A MULTI-TEXTURE APPROACH FOR ESTIMATING IRIS POSITIONS IN THE EYE USING $2.5D$ ACTIVE APPEARANCE MODELS

Hanan Salam
SUPELEC/IETR
Team SCEE, France

Nicolas Stoiber
Dynamixyz
Rennes, France

Renaud Séguier
SUPELEC/IETR
Team SCEE, France

ABSTRACT

This paper describes a new approach for the detection of the iris center. Starting from a learning base that only contains people in frontal view and looking in front of them, our model (based on $2.5D$ Active Appearance Models (AAM)) is capable of capturing the iris movements for both people in frontal view and with different head poses. We merge an iris model and a local eye model where holes are put in the place of the white-iris region. The iris texture slides under the eye hole permitting to synthesize and thus analyze any gaze direction. We propose a multi-objective optimization technique to deal with large head poses. We compared our method to a $2.5D$ AAM trained on faces with different gaze directions and showed that our proposition outperforms it in robustness and accuracy of detection specifically when head pose varies and with subjects wearing eyeglasses.

Index Terms— Active appearance model, iris tracking, gaze detection

1. INTRODUCTION

Although the eye occupies a small region of the face, it carries a significant amount of information. One important information is iris location which is essential for determining the gaze direction. Unfortunately, the importance of iris information comes in parallel with the difficulty of its extraction. This difficulty can be summarized by the following: The iris has a small texture with respect to that of the skin surface, it is occluded when the eye blinks and when the person commits large head movements and even by eyeglasses and hair. Moreover, the variability in its color, scale and location makes it more difficult to be extracted.

Many methods were proposed to detect the gaze direction. We classify them into three categories:

i) IR (Infrared) based methods [1] have the drawback of being strongly dependent of the brightness of the pupils which can be influenced by the eye closure and occlusion, external illumination and the distance of the user from the camera.

ii) Methods based on deformable templates [2, 3] need the definition of an adequate set of initial parameters for the template that should be initialized close to the eye. Furthermore, they usually record a failure with big head poses. Some methods use ellipse fitting [4]. These succeed at finding the location of the iris or the pupil in condition that high resolution images are provided. Another drawback is their incapability to cope with the different states of closure of the eyelids.

iii) Synthesis based approaches are more robust. They rely on the synthesis of an image using a model and computing the position of the iris by comparing the synthesized image to the real one. The interest of such methods is in their ability to model the appearance and the shape of the eye simultaneously in contrary to template matching and ellipse fitting methods that only model its shape. To avoid manual design of the eye model, Active Appearance Models (AAM) [5] are adequate. Building an AAM for iris localization requires training a set of hand labeled faces with different gaze directions. Thus, the appearance of the eye is learned simultaneously with that of the face [6, 7]. In order to arrive at a reliable AAM for iris localization, a large training set should be integrated in order to scan all the joint variation in pose (if the AAM is $2D$) and in gaze direction.

We propose to merge an iris AAM and a local eye model where holes are put in the place of the white-iris region. The iris texture slides under the eye hole permitting to synthesize new gaze directions. We deal with large head poses by implementing a multi-objective optimization: At each frame, the system takes into consideration the eye that appears the most in the camera to robustly detect the gaze. Compared to the state of art, the proposed system works with low resolution images, it does not constrain the user with special requirements (IR illumination, hardware equipment...) and it makes use of the appearance and shape of the eye while avoiding explicit manual design of the model through the use of AAMs. With respect to classical AAM, it has the advantage of restricting the learning database of AAM to people in frontal view and looking in front of them.We can state the contributions of this paper as :

- Combination of two AAMs (eye skin and iris models) to detect iris location;
- A specific AAM of the iris where its shape and appearance are learned apart from the eye skin;
- Explicit use of the head pose in the algorithm to make the system robust to large head poses;
- Detection of the gaze without the need of including people with different gaze directions in the learning database of AAM.

This paper is organized as follows: In section 2 we describe the global system. In section 2.1, we present our algorithm. Then in section 3, we show some experimental results. Finally, in section 4 we draw our conclusions.

2. SYSTEM OVERVIEW

Let us consider that the person is in front of the screen where a webcam is installed. The first step is the detection of the face orientation for which a 2.5D global AAM model [8] is applied. Depending on the detected face orientation, the left and right eyes are unevenly represented in the webcam image. We thus propose to analyze gaze direction using a multi-objective optimization: The contribution of each eye to the final gaze direction is weighted depending on the detected face orientation.

Thanks to Immemo (french ANR project) and IMMERSIVITE (Britanny Region PME project) for funding.

In the following section, we describe how the iris location is calculated for one eye using the proposed multi-texture AAM. Then in section 2.2 we present the multi-objective optimization.

2.1. Multi-texture AAM for iris localization

The basic idea of multi-texture AAM (MT-AAM) is that the interior of the eye is considered as a separate texture from that of the face. By permitting this texture to slide under the skin surface, we succeed at synthesizing any gaze direction and consequently at parameterizing the iris motion. This requires the fusion of 2 AAM, one for the iris texture and one for the surrounding skin (where a hole is put inside the eye).

Local eye skin AAM – This model is built using 22 landmarks that describe the whole eye area including the eyebrows and the texture surrounding the eye. Figure 1(a) is an illustration of the mean texture of our model showing the hole inside the right eye with the annotations to obtain this model.

Iris AAM – Figure 1(b) is an illustration of the mean texture of the iris model with the corresponding annotations. For training these iris images we use a model of 13 landmarks of which 8 describe the elliptical shape of the iris and 1 describes the approximative position of its center; to learn the white texture around the iris, 4 additional landmarks forming a rectangular shape around the iris are placed.

Fusion of the 2 AAM (models) – To search for the iris position in the eye, we combine the eye skin model and the iris texture model. We find the optimal parameters for the eye skin (using the eye skin model) in a prior step. We then use the found parameters to reconstruct the image describing the eye skin. The iris model slides under it with the pose vector T^{iris} describing the iris position in the eye.

$$T^{iris} = \left[S^{iris} t_x^{iris} t_y^{iris} \right] \tag{1}$$

where S^{iris} is the scale of the iris, t_x^{iris} and t_y^{iris} are the horizontal and vertical translation parameters describing the iris points position from the mid-point of the eye. Both models, once merged, constitute the final model describing the eye region (cf. figure 1(c)). A problem of discontinuity between the two models arises. In order to resolve this, we apply a low pass filter (circular averaging with radius 2) on the skin and white parts while preserving the iris. It smoothes the discontinuity between the eyelid, and also reproduces the shadow effect of the eyelid on the iris.

Algorithm – Steps for building our combined AAM for one eye.

1. Localize the eye using the eye skin model.
2. From the webcam image, extract the texture of the eye (g_i) (figure 1(d)).
3. Using the optimal parameters found by the eye skin model, synthesize the eye skin (g_m^{eye}) (figure 1(a)).
4. Until stop condition do:
   (a) Create the model texture of the iris (g_m^{iris}) based on the pose and the appearance parameters of the iris model (figure 1(b)).
   (b) Merge the two textures g_m^{eye} and g_m^{iris} to obtain the texture g_m (figure 1(c)).
   (c) Apply a circular averaging low pass filter
   (d) Compare it to g_i to get the error E: $E = g_i - g_m$
   (e) Tune the pose and appearance of the iris model.

2.2. Multi-objective optimization

Since normally the two eyes have highly correlated appearance and motion, we use only one pose vector and one appearance vector to describe the pose and appearance of both irises. Technically, it should be sufficient to analyze the iris of one eye to obtain its position and appearance in both eyes. Yet, the information from both eyes can conduct to a more robust system unless when the person commits large head movements around the vertical axis where one of the eyes can be partially or completely occluded. This leads to a multi-objective AAM [9]. In this multi-objective system, a single iris model is merged simultaneously with the left and the right eye skin models. Then the resulting models are overlaid on both the right and the left eyes from the camera to get the left and right errors. These are weighted according to the head orientation and summed to get one global error that is minimized using a Genetic Algorithm (GA). This error becomes:

$$E = \alpha E^{left} + \beta E^{right} \tag{2}$$

where E^{left} and E^{right} are the errors corresponding to the left and right eyes respectively. α and β are the weighting factors. They are functions of the head rotation around the z-axis (R_{yaw}), evaluated just after the face detection, and they both follow a double logistic law:

$$\alpha(R_{yaw}) = \begin{cases} 0.5 \text{ if } -d \leq R_{yaw} \leq d \\ 0 \text{ if } -90 \leq R_{yaw} < -22 \\ 1 \text{ if } 22 \leq R_{yaw} < 90 \\ 0.5(1 + l(1 - \exp \frac{-(R_{yaw} - ld)^2}{\sigma^2})) \text{ else} \end{cases}$$

$$\beta(R_{yaw}) = 1 - \alpha \tag{3}$$

where $l = \text{sign}(R_{yaw})$, σ is the steepness factor and d is the band such that the two functions α and β are equal to 0.5. In this way, the face orientation is taken into account by the relevant information from both eyes.

Concerning the Genetic Algorithm, the initial population is generated randomly between the upper and the lower limits of the parameters with a uniform distribution. The fitness function is the penalized sum of the pixel errors of the left and the right eyes presented in equation 2.2.

Since the iris location and scale is constrained by the size of the eye, we add constraints in order to tighten up the search space of

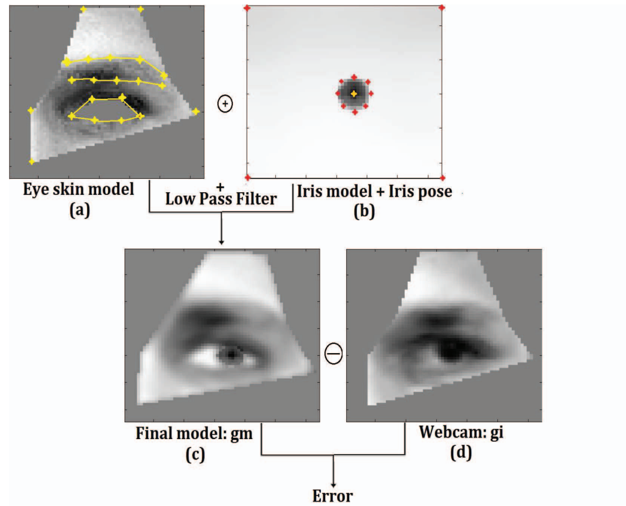


Fig. 1. Merging the eye skin model with the iris one to obtain the final eye model.

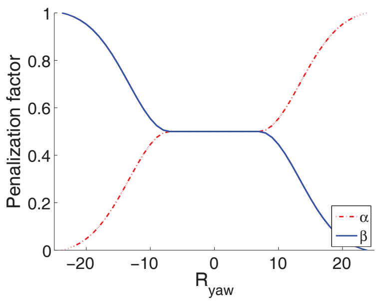
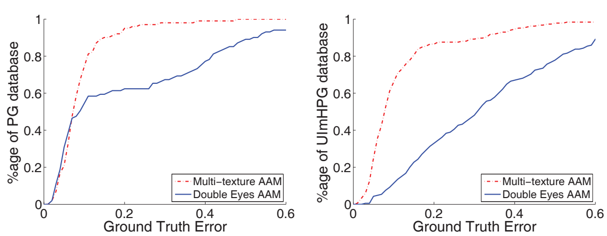


Fig. 2. Double logistic function, d = 7

the GA to a plausible one. These constraints are based on anthropometric averages which give a very tight bound on the size of the iris relative to the size of the eye and on the fact that iris movements have limits imposed on them by the width of the eye. Iris average width is approximated to $1/1.6$ the width of the eye. The scale is varied around an initial scale computed using the width of the iris which is calculated based on the previous relationship ($1/1.6$ the width of the eye) and that of the mean iris. The horizontal and vertical translation parameters can not exceed the half of the eye width and that of the eye height respectively taking the midpoint of the eye corners as the reference point.

3. RESULTS

In order to validate our algorithm we conducted tests on the UIm Head Pose and Gaze (UImHPG) database [10] and the Head Pose - Eye Gaze (HPEG) database [11]. We compare the MT-AAM method with a classical one, the $2.5D$ double eyes AAM. We trained this double eyes AAM (DE-AAM) using a number of faces showing different gaze directions (10 persons from the HPEG with five images per person, one looking forward, one to the extreme left, one to the extreme right and two intermediate gaze directions). This model uses 5 landmarks for each eye and in order to take into consideration the texture surrounding the eyes, 5 landmarks on the bottom of each eyebrow are placed. Concerning the MT-AAM, to find the head orientation, we trained 104 neutral face images of the $3D$ Bosphorous



(a) GTE of HPEG database (b) GTE of UImHPG database

Fig. 3. Multi-texture AAM vs. Double Eyes AAM

database [12]. We trained the local eye skin model using the same learning base as the $2.5D$ double eyes model. To train the iris AAM, we use a group of 23 iris textures starting from the images of [13].

In figure 3 we compare the Ground Truth Error (GTE) versus the percentage of aligned images in the two databases. 3(a) displays results obtained by testing on the learning database (HPEG), while 3(b) displays results obtained by generalizing to new data (UlmHPG database). The GTE is the mean of the distance (Euclidean distance) between ground truth (real location of iris center) marked manually and the iris center given by the gaze detection method, normalized with the interocular distance [14]. We can see how the multi-texture AAM outperforms the DE-AAM. For instance, figure 3(a) shows that for an error less or equal to 15% the multi-texture AAM has detected the right position of the iris on 90% of the images, whereas the DE-AAM shows only 59% at the same error level. Figure 3(b) suggests generalization capabilities for both the $2.5D$ double eyes and the multi-texture models. We have a good detection of 81% for the multi-texture method versus 22% for the DE-AAM method for the same error level of 15%.

In addition, figure 4 shows qualitative results on the 2 databases. On the HPEG database (both training and testing), figure 4(a) shows that the multi-texture method outperforms the classical $2.5D$ double eyes one when it comes to people committing different head poses. The reason behind this is that with the multi-texture AAM we favor the eye that appears the most in the camera using the multi-objective optimization. Concerning people wearing eyeglasses, the learning base of both methods does not contain such subjects. As the figure 4(b) shows on UImHPG database, MT-AAM succeeds to follow the gaze of a person with eye glasses whereas the $2.5D$ DE-AAM method does not. This assures the fact that with our method we are able to restrict the training base where there's no need to include people wearing eyeglasses in order to get reliable results on such subjects. The reason behind this is that excluding the appearance of the interior of the eye for localizing the eyelids makes it easier for our model to cope with other kinds of appearance such as the existence of eyeglasses.

As we see, the classical AAM was not able to capture the gaze of a new subject from a different database. In contrary, the multi-texture AAM was able to do the same task without the need for a database that contains people with different gazes. With the multi-texture AAM, since the model is independent of the appearance of the eyes in the training database, we are able to overcome the problem of generalization.

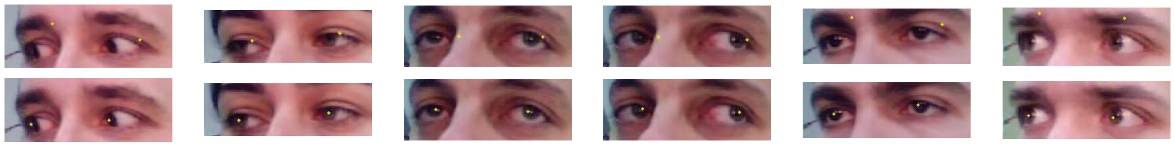
3.1. MT-AAM Performance Discussion

Concerning the complexity of the algorithm, the MT-AAM is built using three models: the global face model, the eye skin model and the iris model (see section 2). So, four steps are needed:

1. Apply the global AAM to get the pose;
2. Apply the left eye skin model to localize the left eye;
3. Apply the right eye skin model to localize the right eye;
4. Search for the iris location by combining the eye skin models with an iris model taking into consideration the head pose from the global AAM model;

On an Intel Xeon-2.5GHZ computer, in real time, the global AAM works at 50 frames/sec. By extrapolation, we can estimate that the MT-AAM will work at $50/4 = 12$ frames/sec which can be accepted for real time applications. This performance can be improved by using one AAM (of course with holes put instead of the eyes) that gives the head pose and localizes the eyelids. The system will then work at 25 frames/sec.

One of the weak points of the MT-AAM is that it is dependent of the results of the eye skin AAM. In other words, if the eye skin AAM diverges, the MT-AAM will not be able to robustly localize the iris. This can be ameliorated by improving the used eye skin model (for example increasing the images in the learning database of the eye skin model to include more variations). Later work will tackle this issue.



(a) Results on the HPEG database: Double eyes model (upper row), multi-texture model (lower row).

(b) Results on UImHPG database (generalization for both models): Double eyes model (upper row), multi-texture model (lower row).

Fig. 4. Comparison between the multi-texture approach (lower row) and the $2.5D$ double eyes AAM approach (upper row).The yellow dots that appear on the images are the outputs of the DE-AAM and the MT-AAM. They signify the iris location obtained by the iris localization method.

4. CONCLUSION

We have presented a new approach for the detection of the iris location in the eye which serves for gaze detection applications. The proposed algorithm acts as a the divide and conquer algorithms where the eye is divided into two parts: the eye skin and the iris-white texture. We localize the eye using an eye skin AAM where holes are put inside the eyes. We then parametrize the iris motion inside the eye by sliding the iris texture under the eye skin model. Using this combined AAM model formulation, any gaze direction can be synthesized and thus detected. Using a multi-objective technique we overcome the problem of large head poses. The experimental results showed that our combined AAM outperforms the classical method of iris detection using AAM when it comes to generalization when the subject commits large head poses, or in the presence of eyeglasses. In future work, we plan to parametrize the eyelids and eyebrows motion where eyelids and eyebrows pose parameters will be equally added to our combined model.

5. REFERENCES

[1] J. P Hansen, J. S Augustin, and H. Skovsgaard, "Gaze interaction from bed," in *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications*, 2011, p. 11.

[2] Qiu Chen, Koji Kotani, Feifei Lee, and Tadahiro Ohmi, "An accurate eye detection method using elliptical separability filter and combined features," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 9, no. 8, August 2009.

[3] H. Wu, Y. Kitagawa, T. Wada, T. Kato, and Q. Chen, "Tracking iris contour with a 3D eye-model for gaze estimation," in *Proceedings of the 8th Asian conference on Computer Vision Part I*, 2007, pp. 688–697.

[4] W. J Ryan, D. L Woodard, A. T Duchowski, and S. T Birchfield, "Adapting starburst for elliptical iris segmentation," in *IEEE second international conference on biometrics : Theory, Applications ans Systems*, Sept. 2008, pp. 1–7.

[5] T. F Cootes, G. J Edwards, and C. J Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681, 2001.

[6] I. Bacivarov, *Advances in the Modelling of Facial Sub-Regions and Facial Expressions using Active Appearance Techniques*, Ph.D. thesis, National University of Ireland, 2009.

[7] P. Ivan, "Active appearance models for gaze estimation," M.S. thesis, Vrije University, Amsterdam, 2007.

[8] A. Sattar, Y. Aidarous, S. Le Gallou, and R. Seguier, "Face alignment by 2.5d active appearance model optimized by simplex," in *International Conference on Computer Vision Systems (ICVS)*, 2007.

[9] A. Sattar and R. Seguier, "Facial feature extraction using hybrid ggenetic-simplex optimization in multi-objective active appearance model," in *Fifth International Conference on Digital Information Management (ICDIM)*, 2010, pp. 152–158.

[10] U. Weidenbacher, G. Layher, P. M Strauss, and H. Neumann, "A comprehensive head pose and gaze database," in *3rd IET International Conference on Intelligent Environments (IE 07)*, pp. 455–458.

[11] S. Asteriadis, D. Soufleros, K. Karpouzis, and S. Kollias, "A natural head pose and eye gaze dataset," in *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*, 2009, p. 1.

[12] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *Proceedings of the First COST 2101 Workshop on Biometrics and Identity Management (BIOD)*, 2008, pp. 47–56.

[13] M. Dobes, L. Machala, P. Tichavskı, and J. Pospisil, "Human eye iris recognition using the mutual information," *Optik-International Journal for Light and Electron Optics*, vol. 115, no. 9, pp. 399–404, 2004.

[14] O. Jesorsky, K.J. Kirchberg, and R.W. Frischholz, "Robust face detection using the hausdorff distance," in *Third International Conference on Audio and Video-based Biometric Person Authentication*, 2001, pp. 90–95.