumination normalization; f, g) subresults of asymmetry evaluation step.

2.2. Illumination normalization

Illumination causes most challenging problems in face analysis. Asymmetry measure is calculated based on the intensity of the face image, thus, the shadows, which are usually quantized as low value regions, play an important role. However, what we need is the information of the face structure, without any effects from shadows or unexpected lighting illumination. Hence, illumination normalization is required in order to enhance the accuracy of the asymmetry measurements.

We apply the approach presented by Xie et al. in [12]. The basic idea is to use the albedo of large scale skin and background, denoted as $R_l(x, y)$ to split the face image I into large-scale and small-scale components.

Based on Lambertian theory, we have:

$$I(x, y) = R(x, y)L(x, y) \tag{1}$$

where R is the albedo of the face and L is the illumination. Estimating this information consists of as an ill-posed problem, hence Xie et al. in [12] apply a transformation to overcome this issue as follows:

$$\begin{aligned} I(x, y) &= R(x, y)L(x, y) \\ &= \left(\frac{R(x, y)}{R_l(x, y))}\right)(R_l(x, y)L(x, y)) \\ &= \rho(x, y)S(x, y) \end{aligned} \tag{2}$$

where ρ contains the intrinsic structure of a face image, and S contains the extrinsic illumination and the shadows, as well as the facial structure. ρ and S are called small-scale features and large-scale features, respectively.

In order to split the image into large-scale and small-scale, the Logarithm Total Variance (LTV) estimation is used. This estimation is introduced in [13] and is the best method to extract illumination-invariant features so far. After splitting the image I into ρ and S, smoothing filter, which is also introduced in [12], are required to be applied on ρ in order to remove unexpected effects from the decomposition in (2).

An example of this step is shown in Figure 2 where c) and d) represent the large-scale and small scale components, respectively, after applying LTV on the image b). The illumination normalized result is shown in e).

2.3. Asymmetry Evaluation

In order to estimate asymmetry, we use the measure introduced by Liu et al. in [11], which is invariant to face expressions. Let us denote the density of the image with I, and the vertically reflected of I with I'. The edges of the densities I and I' are extracted and stored in I_e and I'_e, respectively. Two measurements for the asymmetry are introduced as follows:

Density Difference (D-Face):

$$d(x, y) = \|I(x, y) - I'(x, y)\| \tag{3}$$

Edge orientation Similarity (S-Face):

$$s(x, y) = \cos\left(\theta_{I_e(x,y),I'_e(x,y)}\right) \tag{4}$$

where $\theta_{I_e(x,y),I'_e(x,y)}$ is the angle between the two edge orientations of images I_e and I'_e, at position (x, y). Figure 2 (e) shows the estimated frontal face resulting from the illumination normalization step. In Figure 2 (f) and (g) the D-Face and S-Face are shown, respectively. Based on these measurements, we can estimate the asymmetry of a given face photo since the higher the value of D-Face, the more asymmetric is the face, and the higher the value of S-face, the more symmetric the face. The total difference of D-Face and total dissimilarity of S-Face are calculated as follows:

$$D = \frac{\sum_{x,y\in\Omega} d(x, y)}{\eta_1}; \tag{5}$$

$$S = 1 - \frac{\sum_{x,y\in\Omega} s(x, y)}{\eta_2} \tag{6}$$

where η_1, and η_2 are the normalized thresholds, which scale D and S into (0; 1), and Ω is the estimated region. Since our images are normalized to the fixed size 128×128, and Ω is fixed as in [11], both thresholds η_1, and η_2 are fixed.

Finally, we assign to image I an exponential probability to be computer generated, as follows:

$$P = \lambda e^{-\lambda\sqrt{D^2+S^2}} \tag{7}$$

where λ is a constant (we use $\lambda = 1.0$). If P is over a threshold τ, I is classified as a computer generated human face (we use $\tau = 0.5$).

3. EXPERIMENTAL RESULTS

We have collected the computer generated images from the *Society of Digital Artist*[2] and downloaded football player face images from the database of *Faces for Pro Evolution Soccer 2012*[3]. All of the computer generated images are confirmed that they are purely created by computer. For the natural images, real people and football players images were collected from various sources on the internet. We have also collected other images from Karolinska[4] face database, which contains hundreds of frontal face images. We have created two datasets: Dataset 1 contains very realistic images, which are almost undetectable by human; Dataset 2 contains more images, related to real situations (see Table 1 for details). Examples of images collected for both classes of the two datasets are also shown in Figure 3.

Table 1. Number of images per dataset

	Computer Generated	Photographics
Dataset 1	40	40
Dataset 2	200	200

In our first experiment, we analyze the proposed approach using only asymmetry information achieving 67.5% of accuracy on Dataset 1 and 89.25% on Dataset 2. Shown in Figures 4 and 5 are the ROC chart of False Positive and True Positive rates on Dataset 1 and 2, respectively, while in Tables 2 and 3 corresponding confusion matrices are reported. These results show that geometry information, in this case the asymmetry of human faces, can be effectively used to discriminate computer generated from the natural faces.

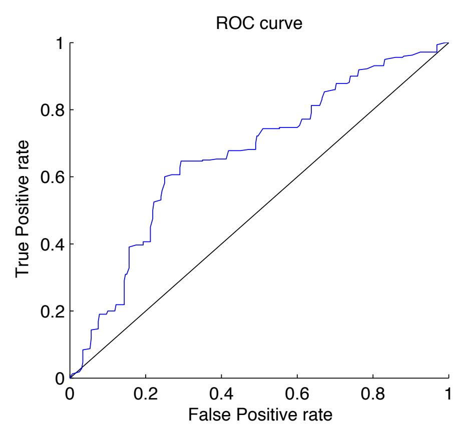

Fig. 4. ROC curve of the proposed method on Dataset 1.

In the second experiment, we compare our method with three state-of-the-art approaches, namely, [1], [3], and [4].

[2]http://CGSociety.org

[3]http://www.pesfaces.co.uk/

[4]http://webscript.princeton.edu/˜tlab/databases/database-2-karolinska-dataset/

Fig. 3. Examples: (a) CG faces of Dataset 1, downloaded from CGSociety.org; (b) real faces of Dataset 1, downloaded on 11-2011 *From left to right: media.celebrity-pictures.ca/Celebrities/Dasha-Astefieva/Dasha-Astefieva-i164695.jpg, picturrs.com/files/funzug/imgs/celebrities/celeb_spsmakeover_11.jpg, img2.timeinc.net/people/i/2009/database/michelle-obama/michelle_obama300.jpg, and www.sailing.org/images/content/committee/Fiona_Barron-passport.jpg*; (c) CG faces of Dataset 2, downloaded from PES 2012 faces database; (d) real faces of Dataset 2, downloaded on 01-2012 *From left to right: en.last-video.com/wp-content/uploads/2010/07/Iniesta-goal.png, www.fcbarcelona.com/web/downloads/fotos/retrats/temp11-12/MESSI.jpg, www.football-rumours.com/images/cristianoronaldo.jpg, and juventus.theoffside.com/files/2009/07/buffon.jpg*.

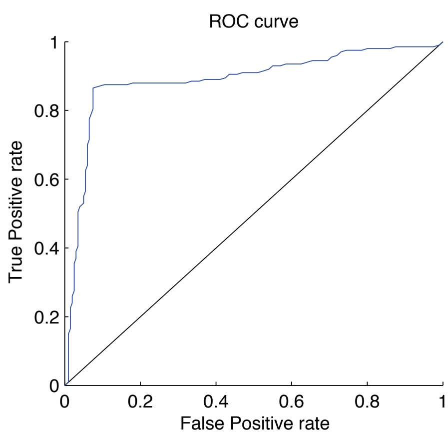

Fig. 5. ROC curve of the proposed method on Dataset 2.

Here, we consider asymmetry information as a feature, and then use Support Vector Machine (SVM) for training and solving the binary classification problem. Shown in Figure 6 are results of comparing these methods using leave-one-out (LLO) cross validation method. It can be noticed that on the challenging Dataset 1, the proposed approach achieves the best performances, while on Dataset 2, there is not much difference among all approaches.

In the last experiment, we use asymmetry information as an additional feature to [1], [3], and [4] and compare results using SVM binary classification (LLO validation). Figure 7

Table 2. Confusion matrix on Dataset 1.

	Computer Generated	Photographics
CG	0.75	0.25
Photographics	0.4	0.6

Table 3. Confusion matrix on Dataset 2.

	Computer Generated	Photographics
CG	0.92	0.08
Photographics	0.135	0.865

and 8 show results on Dataset 1 and Dataset 2, respectively. Performances of state-of-the-art approaches increase on average by 16.25% on the more challenging Dataset 1 when fusing their features with the proposed asymmetry features.

4. CONCLUSIONS

In this study, we presented a novel way to tackle the problem of differentiating between computer generated and photographic human faces. Based on the estimation of the face asymmetry, a given photo is classified as computer generated or not. The results show that our approach can be used as a stand alone method or in combination with other information to improve state-of-the-art techniques. However, some issues still remain open: this approach only works with frontal faces

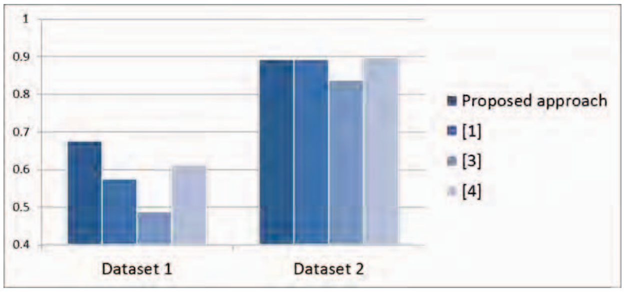

Fig. 6. Comparision of results of the proposed approach with [1], [3], and [4] on both Dataset 1 and Dataset 2.

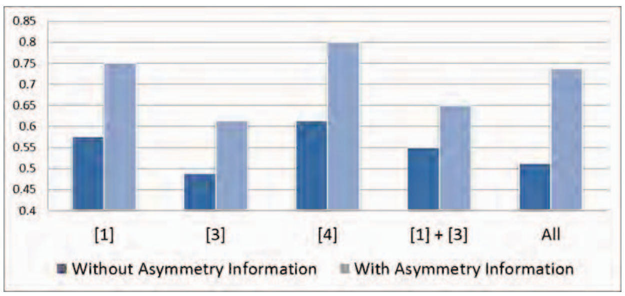

Fig. 7. Improvement on Dataset 1 by adding asymmetry information into existing approaches.

so far, since the normalization step can only normalize rotated faces, but not the turned ones; moreover it is very sensitive to the normalization step. Further works will cope with these problems.

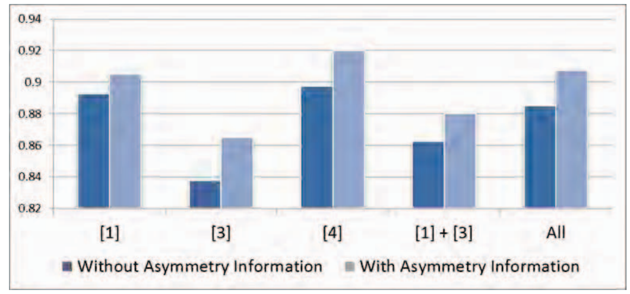

Fig. 8. Improvement on Dataset 2 by adding asymmetry information into existing approaches.

5. ACKNOWLEDGEMENT

The authors are grateful to Prof. Hany Farid of the Dartmouth College for enlightening conversations and fruitful discussions.

6. 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] N. Khanna et al., "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.

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

[5] Jan Luk, Jessica Fridrich, and Miroslav Goljan, "Detecting digital image forgeries using sensor pattern noise," in *SPIE Symposium on Electronic Imaging*, 2006.

[6] V. Conotter and G. Boato, "Analysis of sensor fingerprint for source camera identification," *Electronics Letters*, vol. 47, no. 25, pp. 1366–1367, 2011.

[7] T. T. Ng et al., "Physics-motivated features for distinguishing photographic images and computer graphics," in *ACM International Conference on Multimedia*, 2005, pp. 239–248.

[8] V. Conotter, G. Boato, and H. Farid, "Detecting photo manipulations on signs and billboards," in *IEEE International Conference on Image Processing*, 2010.

[9] H. Farid and M.J. Bravo, "Perceptual discrimination of computer generated and photographic faces," *Digital Investigation*, vol. 8, pp. 226–235, 2012.

[10] I. S. Penton-Voak et al., "Symmetry, sexual dimorphism in facial proportions and male facial attractiveness," *Biological Sciences / The Royal Society*, vol. 268, no. 1476, pp. 1617–1623, 2001.

[11] Y. Liu et al., "Facial asymmetry quantification for expression invariant human identification," *Computer Vision and Image Understanding Journal*, vol. 91, pp. 138–159, 2003.

[12] X. Xie et al., "Normalization of face illumination based on large-and small-scale features," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1807–21, 2011.

[13] T. Chen et al., "Total variation models for variable lighting face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1519–1524, 2006.