# Human–Robot Interaction Through Robust Gaze Following

**Sorin M. Grigorescu and Gigel Macesanu**

**Abstract** In this paper, a probabilistic solution for gaze following in the context of joint attention will be presented. Gaze following, in the sense of continuously measuring (with a greater or a lesser degree of anticipation) the head pose and gaze direction of an interlocutor so as to determine his/her focus of attention, is important in several important areas of computer vision applications, such as the development of nonintrusive gaze-tracking equipment for psychophysical experiments in Neuroscience, specialized telecommunication devices, *Human–Computer Interfaces* (HCI) and artificial cognitive systems for *Human–Robot Interaction* (HRI). We have developed a probabilistic solution that inherently deals with sensor models uncertainties and incomplete data. This solution comprises a hierarchical formulation of a set of detection classifiers that loosely follows how geometrical cues provided by facial features are used by the human perceptual system for gaze estimation. A quantitative analysis of the proposed architectures performance was undertaken through a set of experimental sessions. In these sessions, temporal sequences of moving human agents fixating a well-known point in space were grabbed by the stereovision setup of a robotic perception system, and then processed by the framework.

## 1 Introduction

Head movements are commonly interpreted as a vehicle of interpersonal communication. For example, in daily life, human beings observe head movements as an expression of agreement or disagreement in a conversation, or even as a sign of confusion. On the other hand, gaze shifts are usually an indication of intent, as they commonly precede action by redirecting the sensorimotor resources to be used. As a

S.M. Grigorescu (✉) · G. Macesanu
Department of Automation, Transilvania University of Brasov, Mihai Viteazu 5,
500174 Braşov, Romania
e-mail: s.grigorescu@unitbv.ro

G. Macesanu
e-mail: gigel.macesanu@unitbv.ro



P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_10

**Fig. 1** Gaze following in the context of joint attention for HRI, using the ROVIS system on a Neobotix® MP 500 mobile platform

consequence, sudden changes in gaze direction can express alarm or surprise. Gaze direction can also be used for directing a person to observe a specific location. To this end, during their infancy, humans develop the social skill of *joint attention*, which is the means by which an agent looks at where its interlocutor is looking at by producing an eye-head movement that attempts to yield the same focus of attention. Over nine months of age, infants are known to begin to engage with their parents/caregivers in an activity in which both look at the same target through joint attention.

As artificial cognitive systems with social capabilities become more and more important due to the recent evolution of robotics towards applications where complex and human-like interactions are needed, basic social behaviors such as joint attention have increasingly become important research topics in this field. Figure 1 illustrates the ROVIS[1] (*Robust Vision and Control Laboratory*) gaze following system at work, under the context of joint attention for *Human Robotic Interaction* (HRI). Gaze following thus represents an important part of building a social bridge between humans and computers. Researchers in robotics and artificial intelligence have been attempting to accurately reproduce this type of interaction in the last couple of decades, and, although much progress has been made [1], dealing with perceptual uncertainty still renders it difficult for these solutions to work adaptively.

Gaze following is an example for which the performance of artificial systems is still far from human adaptivity. In fact, the gaze following adaptivity problem can be stated as follows: how can gaze following be implemented under nonideal circumstances (perceptual uncertainty, incomplete data, dynamic scenes, etc.)? Figure 2 demonstrates how incomplete data, arguably the issue where the lack of adaptivity and underperformance of artificial systems are most apparent, might influence the outcome of gaze following.

In the following text, we propose a robust solution to facial feature detection for human–robot interaction based on (i) a feedback control system implemented at the image processing level for the automatic adaptation of the system's parameters, (ii) a

[1] http://rovis.unitbv.ro.

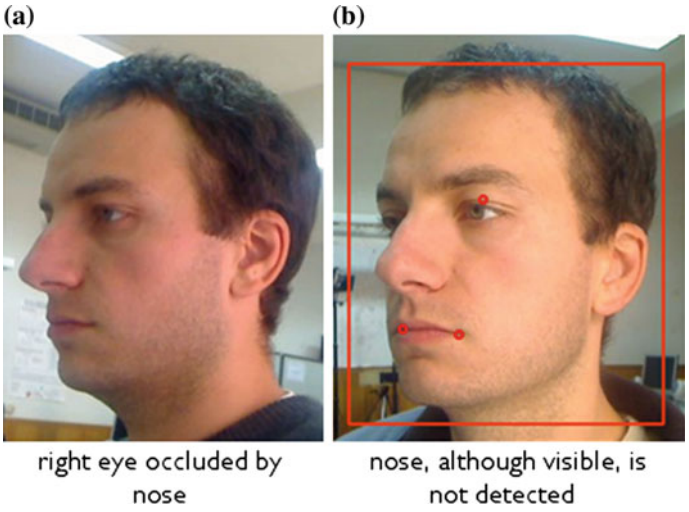


**Fig. 2** Examples of probable gaze following failure scenarios due to incomplete data: facial features occluded in profile views (**a**), or failure of feature detection algorithms (**b**)

cascade of facial features classifiers, and (iii) a *Gaussian Mixture Model* (GMM) for facial points segmentation. The goal is to obtain a real-time gaze following estimator which can cope with uncertainties and incomplete data. The proposed system aims at the robust computation of the human gaze direction in the context of joint attention for HRI.

## 2 Related Work

### *2.1 Gaze Following*

In recent years, the problem of gaze following has been extensively studied. Physiological investigations have demonstrated that the brain estimates the gaze as a mixture of eye direction and head position and orientation (pose) [2]. By itself, head pose provides an estimate that represents a coarse approximation of gaze direction that can be used in situations in which the eyes are invisible (e.g., when observing a distant person, or when sunglasses occlude the eyes) [3]. When the eyes are not occluded, the head pose is an extra marker that can be used to estimate the direction of the gaze. The gaze direction estimation problem, as it is solved by the human brain, can therefore be subdivided into two fundamental and *sequential* subproblems: *head pose estimation* and *eye gaze estimation*.

The consequences of such a solution are twofold: partial information can be used to already arrive to an estimate; however, this happens at the expense of biasing. As

**Fig. 3** Wollaston illusion: although the eyes are the same in both images, the perceived gaze direction is dictated by the orientation of the head. (Adapted from [2, 3])

an illustration of this drawback, in Fig. 3 is shown [2] that the interpretation of the gaze for an observer is deviated in the direction of the head. In any case, the error propagated by erroneously estimating one of the features is greatly compensated by the fact that the human brain is able to yield an estimate *even when only presented with partial or incomplete information*. Moreover, visual features used to detect a face or an eye do not need to be the same for both cases, so they can be detected independently, which makes the problem more tractable.

Consequently, the following paragraphs will present a summarized survey of solutions for each subproblem.

In the survey by [3], solutions for head pose estimation are divided into eight categories: seven represent pure methods, while the remaining are hybrid methods, i.e., combinations of the other methods. The article ends by presenting a quantitative comparison of the performance of these methods.

As mentioned in this survey, most of the computer vision based head pose calculation algorithms have diverged greatly from the results of psychophysical experiments as to how the brain tackles this problem. In fact, the former are concentrated on *appearance-based* methods, while the latter takes into account how the human perceives the pose of the head based on *geometrical cues* [3].

Geometrical approaches, as shown in Fig. 4, attempt to detect head features as accurately as possible in order to compute the pose of the head. An example of a geometrical approach for head pose estimation is presented in [4], where monocular images are used as input information. The proposed algorithm makes minimal assumptions, compared with other methods, about the facial features structure. Knowing the positions of the nose, eyes, and mouth, the facial normal direction can be obtained from one of the next two methods [4], also used in our work:

1. Using two relations: the nose tip and the line between the far corners of the mouth ($R_1 = \frac{l_m}{l_f}$); the line between one eye with the correspondent far corners of the mouth and the distance given by the nose tip; and the line connecting one eye with the far corners of the mouth ($R_2 = \frac{l_n}{l_f}$);
2. Using the line between the eye extremities and the far mouth corners.

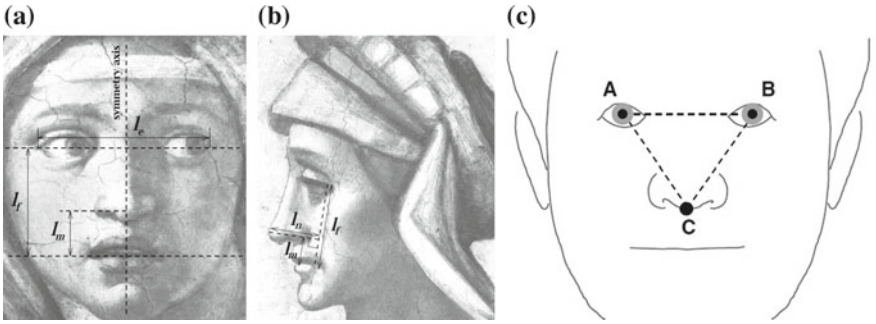


**Fig. 4** Geometrical relations between facial features (Adapted from [4]). 3D gaze orientations can be computed using the distances between detected facial features, such as the eyes, nose and mouth

The derivation of the roll, pitch, and yaw for a human head is presented in [5]. The assumption from this article is that the four points that describe the eye are collinear. The position is obtained using the line through the four eye points and the nose tip. The main difficulties with this method are related to the pitch direction estimation, which uses an anthropometric face analysis [5]. The yaw and the pitch are obtained from eye corners and the intrinsic camera parameters (focal length).

The method proposed by [6] uses the model of the face and the eye, deduced from anthropometric features in order to determine the head orientation. This method uses only three points (e.g., eye centers and the middle point between the nostrils) to perform the desired task. Their model uses the following assumption: $d(A, C) = d(B, C)$; $d(A, B) = kd \cdot d(A, C)$; $d(A, B) = 6,5$ cm, where $A$ and $B$ are the central points of each eye and $C$ is the middle point between the nostrils.

Another solution for head pose estimation is introduced in [7]. The main idea here is to consider an isosceles triangle, with corners in both eyes and in the center of the mouth. The direction of the head is computed if we assume that one side of the triangle lies on the image plane, such that applying a trigonometric function we can estimate the angle between the triangle plane and the image plane [7].

Finally, an alternative method for head estimation is supposed to use multiple cameras [8] with accurate calibration information available. Skin color segmentation is performed on each camera, and then data fusion is performed, resulting in a 3D model of the head. The orientation of the head is estimated based on a particle filter.

## *2.2 Facial Features Extraction*

Feature detection represents a subtopic within the head pose estimation problem. An accurate estimate for the eye, nose, or the mouth represents an intermediate stage, in which essential information used by the geometrical approach for head pose estimation is computed. Methods for gaze estimation, presented in the following section,

include eye feature detection. Detection of other important facial features, such as the mouth and the nose, is discussed next.

Mouth recognition is dealt with methods such as the ones suggested in [9, 10]. A common approach for detecting the mouth is by pre-segmenting the color red on a specific patch of the image. Both methods use a ROI (Region of Interest) extracted after head segmentation, in which the mouth is approximately segmented, after a color space conversion is performed (such as RGB to HSI (*Hue, Saturation, Intensity*) [9], or RGB to *Lab* [10]). On the other hand, nose detection algorithms use Boosting classifiers, commonly trained with Haar-like features [11], or the 3D information of the face, as in [12].

As suggested in [13], most of the methods used for eyes detection and segmentation can be divided into shape-based, appearance-based and hybrid methods. The shape-based technique uses the detection of the iris, the pupil, or the eyelids to locate the eye. Particular features, such as the pupil (dark/bright pupil region) or cornea reflections are used in appearance-based approaches, while the hybrid method tries to combine the advantages of both methods.

The shape-based algorithm proposed in [14], built on the isophote curvature concept, i.e., the curve that connects points of the same intensity, is able to deliver accurate eye localization from a web camera. The main advantage of using this concept is that the shape of the isophotes is invariant to rotation or to linear illumination changes. The eye location can be determined using a combination of Haar features, dual orientation Gabor filters and eye templates, as described in [15].

Unsupervised learning algorithms, such as the *Independent Component Analysis* (ICA), are used in [16] for eyes extraction, based on the fact that the eye is a stable facial feature. The two stages technique determines first a rough eye ROI using ICA and the gray-level image intensity variance, and second, the eye center point is computed from image intensity data.

Finally, an alternative method which uses two visual sensors is proposed in [17]: a wide-angle camera for face detection and rough eyes estimation and an active pan–tilt–zoom camera to focus on the rough detected ROIs. The method considers the face as a 3D terrain surface and the eye areas as ”pits” and ”hillsides” regions. The eyes 2D positions are chosen using a (GMM). A similar dual stereo camera system is also proposed in [18], where a wide-angle camera detects the face and an active narrow *Field of View* (FoV) system tracks the eyes at high resolution.

As mentioned above, most methods tackle the problem of gaze direction estimation using either head pose or eyes direction estimation. However, papers such as [14, 19, 20] present hybrid approaches that combine head pose and eye direction estimation for obtaining the subject's gaze direction.

In [14], a hybrid solution for eye detection and tracking, combining the detection results with a *Cylindrical Head Model* (CHM) for head direction estimation, is presented. In [19], the gaze's direction is computed in two stages, after a camera calibration process: first the eyes orientation vector is determined with respect to the head's coordinate system and, second, the final gaze direction estimate is given by a fusion between the determined eyes and head's poses. Both approaches have lim-

itations in estimating the gaze's orientation when either the eyes or the poses of the head are imprecise.

The technique from [20] describes a human gaze direction algorithm from a combination of *Active Appearance Models* (AAM) and a CHM. Although the approach seems to perform well in off-line experiments, real-time scenarios are not presented. One other notable facial features extractor is the Flandmark system [21], which, despite its real-time capabilities and ability to detect and track facial features from frontal faces, fails to recognize features when the pose of the head has a slight offset from the frontal view.

## 3 Controlling a Machine Vision System

In a robotics application, the purpose of the machine vision system is to perceive the environment through a camera module.

An image processing chain is usually composed of low (e.g., image enhancement, segmentation) and high (e.g., object recognition) level image processing methods. In order for the high level operations to perform properly, the low level ones have to deliver reliable information. In other words, object recognition methods require reliable input coming from previous operations [22].

In order to improve the image processing chain, we propose to control the low level vision operation through a feedback loop derived from the higher level components. In [23, 24], the inclusion of feedback structures within vision algorithms for improving the overall robustness of the chain is suggested.

The core idea of the feedback control system for adapting the low level vision operations is presented in Fig. 5, where the control signal $u$, or *actuator variable*, is a parameter which controls the processing method, whereas the *controlled variable* $y$ is a measure of image processing quality.

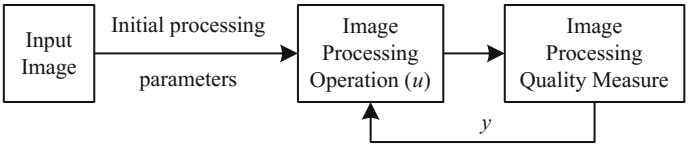


**Fig. 5** Feedback adaptation of a computer vision algorithm. The image processing quality measure $y$ is used as a feedback control variable for adapting the parameters of the vision algorithms using the actuator $u$

## 4 Image Processing Chain

The gaze following image processing chain, depicted in Fig. 6, contains four main steps. We assume that the input is an 8-bit gray-scale image $I = J^{V \times W}$, of width $V$ and height $W$, containing a face viewed either from a frontal or profile direction, where $J = \{0, \ldots, 255\}$. $(v, w)$ represents the 2D coordinates of a specific pixel. The face region is obtained from a face detector.

First, a set of facial features ROI hypotheses $\mathbf{H} \in \{h_{le}, h_{re}, h_n, h_m\}$, consisting of possible instances of the left $h_{le}$ and right $h_{re}$ eyes, nose $h_n$ and mouth $h_m$, are extracted using a local features estimator which determines the probability measure $p(\mathbf{H}|I)$ of finding one of the searched local facial region. The number of computed ROI hypotheses is governed by a probability threshold $T_h$, which rejects hypotheses with a low $p(\mathbf{H}|I)$ confidence measure. The choice of the $T_h$ threshold is not a trivial task when considering time critical systems, such as the gaze estimator, which, for a successful HRI, has to deliver in real-time the 3D gaze orientation of the human subject. The lower $T_h$ is, the higher the computation time. On the other hand, an increased value for $T_h$ would reject possible "true positive" facial regions, thus leading to a failure in gaze estimation. As explained in the following, in order to obtain a robust value for the hypotheses selection threshold, we have chosen to adapt $T_h$ with respect to the confidences provided by the subsequent estimators from Fig. 6, which take as input the facial regions hypotheses. The output probabilities coming from these estimation techniques, that is, the spatial estimator and the GMM for point-wise feature extraction, are used in a feedback manner within the extremum seeking control paradigm.

Once the hypotheses vector $\mathbf{H}$ has been built, the facial features are combined into the spatial hypotheses $\mathbf{g} = g_0, g_1, \ldots, g_n$, thus forming different facial region combinations. Since one of the main objectives of the presented algorithm is to identify facial points of frontal, as well as profile faces, a spatial vector $s_i$ is composed either from four, or three, facial ROIs:

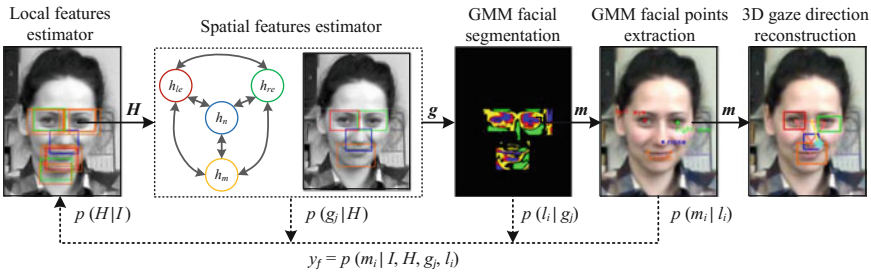


**Fig. 6** Block diagram of the proposed gaze following system for facial feature extraction and 3D gaze orientation reconstruction. Each processing block within the cascade provides a measure of feature extraction quality, fused within the controlled variable $y_f$ (see Eq. 2)

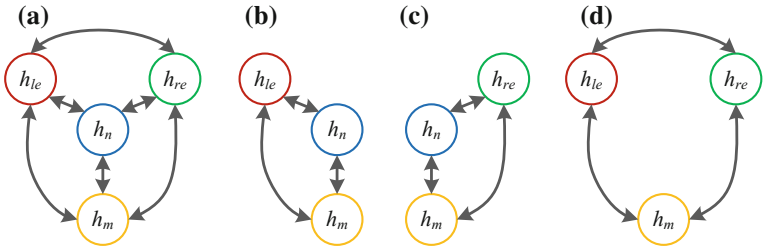


**Fig. 7** Different spatial combinations of features used for training the four classifiers. **a** All four facial features. **b**, **c**, **d** Cases where only three features are visible in the sample image

$$g_i = \{h_0, h_1, h_2, h_3\} \cap \{h_0, h_1, h_2\}, \tag{1}$$

where $h_i \in \{h_{le}, h_{re}, h_n, h_m\}$.

The extraction of the best spatial features combination can be seen as a graph search problem $g_j = f : G(\mathbf{g}, \mathbf{E}) \rightarrow \Re$, where $\mathbf{E}$ are the edges of the graph connecting the hypotheses in $\mathbf{g}$. The considered features combinations are illustrated in Fig. 7. Each combination has a specific spatial probability value $p(g_j|\mathbf{H})$ given by a spatial estimator trained using the spatial distances between the facial features from a training database.

Once the spatial distributions of the probable locations of the facial features ROIs are available, their pointwise location $m_i$ is determined using a GMM segmentation method. Its goal is to extract the most probable facial pointwise locations $m_i$ given the GMM pixel likelihood values $p(l_i|g_j)$. The most relevant point features for computing the 3D gaze of a person are the centers of the eyes, tip of the nose, and corners of the mouth.

The described data analysis methods are used to evaluate a feature space composed of the local and spatial features.

Having in mind the facial feature points extraction algorithm described above, it can be stated that the confidence value $y_f$ of the processing chain in Fig. 6 is a probability confidence measure obtained from the estimators cascade:

$$y_f = p(m_i|I, \mathbf{H}, g_j, l_i). \tag{2}$$

Since the whole described processing chain is governed by a set of parameters, such as the threshold $T_h$ for selecting the vector $\mathbf{s}$, we have chosen to adapt it using an extremum seeking control mechanism and the feedback variable $y_f$, derived from the output of the gaze following structure illustrated in Fig. 6. The final 3D gaze orientation vector $\vec{\varphi}(m_i)$, representing the roll, pitch, and yaw of the human subject, is determined using the algorithm proposed in the work of Gee and Cipolla [4].

# 5 Performance Evaluation

## 5.1 Experimental Setup

In order to test the performance of the proposed gaze following system, the following experimental setup has been prepared.

The system has been evaluated on the *Labeled Faces in the Wild* (LFW) database [25]. LFW consists of 13,233 images, each having a size of $250 \times 250px$. In addition to the LFW database, the system has been valuated on an Adept Pioneer® 3-DX mobile robot equipped with an RGB-D sensor delivering $640px \times 480px$ size color and depth images. The goal of the scenarios is to track the facial features of the human subject in the HRI context. The error between the real and estimated facial feature's locations was computed offline.

For evaluation purposes, two metrics have been used:

- the mean normalized deviation between the ground truth and the estimated positions of the facial features:

$$d(\mathbf{m}, \hat{\mathbf{m}}) = \tau(\mathbf{m}) \frac{1}{k} \sum_{i=0}^{k-1} \|m_i - \hat{m}_i\|, \tag{3}$$

  where $k$ is the number of facial features, $\mathbf{m}$ and $\hat{\mathbf{m}}$ are the manually and estimated annotated positions of the eyes, nose and mouth, respectively, and $\tau(\mathbf{m})$ is a normalization constant:

$$\tau(\mathbf{m}) = \frac{1}{\|(m_{le} + m_{re}) - m_m\|}. \tag{4}$$

- the maximal normalized deviation:

$$d^{\max}(\mathbf{m}, \hat{\mathbf{m}}) = \tau(\mathbf{m}) \max_{j=0,\ldots,k-1} \|m_i - \hat{m}_i\|. \tag{5}$$

## 5.2 Competing Detectors

The proposed gaze following system has been tested against three open source detectors.

(1) *Independent facial feature extraction*: The detector is based on the Viola–Jones boosting cascades and returns the best detected facial features, independent of their spatial relation. The point features have been considered to be the centers of the computed ROIs.
The boosting cascades, one for each facial feature, have been trained using a

few hundred samples for each eye, nose, and mouth. The searching has been performed several times at different scales, with Haar-like features used as inputs to the basic classifiers within the cascade. From the available ROI hypotheses, the one having the maximum confidence value has been selected as the final facial feature.

(2) *Active Shape Models*: An *Active Shape Model* (ASM) calculates a set of feature points along the facial features contours of the eyes, nose, mouth, eyebrows, or chin. An ASM is initially trained using a set of manually marked contour points. The open source AsmLib, based on OpenCV, has been used as candidate detector. The ASM is trained using manually marked face contours. The trained ASM model determines variations in the training dataset using *Principal Component Analysis* (PCA), which enables the algorithm to estimate if the contour is a face.

(3) *Flandmark*: *Flandmark* [21] is a deformable part model detector of facial features, where the detection of the point features is treated as an instance of struc-

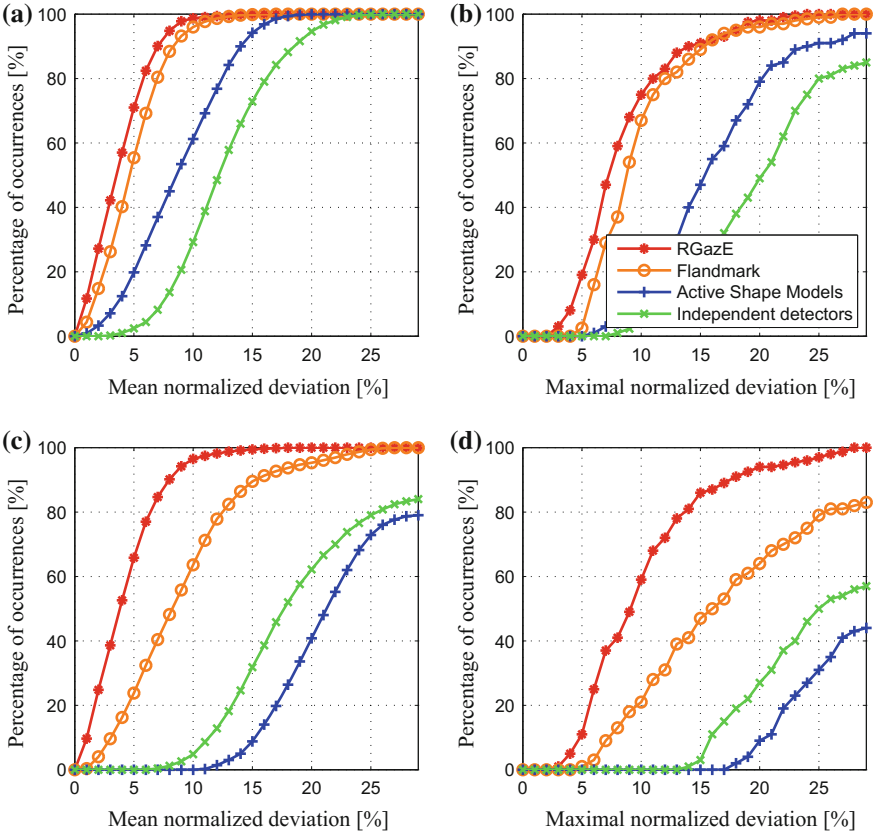


**Fig. 8** Cumulative histograms for the mean and the maximal normalized deviation shown for all competing detectors applied on video sequences with frontal (**a**, **b**) and profile (**c**, **d**) faces

tured output classification. The algorithm is based on a *Structured Output Support Vector Machine* (SO-SVM) classifier for the supervised learning of the parameters for facial points detection from examples.

In comparison to our gaze following system, which uses a segmentation step for determining the pointwise location of the facial features, flandmark considers the centers of the detected ROIs as the point location of the eyes, nose, and mouth.

The mean and maximal deviation metrics were used to compare the accuracy of the four tested detectors with respect to the ground truth values available from the benchmark databases. Especially for the evaluation of the computation time, the algorithm has also been tested on a mobile robotic platform.

The cumulative histograms of the mean and maximal normalized deviation are shown in Fig. 8 for frontal and profile faces. In all cases, the proposed estimator delivered an accuracy value superior to the ones given by the competing detectors. If the accuracy difference between our algorithm and Flandmark is relatively low for the case of frontal faces, it actually increases when the person's face is imaged from a profile view.

An interesting observation can be made when comparing the independent detectors with the ASM one. Although the ASM outperforms independent facial feature extraction on frontal faces, it does not perform well when the human subjects are viewed from the lateral. This is due to the training nature of the ASM, where the input training data is made of points spread on the whole frontal area (e.g., eyes, eyebrows, nose, chin, cheeks, etc.).

## 6 Conclusion

In this paper, a robust facial features detector for 3D gaze orientation estimation has been proposed. The solution is able to return a reliable gaze estimate, even if only a partial set of facial features is visible. The paper brings together algorithms for facial feature detection, machine learning, and control theory. During the experiments, we investigated the system's response and compare the results to ground truth values. As shown in the experimental results section, the method performed well with respect to various testing scenarios. As future work, the authors consider the possibility of extending the framework for the simultaneous gaze estimation of multiple interlocutors and the adaptation of algorithm with respect to the robot's egomotion.

**Acknowledgements** We hereby acknowledge the structural founds project PRO-DD (POS-CCE, O.2.2.1., ID 123, SMIS 2637, ctr. No 11/2009) for providing the infrastructure used in this work.

## References


1. Scassellati, B.: Theory of mind for a humanoid robot. Auton. Robots **12**(1999), 13–24 (2002)
2. Langton, S.R.H., Honeyman, H., Tessler, E.: The influence of head contour and nose angle on the perception of eye-gaze direction. Atten. Percept. Psychophys. **66**(5), 752–771 (2004)
3. Chutorian, E., Trivedi, M.: Head pose estimation in computer vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 607–629 (2009)
4. Gee, A., Cipolla, R.: Determining the gaze of faces in images. Image Vis. Comput. **12**(10), 639–647 (1994)
5. Horprasert, T., Yacoob, Y., Davis, L.: Computing 3-d head orientation from a monocular image sequence. In: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, pp. 242–247, Oct 1996
6. Kaminski, J., Knaan, D., Shavit, A.: Single image face orientation and gaze detection. Mach. Vis. Appl. **21**(3), 85–98 (2009)
7. Nikolaidis, A., Pitas, I.: Facial feature extraction and pose determination. Pattern Recogn. **33**(11), 1783–1791 (2000)
8. Canton-Ferrer, C., Casas, J., Pardas, M.: Head orientation estimation using particle filtering in multiview scenarios. In: Multimodal Technologies for Perception of Humans, vol. 4625, pp. 317–327. Springer, Berlin (2008)
9. Pantic, M., Tomc, M., Rothkrantz, L.: A hybrid approach to mouth features detection. In: 2001 IEEE International Conference on Systems, Man, and Cybernetics, vol. 2, pp. 1188–1193 (2001)
10. Skodras, E., Fakotakis, N.: An unconstrained method for lip detection in color images. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1013–1016 (2011)
11. Gonzalez-Ortega, D., Diaz-Pernas, F., Martinez-Zarzuela, M., Anton-Rodriguez, M., Diez-Higuera, J., Boto-Giralda, D.: Real-time nose detection and tracking based on adaboost and optical flow algorithms. In: Intelligent Data Engineering and Automated Learning, vol. 5788, pp. 142–150. Springer, Berlin (2009)
12. Werghi, N., Boukadia, H., Meguebli, Y., Bhaskar, H.: Nose detection and face extraction from 3d raw facial surface based on mesh quality assessment. In: 36th Annual Conference on IEEE Industrial Electronics Society, pp. 1161–1166 (2010)
13. Hansen, D., Ji, Q.: In the eye of the beholder: a survey of models for eyes and gaze. IEEE Trans. Pattern Anal. Mach. Intell. **32**(3), 78–500 (2010)
14. Valenti, R., Sebe, N., Gevers, T.: Combining head pose and eye location information for gaze estimation. IEEE Trans. Image Process. (2011)
15. Ke, L., Kang, J.: Eye location method based on haar features. In: 2010 3rd International Congress on Image and Signal Processing, vol. 2, pp. 925–929 (2010)
16. Hassaballah, M., Kanazawa, T., Ido, S.: Efficient eye detection method based on grey intensity variance and independent components analysis. Comput. Vis. IET **4**(4), 261–271 (2010)
17. Reale, M., Canavan, S., Yin, L., Hu, K., Hung, T.: A multi-gesture interaction system using a 3-d iris disk model for gaze estimation and an active appearance model for 3-d hand pointing. IEEE Trans. Multimedia **13**(3), 474–486 (2011)
18. Beymer, D., Flickner, M.: Eye gaze tracking using an active stereo head. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 451–458 (2003)
19. Ronsse, R., White, O., Lefevre, P.: Computation of gaze orientation under unrestrained head movements. J. Neurosci. Methods **159**, 158–169 (2007)
20. Sung, J., Kanade, T., Kim, D.: Pose robust face tracking by combining active appearance models and cylinder head models. Int. J. Comput. Vis. **80**, 260–274 (2008)
21. Uřičář, M., Franc, V., Hlaváč, V.: Detector of facial landmarks learned by the structured output SVM. In: Csurka, G., Braz, J. (eds.) VISAPP '12: Proceedings of the 7th International Conference on Computer Vision Theory and Applications, vol. 1, pp. 547–556. SciTePress—Science and Technology Publications, Portugal, Feb 2012

22. Hotz, L., Neumann, B., Terzic, K.: High-level expectations for low-level image processing. In: KI 2008: Advances in Artificial Intelligence. Springer, Berlin (2008)
23. Ristic, D.: Feedback structures in image processing. Ph.D. dissertation, Bremen University, Institute of Automation, Bremen, Germany, Apr 2007
24. Grigorescu, S.M.: Robust machine vision for service robotics. Ph.D. dissertation, Bremen University, Institute of Automation, Bremen, Germany, June 2010
25. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments, University of Massachusetts, Amherst. Technical Report 07–49, Oct 2007