ESTIMATION OF EYE GAZE DIRECTION ANGLES BASED ON ACTIVE APPEARANCE MODELS

Petros Koutras and Petros Maragos

School of E.C.E., National Technical University of Athens, 15773 Athens, Greece

Email: {pkoutras, maragos}@cs.ntua.gr

ABSTRACT

In this paper we demonstrate efficient methods for continuous estimation of eye gaze angles with application to sign language videos. The difficulty of the task lies on the fact that those videos contain images with low face resolution since they are recorded from distance. First, we proceed to the modeling of face and eyes region by training and fitting Global and Local Active Appearance Models (LAAM). Next, we propose a system for eye gaze estimation based on a machine learning approach. In the first stage of our method, we classify gaze into discrete classes using GMMs that are based either on the parameters of the LAAM, or on HOG descriptors for the eyes region. We also propose a method for computing gaze direction angles from GMM log-likelihoods. We qualitatively and quantitatively evaluate our methods on two sign language databases and compare with a state of the art geometric model of the eye based on LAAM landmarks, which provides an estimate in direction angles. Finally, we further evaluate our framework by getting ground truth data from an eye tracking system Our proposed methods, and especially the GMMs using LAAM parameters, demonstrate high accuracy and robustness even in challenging tasks.

***Index Terms*—** Eye gaze estimation, gaze direction angles, face and eyes modeling, active appearance models, histograms of oriented gradients, gaussian mixture models, eye-tracker estimation.

1. INTRODUCTION

The role of Eye Gaze: Eyes are primarily a human sensor for visual data input. They can also express information and specifically the gaze direction which defines the place where a person focuses his attention. Eyes, as well as the other face parts, participate in the expression of emotions [1]. Specifically, there are some emotions which occur mainly with a change in the eyes region. Generally, people can easily determine the eye gaze direction by rotation of the head and eyes angles. Multiple application fields could benefit from the use of a system for automatic estimation of eye gaze. For example, it could be integrated in car security systems [2] and in computer interaction systems [3] for people with special needs. Eye gaze estimation could be used to design systems for automatic sign language recognition [4, 5, 6] as well. Finally, it could be the basis for the development of computer interfaces [6, 7], which would help people with speech and aural disabilities to interact with computers.

Background: Eye Gaze can be defined in two different ways: as the direction of gaze vector [2, 8, 9] or as a point of interest in a reference plane (such as a monitor), on which our eyes focus [10, 11]. Both these approaches have been widely used, depending on the purpose of application [12]. The main steps and individual problems that arise in the estimation of eye gaze are eyes detection and tracking, head pose estimation and their combination using a model for the calculation of the final gaze. Generally, the eye gaze estimation methods can be classified into intrusive and non-intrusive. The intrusive techniques require the use of additional equipment and thus affect the user behavior. In contrast, non-intrusive methods do not use any type of additional equipment which interacts directly with the user and influence his behavior. These methods use Computer Vision techniques to extract information from a set of images and are quite promising for the development of a widely used gaze estimation system with a relatively low cost [13].

For the eye gaze estimation several types of techniques have been proposed: 3D geometrical models, 2D feature based regression models and appearance based models [14]. The 3D geometrical models, in most cases, use infrared light and achieve an estimation with high accuracy [15, 13]. Nonetheless, typical cameras are also used [2, 8, 12], but they require high-resolution images focused on the eye region, because a small error in the estimation of one parameter can lead to large errors in the estimation of gaze. Regression models are also based on feature extraction and are sensitive to feature errors, but to a lesser extent [16, 9, 17]. Finally, models based on appearance are more robust and can operate well with low quality images, since they are not based on local features but use information from the whole eye region [10, 18, 11, 5, 19]. Additionally, they do not require any camera calibration, however their estimation accuracy is lower than previous mentioned models.

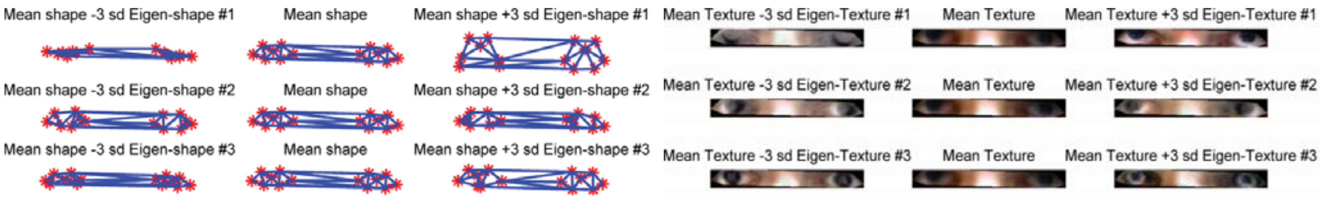
This paper's contributions are summarized as follows. In Section 2 we deal with the problem of tracking and modeling the face and the eyes' region using AAMs. In Section 3 we proceed to the estimation of eye gaze by fitting statistical models either on the AAMs parameters or on HOG features extracted from the eye's region. Our method, in addition to the classification into discrete directions, can provide a continuous estimation for eye gaze by computing gaze direction angles from GMM log-likelihoods. The proposed method performed well even in difficult tasks like sign language videos (Section 4, which have low face resolution and face occlusions. The evaluation, based on an Eye Tracker estimation (Section 5), confirms the effectiveness and accuracy of our methods, as well.

2. FACE AND EYES MODELING

2.1. Active Appearance Models

For the task of face tracking and modeling we have used the Active Appearance Models (AAMs), which were originally proposed by Cootes et al. [20, 21] and later improved by Baker and Matthews

This research was supported by the project "COGNIMUSE" which is implemented under the "ARISTEIA" Action of the Operational Program "Education and Lifelong Learning" and is co-funded by the European Social Fund (ESF) and National Resources. It was also partially supported by the European Union under the projects "MOBOT" with grant FP7-600796 and "DIRHA" with grant FP7-288121.



(a) Shape Model (b) Texture Model

Fig. 1: Variance of the first 3 eigenshapes and eigentextures of local AAM for eyes region into $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$, where λ_i is the corresponding eigenvalue. Best viewed in color.

[22, 23], who added fast and efficient algorithms for fitting AAMs in new images. The AAMs are our basic tool for the development of eye gaze estimation methods. In some cases we use directly the AAMs parameters, while others methods rely on the result of shape mask fitting, in order to extract various features for the face. Also, in some approaches AAMs can be used as an initial preprocessing stage before the main algorithm is applied.

Global AAMs: For the sign language videos we trained *subject-specific* AAMs for the whole face (Global AAM), which capture most variance of the face and especially the eyes movements. The tracking initialization is achieved using the Viola-Jones face detection [24] while for the training of AAMs we manually annotate an image set with the face mask landmarks. We use 86 points for the graph of face shape, including points for the forehead, in order to increase the model accuracy. We keep 90% and 80% of the variance for the shape and texture model respectively. For fitting the AAMs into novel images we use the inverse compositional algorithm, proposed by Baker and Matthews [23]. The fitting process estimates the concatenated shape and texture parameters vector $\mathbf{q} = [\tilde{\mathbf{p}}^T, \tilde{\boldsymbol{\lambda}}^T]^T$ that minimizes the norm $\|E(\mathbf{q})\|_2^2$, where $\tilde{\mathbf{p}}$ are the shape parameters $\mathbf{p}$ concatenated with the parameters of a 4 d.o.f. similarity transform and $\tilde{\boldsymbol{\lambda}}$ are the texture parameters $\boldsymbol{\lambda}$ enhanced with the parameters of a global affine texture transformation. $E(\mathbf{q})$ is the error image defined as the difference between the reconstructed texture and the image texture warped on the mean shape $\mathbf{s_0}$. This optimization problem is linear in texture parameters but non-linear for the shape parameters. For its solution we use the adaptive and constrained algorithms for inverse compositional AAMs fitting, proposed in [25].

Eyes Region Local AAMs: In order to decompose and model better the eyes region, we train and fit a local AAM, initializing the points of its shape graph from the corresponding landmarks of the Global AAM. We have kept 6 eigenshapes and 12 eigentextures. Figure 1 shows the variance of the first 3 eigenshapes and eigentexture of the local AAM on GSL database. Notice that the first eigenshape models the opening and closing of the eyes, while the second and the third describe the motion of the iris. Moreover, the texture model describes also successfully the changes in the position of the iris.

2.2. The Database

We have implemented and evaluated our proposed techniques, by using videos from two databases: 1) Greek Sign Language (GSL) [26] by the Institute for Language and Speech Processing ("ILSP") and 2) American Sign Language database BU400 [27]. In sign language eye movements play an important role, because the eyes are one of the main ways for changing prosody in sign language and to direct our "speech" to another person. But there are some technical limitations in the use of sign language videos. First, most video databases do not have sufficient annotation for checking the results accuracy. Also, eye gaze estimation system must be designed without requiring a specific calibration, because if a video has been recorded at previous time, it will may be impossible to extract such information. Finally, because in sign language hands movement plays the dominant role, the whole upper body of the signer has been recorded. This results in low resolution of the image in the face region and much more around the eyes, even in cases where the initial resolution of the video is quite high. In Fig. 2 we see indicative tracking results on GSL database. Note that the fitting of the AAM around the eyes region is less influenced by the face occlusions.

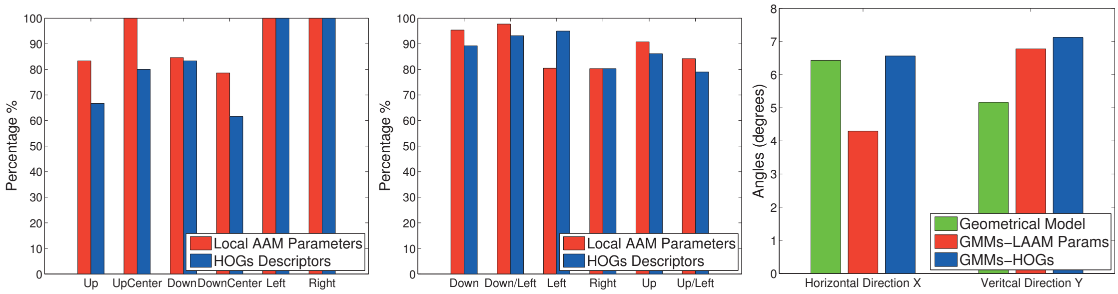
(a) (b) (c) (d) (e) (f)

Fig. 2: Global AAM tracking results on GSL database.

3. EYE GAZE ESTIMATION USING GMMS AND EYES APPEARANCE

For the eye gaze classification we employ an approach based on Gaussian Mixture Models (GMMs). For this purpose, it is needed to define "manually" a training set with images for each class, which represents a different gaze direction. In our approach we used six classes describing the basic directions of gaze: "Up","Center and Up", "Down","Center and Down","Left","Right". Then, for each class an independent GMM is trained using the EM algorithm, which expresses the gaze classification probability in the above classes.

For the classification process, we used as feature vector the parameters of the Local AAM for eyes region. This vector comprises the parameters $\mathbf{p}, \boldsymbol{\lambda}$, which control the LAAM shape and texture, but not those which are related with similarity transformation and global affine texture transformation. We have also included the first two shape parameters of the Global AMM since they mainly contain information about the head pose [4, 28]. We have also experimented with Histograms of Oriented Gradients (HOGs), which describe the structure of the shape in an image and have been used with great success in object recognition, such as the pedestrian detection [29, 30]. There are also used to express the eyes region changes for facial expressions recognition tasks, which are needed in the automatic sign language recognition [5]. So, we expect to have similar successful results for the eye gaze estimation problem. For HOGs features extraction it is necessary first to determine the eyes region. This can be done by defining on the face image a rectangle based on the points of AAM shape mask, which encloses the eyes region. Then, for each eyes image we extract a vector of HOGs based features, which will be used for gaze classification. For the new testing samples, we first



(a) GSL Database (b) BU400 Database (c) Mean Absolute Error relative to Eye-Tracker

Fig. 3: **a,b)** Percentages for successful eye gaze classification based on Local AAM parameters (red) or HOGs descriptors for eyes region (blue), using GMMs. **c)** Mean Absolute Error (MAE) for horizontal and vertical angle estimation, relative to Eye Tracker estimation. Note that the lower performance in vertical direction is related to the small changes in eye tracker estimation due to the limited screen height.

compute the log-likelihood for each of the 6 gaze categories and then for each direction (horizontal or vertical) we select the class with the highest likelihood.

Estimation of Eye Gaze Angles from GMM Likelihoods: The described gaze direction classification provides us a quite good discrete estimation for eye gaze. However, in many cases it is necessary to have a continuous estimate for gaze angles, so the accuracy given by the 6 basic classes is not enough. For this reason, we propose a method, which makes use of GMMs log-likelihood for each one of the 6 basic classes, in order to compute angles for the horizontal and vertical gaze direction. The proposed method uses two different approaches for the computation of the angles. The first one, assumes large changes in the gaze angles. It is based on the estimation of the gaze vector projection on the image plane, using the likelihood ratio between the 6 basic classes. The other one is related with small gaze changes around the straight direction, and is computed by a linear regression between GMMs likelihood and gaze direction angles.

For the determination of the angle θ_{ver} in the vertical direction, we first compute the following likelihood ratios (LL_i is defined as the likelihood of each class minus the overall minimum value): $V_{up} = \frac{LL_{up} - LL_{upcenter}}{LL_{up}}$ and $V_{down} = \frac{LL_{down} - LL_{downcenter}}{LL_{down}}$. The total vertical projection will be $V_{UD} = V_{up} - V_{down}$, while the vertical direction angle will be given by: $\theta_{ver} = \arctan(V_{UD})$. Afterwards, we compute the ratio between the two central directions: $D_{center} = \frac{LL_{upcenter} - LL_{downcenter}}{\max(LL_{upcenter}, LL_{downcenter})}$. In the estimation of angle θ_{ver} we have considered that there are two different central directions. So, we add in the above result a correction term, which expresses the difference between the two central directions (κ_{center} is an experimentally defined small constant angle between $7^\circ - 10^\circ$): $\theta_{ver} = \arctan(V_{UD}) + D_{center} \cdot \kappa_{center}$. In a similar manner, we compute the ratios V_{left}, V_{right} for the horizontal direction: $V_{left(right)} = \frac{LL_{left(right)} - LL_{center}}{LL_{left(right)}}$. As *center* it is selected the central direction with the highest likelihood. So, the total horizontal projection is $V_{RL} = V_{right} - V_{left}$ and the relevant horizontal angle will be given by: $\theta_{hor} = \arctan(V_{RL})$.

However, the method described above performs weakly in estimating gaze angles, that lie very close to the straight direction. So, two angles ϕ_{ver} and ϕ_{hor} are introduced for describing small variations of gaze around the central direction. This angles are computed using linear regression on class likelihoods: $\phi_{ver} = \left(\frac{LL_{up}}{LL_{upcenter}} - \frac{LL_{down}}{LL_{downcenter}}\right)\lambda_{ver}$ and $\phi_{hor} = \left(\frac{LL_{right} - LL_{left}}{LL_{center}}\right)\lambda_{hor}$. The angles $\lambda_{ver}, \lambda_{hor}$ are two small angles about $5^\circ - 8^\circ$, defined also experimentally. For the angle ϕ_{ver} it should be added the angle between the two central directions. So, for large gaze variations, angles are computed by the previous equations, while for small changes around the central direction will be given by: $\theta_{ver} = \phi_{ver} + D_{center} \cdot \kappa_{center}$ and $\theta_{hor} = \phi_{hor}$.

4. EVALUATION ON SIGN LANGUAGE VIDEOS

Evaluation on GSL: Figure 3a presents percentages for successful gaze classification, in each of the 6 classes, on GSL database. As classifier input we have selected either the parameters of the local AAM (red) or the HOG features (blue), and we have employed mixture models with 2 Gaussians ($K = 2$). We can observe that recognition rates reach the perfect for the classes "UpCenter","Left","Right", while for the other classes are also quite high. This happens because the above 3 classes contain more obvious changes in the eyes region relative to the other classes. At this point, we should note that in cases where the eyes appear closed, we considered that the gaze direction is "Down". As we have mentioned, the recognition rates for these 6 very basic eye gaze classes are quite high, proving in this way the success of the proposed method. However, if the classes number increases then, as it is expected, the rates will decrease considerably. In addition, classify the gaze into more classes requires the annotation of the corresponding training set. This is not always straightforward, because a more accurate classification for eye gaze is a difficult task even for humans. Regarding the employed feature vector, we observe that in general the recognition rates remain quite high. However, excluding " Left" and " Right", the other classes rates appear lower compared to those using the LAAM parameters. Also, we should mention that the vector of HOGs features is significantly larger (81 values) than that of LAAM parameters (20 values). So, in most cases, gaze classification based on LAAM parameters seems to yield better results, with the use of a much smaller amount of parameters. This is expected, because HOGs describe the local changes of an image shape, in contrast to the AMMs, which give a representation of a given object based on information from other similar images. On the other hand, the extraction process for the HOGs is computably easier than AAMs training and fitting. The required eyes region can be found by fitting a more simple model, like Active Shape Models [21], or using other facial points detectors [31, 32], e.g. a Haar-like detector in a similar manner as in face detection [33, 34, 35].

Estimation of gaze angles on GSL: We have applied the proposed method for gaze angles estimation on the GSL database. First, we compute, as before, the log-likelihood for each class GMM, using

Fig. 4: Examples of estimating gaze direction angles from GMM likelihoods for each class using the **LAAM parameters (red)** or the **HOGs descriptors (blue)**, on GSL database. We also see the gaze estimation using a **geometrical model (green)** [2]. Best viewed in color.

as feature vector either the LAAM parameters or HOGs descriptors. Then, direction angles are estimated according to the above algorithm. Figure 4 presents indicative results of these methods. We also compare our method with a state of the art geometrical model for eye gaze estimation, combining available information about the head pose. The initial model is based on the Ishikawa et al.[2] model for driver gaze tracking. However, we have made some modifications in order to increase robustness with low resolution eye images and to be able to easily calibrate the whole system using only one image of the straight gaze direction. In general, we can see that the results are quite accurate and agree in a large degree with human sense. The proposed method managed to give quite good results in difficult cases, like face occlusions, where geometrical models have difficulties to give a correct estimate. On the other hand, the estimated angles tend to magnify the extreme gaze changes, as the up/down (or left/right) gaze turn. Finally, depending on the feature vector employed, it seems clearly that the LAAM parameters perform better than HOG descriptors. The latter, in some cases, fail to estimate correctly the gaze direction (Figure 4c, 4f). This fact was observed during the classification process as well, but it becomes more apparent with the continuous angles estimation.

Evaluation on BU400: Next we evaluate our classification methods on BU400 database which has eye gaze annotation for the basic angles. Spesifically, we classify the gaze into 6 different directions: “Down”, “Down/Left”, “Left”, “Right”, “Up” and “Up/Left”. Figure 3b presents accuracy percentages for the gaze classification, on BU database. In this dataset we employ GMMs with $K = 7$ Gaussians due to the low resolution videos and the high variance in both head pose and eye gaze directions. We confirm that the method based on the LAAM parameters performs better than the HOG features for almost all the gaze classes.

5. EVALUATION BASED ON EYE TRACKER ESTIMATION

In this last section, we will compare our continuous estimation for eye gaze direction angles, with the one that is provided by an eye tracking system. As we have seen, the above mentioned methods have good quality performance over sign language videos, which have the limitations that we have already discussed. Unfortunately,

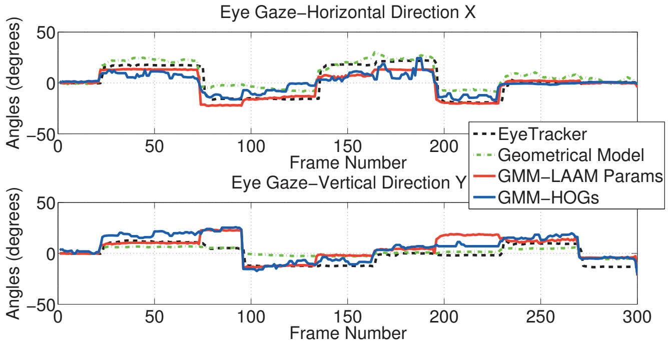


Fig. 5: Example of horizontal and vertical angle estimation for gaze variations inside screen boundaries. Best viewed in color.

we cannot provide any quantitative evaluation of these results, because it is difficult to find a database with both sufficient continuous annotation for eye gaze and spontaneous scenes, such as sign language videos. For this reason, we are going to quantitatively evaluate the described methods on a somewhat different task: the eye gaze estimation when someone sits in front of his computer monitor. This is a widely used case, for which we can obtain ground truth estimation via an eye tracking system. Specifically, we have used the commercial Eye Tracking System TM3 provided by EyeTechDS. Figure 5 presents an example of the estimated angles for gaze variations inside screen boundaries, while in Fig. 3c we see the Mean Absolute Error (MAE) relative to Eye Tracker estimation. For the estimation using HOGs, we have applied a median post-filter because the original data had a lot of noise. It is clear that for the horizontal direction all methods’ results are close to the tracker’s estimation. Although, GMMs with LAAM parameters have the best performance, with mean absolute error value around 4°. For vertical direction, there are deviations between methods and tracker estimation. In this case, according to MAE, the Geometrical Model seem to perform better. However, from the angle curve we see that it happens because its estimation is very close to zero while the other two methods track better the gaze changes. The lower performance in vertical direction is partly explained by the fact that the screen height is not big enough, in order to have large variations both in gaze angles and eyes region. In addition, let us note that for all methods the MAE is relatively low, under 7°. Closing, we can say that “GMMs-LAAM” method has, in general, the overall better performance, as we have also mentioned and for the GSL database.

6. CONCLUSIONS

In this work, we approach the problem of continuous estimation of eye gaze angles, with application to low resolution face images from sign language videos. We proceeded to a discrete gaze classification using GMMs, based on either LAAM parameters or HOG descriptors. Evaluation on sign language databases showed that both representations are promising, although LAAM seems to perform better. Additionally, we proposed a method for continuous estimation of gaze angles from GMM log-likelihoods with successful results, as well. We have also compared our method with a state of the art geometrical model for gaze estimation, based on LAAM landmarks. Qualitative experimental evaluation on sign language videos verify the effectiveness of our method. Finally, we evaluated our methods using ground-truth data from an Eye Tracker and showed that for all cases the Mean Absolute Error (MAE) for both gaze directions is relatively low and specifically under 7°.

7. REFERENCES

[1] Laura Florea, Corneliu Florea, Ruxandra Vrânceanu, and Constantin Vertan, "Can your eyes tell me how you think? a gaze directed estimation of the mental activity," in *Proc. British Machine Vision Conf.*, 2013.

[2] T. Ishikawa, S. Baker, I. Matthews, and T. Kanade, "Passive Driver Gaze Tracking with Active Appearance Models," in *Proc. World Congress on Intelligent Transportation Systems*, 2004.

[3] R. Valenti, J. Staiano, N. Sebe, and T. Gevers, "Webcam-based visual gaze estimation," in *Proc. Int'l Conf. Image Analysis and Processing ICIAP*, 2009.

[4] U. von Agris, J. Zieren, U. Canzler, B. Bauer, and K.-F. Kraiss, "Recent developments in visual sign language recognition," *Universal Access in the Information Society*, vol. 6, no. 4, pp. 323–362, 2008.

[5] N. Michael, C. Neidle, and D. Metaxas, "Computer-based recognition of facial expressions in ASL: From face tracking to linguistic interpretation," in *Proc. Workshop on Representation and Processing of Sign Languages, LREC*, 2010.

[6] S. Sclaroff, M. Betke, G. Kollios, J. Alon, V. Athitsos, R. Li, J. J. Magee, and T.-P. Tian, "Tracking, Analysis, and Recognition of Human Gestures in Video," in *Proc. Int'l. Conf. on Document Analysis and Recognition*, 2005.

[7] L. Maat and M. Pantic, "Gaze-X: Adaptive, Affective, Multimodal Interface for Single-User Office Scenarios," in *Proc. ACM Intl Conf. Multimodal Interfaces*, 2006.

[8] Y. Matsumoto and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement," in *Proc. IEEE Int'l. Conf. on Automatic Face and Gesture Recognition*, 2000.

[9] M. Takatani, Y. Ariki, and T. Takiguchi, "Gaze estimation using Regression Analysis and AAMs parameters selected based on Information Criterion," in *Proc. Int'l. Workshop on Gaze Sensing and Interactions, ACCV*, 2010.

[10] T. D. Rikert and M. J. Jones, "Gaze Estimation using Morphable Models," in *Proc. IEEE Int'l. Conf. on Automatic Face and Gesture Recognition*, 1998.

[11] D. W. Hansen, J. P. Hansen, M. Nielsen, A. S. Johansen, and M. B. Stegmann, "Eye typing using Markov and Active Appearance Models," in *Proc. IEEE Workshop on Applications of Computer Vision*, 2004.

[12] J.G. Wang, E. Sung, and Venkateswarlu R., "Eye gaze estimation from a single image of one eye," in *Proc. Int'l Conf. on Computer Vision*, 2003.

[13] T. Ohno, N. Mukawa, and A. Yoshikawa, "Freegaze: A gaze tracking system for everyday gaze interaction," in *Proc. Symposium on Eye tracking Research and Applications*, 2002.

[14] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, 2010.

[15] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. on Biomedical Engineering*, vol. 53, no. 6, pp. 1124–1133, 2006.

[16] J. Zhu and J. Yang, "Subpixel eye gaze tracking," in *Proc. IEEE Int'l. Conf. on Automatic Face and Gesture Recognition*, 2002.

[17] Krystian Radlak, Michal Kawulok, Bogdan Smolka, and Natalia Radlak, "Gaze direction estimation from static images," in *Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2014.

[18] G. Bebis and K. Fujimura, "An Eigenspace Approach to Eye-Gaze Estimation," in *Proc. ISCA Int'l. Conf. on Parallel and Distributed Computing Systems*, 2000.

[19] Y. Sugano, Y. Matsushita, and Y. Sato, "Appearance-based gaze estimation using visual saliency," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 329–341, 2013.

[20] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active Appearance Models," in *Proc. European Conference Computer Vision*, 1998.

[21] T. F. Cootes and C. J. Taylor, "Statistical Models of Appearance for Computer Vision," 1999, Technical Report, University of Manchester.

[22] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," *Int'l. Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.

[23] I. Matthews and S. Baker, "Active Appearance Models revisited," *Int'l. Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[24] P. Viola and M. J. Jones, "Robust real-time face detection," *Int'l. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[25] G. Papandreou and P. Maragos, "Adaptive and constrained Algorithms for Inverse compositional AAM Fitting," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.

[26] E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Glauert, R. Bowden, A. Braffort, P. Maragos, and F. Lefebvre-Albaret, "Sign Language technologies and resources of the Dicta-Sign project," in *Proc. Workshop on Representation and Processing of Sign Languages, LREC*, 2012.

[27] P. Dreuw, C. Neidle, V. Athitsos, S. Sclaroff, and H. Ney, "Benchmark databases for video-based automatic sign language recognition," in *Proc. LREC*, 2008.

[28] E. Antonakos, V. Pitsikalis, I. Rodomagoulakis, and P. Maragos, "Unsupervised classification of extreme facial events using active appearance models tracking for sign language videos," in *Proc. Int'l Conf. Image Processing*, 2012.

[29] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005.

[30] N. Dalal, "Finding People in Images and Videos, PhD Thesis," 2006, Institut National Polytechnique de Grenoble/INRIA.

[31] Michel Valstar, Brais Martinez, Xavier Binefa, and Maja Pantic, "Facial point detection using boosted regression and graph models," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010.

[32] Brais Martinez, Michel François Valstar, Xavier Binefa, and Maja Pantic, "Local evidence aggregation for regression-based facial point detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 5, pp. 1149–1163, 2013.

[33] H. Cevikalp, B. Triggs, and V. Franc, "Face and landmark detection by using cascade of classifiers," in *Proc. IEEE Int'l. Conf. on Automatic Face and Gesture Recognition*, 2013.

[34] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2012.

[35] M. Kawulok and J. Szymanek, "Precise multi-level face detector for advanced analysis of facial images," *Image Processing, IET*, vol. 6, no. 2, pp. 95–103, 2012.