Head Yaw Estimation using Frontal Face Detector

José Mennesson[1], Afifa Dahmane[2], Taner Danisman[3] and Ioan Marius Bilasco[1]
[1]Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000, Lille, France
[2]Computer Science Department, USTHB University, Algiers, Algeria
[3]Department of Computer Engineering, Akdeniz University, Antalya, Turkey

Keywords: Head Pose, Yaw Angle Estimation, Frontal Face Detector, 3D Head Rotation, Ellipsoid.

Abstract: Detecting accurately head orientation is an important task in systems relying on face analysis. The estimation of the horizontal rotation of the head (yaw rotation) is a key step in detecting the orientation of the face. The purpose of this paper is to use a well-known frontal face detector in order to estimate head yaw angle. Our approach consists in simulating 3D head rotations and detecting face using a frontal face detector. Indeed, head yaw angle can be estimated by determining the angle at which the 3D head must be rotated to be frontal. This approach is model-free and unsupervised (except the generic learning step of VJ algorithm). The method is experimented and compared with the state-of-the-art approaches using continuous and discrete protocols on two well-known databases : FacePix and Pointing04.

1 INTRODUCTION

Head yaw estimation is one of the key components for estimating the orientation of the face. Yaw rotation is a rotational component of 3DoF (yaw, pitch, roll) and defined as rotation about the vertical z-axis. Accurate estimation of the yaw angle has particular importance in several domains such as analysing drivers' driving behaviour, video surveillance and facial analysis (e.g. face recognition, face detection, face tracking, gender recognition and age estimation). In such contents, the head pose can quickly change from frontal upright position and generate out-of-plane rotations. Out-of-plane rotations (yaw and pitch) are more challenging than in-plane rotations (roll) as they largely influence the performances of underlying analysing systems. Therefore, there is a need to recover the head pose information when the face is not directly facing to the camera.

Common typologies of approaches in head pose estimation are summarized in (Murphy-Chutorian and Trivedi, 2009). These methods can be roughly grouped into two categories: Model-based and appearance-based methods. Model-based methods use 3D information for head pose estimation while appearance-based methods infer the relationship among the 3D points and their projections on 2D. According to the survey, the performance of head-pose estimation systems significantly degrades in out-of-plane rotations than in-plane rotations. As indicated in (Jung and Nixon, 2010) existing 2D models are not effective since they do not represent the curved surfaces and they do not cope well with large variations in 3D. Therefore, they are not robust to out-of-plane rotations.

(Kwon et al., 2006) used cylindrical head model to recover the 3D head pose information from a set of images. They first detect the face and then generate an initial template for the head pose and cylindrical head model. They dynamically update this template to recover the problems in tracking. Head motion is tracked based on optical flow in sequential images. For this reason, this method requires a set of sequential frames and cannot be applied to single images.

(Narayanan et al., 2014) studied yaw estimation using cylindrical and ellipsoidal face models. Their study on ellipsoidal framework provides MAE between 4° and 8° outperforming manifold-based approaches on FacePix dataset. Methods dealing with yaw detection become more and more complex by combining various techniques requiring most of the time specific machine learning tools. This kind of methods can suffer from the fact that combining methods that are not completely precise and robust individually, may eventually result in a lack of global precision and tuning the system becomes complicated

Mennesson, J., Dahmane, A., Danisman, T. and Bilasco, I.
Head Yaw Estimation using Frontal Face Detector.
DOI: 10.5220/0005711905170524
In Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 4: VISAPP, pages 517-524
ISBN: 978-989-758-175-5

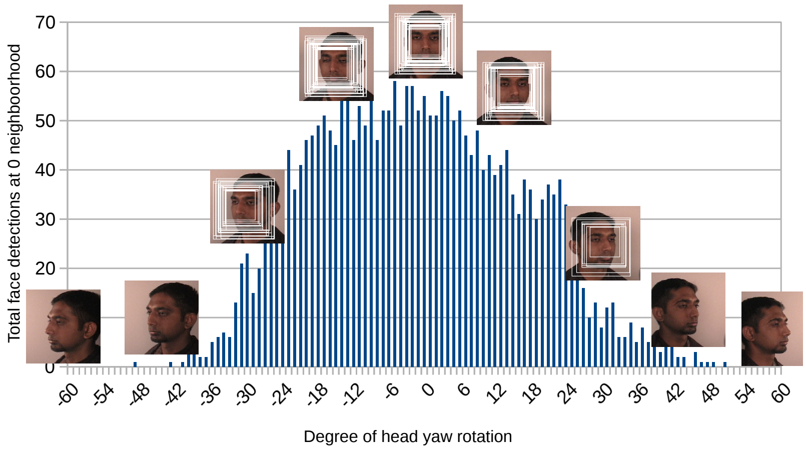

Figure 1: Changes in total VJ face detections for different yaw angles at zero neighborhood. Each bar shows the number of detections obtained from different scales of the VJ detector for specified degree of rotation.

as compromises between the drawbacks of underlying methods might not be straight-forward. Besides, the learning process in such methods can conduct to specific solutions that are tuned for a specific training set and may behave differently in more generic context such as in-the-wild unconstrained settings.

We discuss apart approaches that are dealing with the discrete yaw estimation and continuous yaw estimation. Discrete pose represents the orientation at fixed intervals (e.g., $\pm 15°$) and they are only capable of estimating coarse pose space. On the other hand continuous estimations can handle fine (precise) poses. In each of this two categories we are considering approaches that study the yaw estimation. The yaw discrete estimation approaches can be considered, most of the time, as regular classification problems where each discrete pose specific training and analysis is done. In the yaw continuous estimation regression methods and/or tracking mechanisms initialized from known/predefined frontal are commonly used.

Using a similar idea as (Danisman and Bilasco, 2015) which estimate roll pose estimation, we focus on yaw pose estimation from still images using frontal Viola-Jones (Viola and Jones, 2001) face detector using a two-stage approach. The frontal Viola-Jones detector responds positively to faces that are nearly frontal. However, the number of candidate face regions responding to the Viola-Jones detector is generally more important for frontal faces than for non-frontal faces and falls zero for profile faces. Figure 1 shows total VJ face detections for different yaw angles at zero neighbourhood. It is clear that the maximum of detected frontal faces is obtained with a yaw angle near zero degree (a frontal face). When one moves away from frontal face, the number of detected frontal faces decreases to zero with a Gaussian decay. In order to take advantage of this fact, we present the same face under different perspectives corresponding to candidate yaw angles to the Viola-Jones detector. This idea is illustrated in Figure 2 where a face (whose yaw angle is $-30°$) is projected onto a 3D ellipsoid and rotated artificially from $-90°$ to $90°$. As a nearly-frontal face still activates the Viola-Jones detector, we consider the whole span of perspectives angles that responded positively to Viola-Jones in order to estimate the yaw by considering several acceptance criteria : continuous detection over a given yaw range, average over all positive candidate angles. Another says is that regardless of the yaw angle of the analysed face, we apply a set of transformation to the face and we study how the VJ detector behaves with regard to applied transformation. The inverse yaw transformation yielding the best behaviour with regard to a given detector (Viola-Jones frontal face detector in our case) and a given criteria (number of consecutive detections, for instance) is a fairly good candidate for the yaw detection. The main idea is not to characterize the object, but the inverse transformation applied on the object in order to bring the object in a state where expected behaviour is met. The strength of this approach resides also in the use of a well-known method largely studied and used in the literature. In addition, this method does not require specific learning step other than the one used for training the underlying frontal Viola-Jones face detector.

The remaining part of this paper is organized as follows. Section 2 discuss the existing typologies of yaw estimation approaches. Section 3 presents our approach including head segmentation, 2D projection

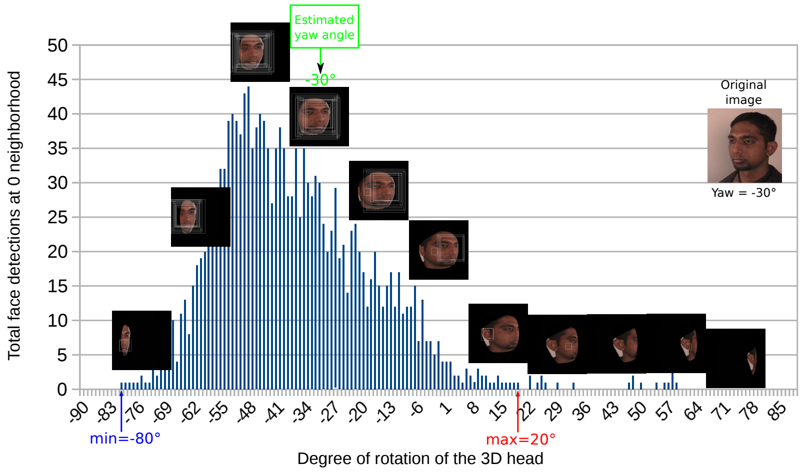

Figure 2: Changes in total VJ face detections for different yaw angles at zero neighborhood. Each bar shows the number of detections obtained from different scales of the VJ detector for specified degree of rotation of the ellipsoid.

on ellipsoid, face detection and head yaw estimation. Section 4 explains the experimentation datasets and presents evaluation results on a continuous and discrete perspective.

2 STATE OF THE ART

A common functional taxonomy covering the large variations in face orientation estimation studies can be found in (Murphy-Chutorian and Trivedi, 2009). In the current work, we selected representative approaches that report results on yaw estimation on public datasets such as FacePix (Black et al., 2002) and Pointing04 (Gourier et al., 2004). In particular, we selected the works giving the best results in (Murphy-Chutorian and Trivedi, 2009) and in (Dahmane et al., 2014) on these two databases. Among the numerous approaches for head pose estimation, one can cite :

Model-based approaches include geometric and flexible model approaches. For example, (Narayanan et al., 2014) propose to use cylindrical and ellipsoidal face models to estimate yaw angle. (Tu et al., 2007) perform head pose estimation based on a pose tensor model.

Regression based approaches consider pose angles as regression values. In (Stiefelhagen, 2004), a neural network is trained for fine head pose estimation over a continuous pose range. (Gourier et al., 2007) propose to train an associative neural network using data from facial feature locations while (Ji et al., 2011) use convex regularized sparse regression.

Manifold Embedding approaches produce a low dimensional representation of the original facial features and then learn a mapping from the low dimensional manifold to the angles. In (Balasubramanian et al., 2007), biased manifold embedding for supervised manifold learning is performed and (Liu et al., 2010) propose a K-manifold clustering method, integrating manifold embedding and clustering.

Symmetry-based approaches use the symmetrical properties of the head to estimate yaw angle as in (Dahmane et al., 2014).

3 OUR APPROACH

Our approach is based on the classical Viola-Jones (Viola and Jones, 2001) frontal face detector. Indeed, this detector is able to detect frontal faces in images. Assuming that a frontal face is a face with a head yaw angle $\in$ [-45°;45°], head yaw can be estimated by artificially turning head in 3D space from -90° to 90° about the vertical z-axis and detect at each step if there is a frontal face in the image plane. Head yaw angle can be estimated by determining the angle at which the 3D head must be rotated to be frontal. Using this assumption, our method is highlighted in Figure 3.

First, the head must be segmented as much accurately as possible because our method depends highly on this step. The head can be segmented for example using GrabCut (Rother et al., 2004) or a skin detector (Zaidan et al., 2014). In this paper, GrabCut is chosen because it is a well-known and widely used method to segment images. Then, the head is cropped and projected on a 3D sphere to simulate the 3D shape of the head. It can be done using complex methods which estimate the real shape of the face as in (Blanz and Vetter, 1999). In order to keep the method as efficient as possible, we choose to project the face on an ellip-

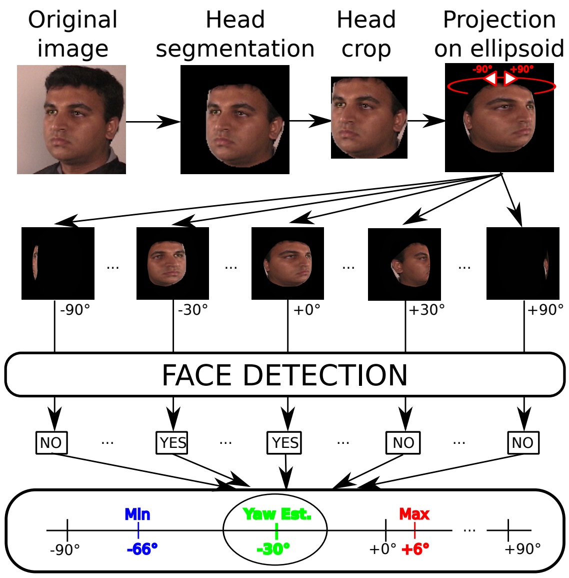

Figure 3: Overview of our method.

soid. The ellipsoid is then turned from $-90°$ to $90°$ about the z-axis to simulate a 3D head face turn and, at each angle, face detection is performed. Results of detection and non-detection are stored in a binary vector. After a filtering step, we consider the minimum and the maximum angles where the face has been detected. Several criteria can be applied (consecutive frames detection, mode, mean, etc.), but here we retain the mean of the detection angles.

3.1 Head Segmentation

Our approach depends on the segmentation of the face. In fact, if there is unbalanced background on the right and on the left of the face, this will skew the estimation of the head yaw. In order to solve this problem, the face must be accurately segmented before the projection on the ellipsoid.

One of the most well-known and robust algorithm to segment objects in images is GrabCut (Rother et al., 2004). This algorithm needs to be initialized with an area (a rectangle for example) which probably corresponds to the foreground (here, it is the head of the person). Everything outside the rectangle is considered as the background.

In this paper, we assumed that the face is centered. So, the area given to the GrabCut is a rectangle in the center of the image. The size of the rectangle is half of the size of the image. This constraint can be lowered if we assume that the yaw angle $\in [-45°;45°]$ and an initial VJ can be used instead. Outside this range, other versions of VJ or other techniques must be employed to find the first candidate face boundary box. RACV library (Auguste, 2014) is used to optimize the head rectangle given to GrabCut using the skin proportion therein. Finally, the convex hull of the head is computed.

3.2 Projection on Ellipsoid

In order to simulate the rotation of the face, the segmented face is projected on an ellipsoid. In this paper, we have used OpenGL (Open Graphics Library) as in (Aissaoui et al., 2014). The height of the ellipsoid is 1.5 times the width in order to approximate the general proportions of faces. The ellipsoid is rotated from $-90°$ to $90°$ about the z-axis to cover all the angles where the frontal face can be detected. The incremental parameter of the rotation angle, denoted *step-factor*, can vary from $1°$ to $10°$ in order to accelerate the process. The effect of this parameter on the results is shown in section 4. Finally, the image we consider is a projection of the 3D ellipsoid on the image plane.

3.3 Frontal Face Detection

Whenever the face is rotated, a face detection is performed. In this paper, the classical VJ detector (Viola and Jones, 2001) who proposes to use Haar Feature-based Cascade Classifiers to detect faces, is used. Each frontal face detection is then stored in a binary vector.

In order to increase the chance of a matching size with the model for detection, we use a small step for resizing (scale factor=1.1). To eliminate false positives and get the proper face rectangle out of detections, the minimum neighbours parameter is set to 1.

3.4 Head Yaw Estimation

As it is said before, a binary vector containing the detection results is obtained. In Figure 4, these vectors are represented by binary images where white (resp. black) pixel depicts that a face is detected (resp. not detected). Each image corresponds to a different clip, each line corresponds to a head image with a particular yaw angle and each columns corresponds to a rotation angle of the ellipsoid. If we assume that there is no error in the detection results, head yaw angles can be estimated by considering the minimum and the maximum angles the head is detected. Indeed, the VJ detector being symmetric, it will detect the face as well if it is slightly rotated to the right or left (See Figure 4 - Clip7 and Clip25). A morphological opening of size 3 to suppress noise is performed. Estimated

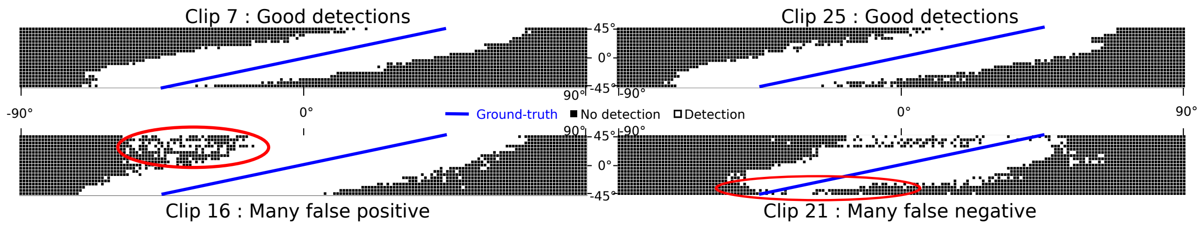

Figure 4: Detection results on FacePix dataset. White pixel: detection, black pixel: non detection, blue line: ground-truth.

head yaw is the angle between the minimum and the maximum angles at which the frontal face is detected.

Several solutions can be proposed :

- HYE1 : In order to make our approach robust to false positive detection (see Figure 4 - Clip16), connected detections are labeled. Then, only the largest connected component is considered to estimate head yaw angle as in Figure 5.

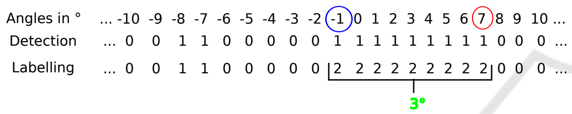

Figure 5: Labeling connected detections and choosing the largest connected component.

- HYE2 : The center of mass of the binary vector can be computed. This version is more robust to false negative detections (see Figure 4 - Clip21).

In the following section, we are conducting experiments to evaluate our methods. The objectives are to compare the results of our approach with the state of the art and to evaluate the influence of the parameters, initial yaw angles and different face shapes on the results.

4 EXPERIMENTS

Tests are made on FacePix (Black et al., 2002) and Pointing04 (Gourier et al., 2004) datasets. Two protocols (continuous and discrete) have been defined. The continuous like protocol aims to evaluate the capacity of the proposed method to offer fine grain characterization of yaw movements. The discrete like protocol considers the yaw movements problem as a classification problem, where 15° yaw interval classes are considered.

4.1 Datasets

In order to validate our approach, head yaw estimation is tested on several databases :

- FacePix Dataset

This dataset contains 3 sets of face images for each of 30 people included in the dataset. In Figure 6, we present frontal views of four persons that where included in Figure 4 experiments. We have selected this four persons in order to study closely the impact of eye glasses, long hair and skin color on the results. However, the whole dataset is considered for the results presented in the following. The first set contains 181 color face images which corresponds to yaw angle which vary across a range from $+90°$ to $-90°$. The second and the third sets contain just frontal faces and just concern illumination variations.

In this paper, we consider a subset of the first set which contains only faces which are in $[-45°;45°]$ (i.e. 2730 images) due to the frontal face detector used. Indeed, if too much information about the face is missing, the frontal face detector is no longer able to detect the face, even if the head is artificially rotated.

Figure 6: Several images from facePix dataset.

- Pointing04 Dataset

The head pose database contains 15 sets of images. Each set consists in two series of 93 images of the same person at different poses. There are 15 people in the database, wearing glasses or not and having various skin color. The pose, or head orientation is determined by 2 angles (yaw,pitch), which varies from $-90°$ to $+90°$. The first serie is used and only faces which are in $[-45°;45°]$ (i.e. 105 images) are considered due to the face detector used.

4.2 Evaluation

In this paper, the following experiments are conducted on continuous (i.e. evaluated every degree on FacePix dataset) and discrete space (i.e. evaluated every 15 degree on Pointing04 dataset).

4.2.1 Experiments on FacePix

In these tests, continuous experiments are conducted using well-known measures as the Mean Average Error (MAE), the Root Mean Square Error (RMSE) and the Standard Deviation (STD).

In Table 1, one can see that our approach provides comparable MAE with classical approaches of the state of the art. But our approach is model-free, person-free and unsupervised (except the generic learning step of VJ algorithm). One can also add that a significant drawback of manifold learning techniques is the lack of a projection matrix to treat new data points (Balasubramanian et al., 2007). The HYE2 method gives the best result on this database. The following charts show the results computed with this solution.

Table 1: FacePix : Comparison with the state of the art.

Method	MAE
(Balasubramanian et al., 2007)	
Biased Isomap	5.02°
Biased LLE	2.11°
Biased LE	1.44°
(Liu et al., 2010)	
Manifold clustering	3.16°
(Ji et al., 2011)	
Regression	6.1°
(Dahmane et al., 2014)	
Symmetry classification	3.14°
(Narayanan et al., 2014)	
CE	5.55°
Center CE	5.26°
Boundary CE	5.28°
Proposed Method	
HYE1	5.2°
HYE2	**4.8°**

In Figure 7, a boxplot which represents estimated yaw angles with regard to the groundtruth is shown. It is clear that most of estimated yaw angles are near the groundtruth. Widest errors are found nearby $-45°$ and $45°$. This is due to the VJ detector which has difficulties to detect frontal faces at these angles even if the face is turned artificially. Indeed, there can be a lack of face information (e.g. an eye is hidden) which can prevent the frontal face detection when the head is turned to -45° or 45°.

In order to evaluate our method in function of persons and head yaw angles, MAE measures are computed and shown in Figure 8 and 9. Figure 8 shows that most angles (between $-35°$ and $33°$) have a MAE under 5.5°. The worst results are obtained when head yaw angles are greater than 35° due to the frontal face detector properties.

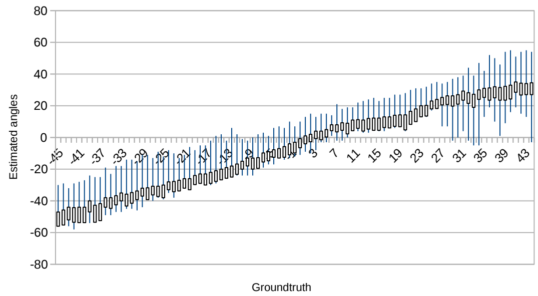

Figure 7: FacePix : boxplot of head yaw angle estimation.

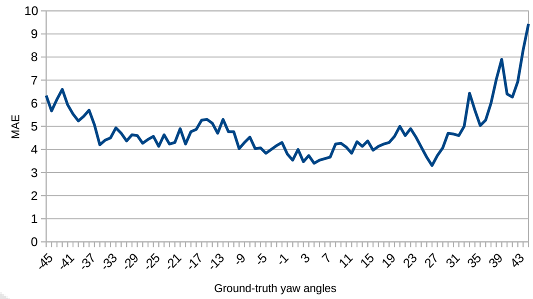

Figure 8: FacePix : Histogram of head yaw angles estimation MAE for ground-truth yaw angles.

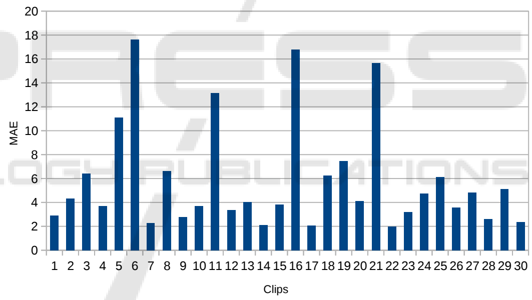

Figure 9: FacePix : Histogram of head yaw angles estimation MAE for Clip1-30.

In Figure 9, most Clips have low MAE ($< 8°$) except Clip5, 6, 11, 16, 21. Frontal faces in these clips are sometimes difficult to detect. For example, in Clip21 (see Figure 6 and 4), the person wear glasses, and in Clip11, the person closes almost completely her eyes, which prevent neat face detection.

We have seen in section 3.2 that our method depends of a parameter which controls the incremental step of the rotation of the face (i.e. *step-factor*). Figure 10 shows the evolution of MAE, MRSE and STD obtained on FacePix using several *step-factor* settings. One can easily see that the results vary only slightly if *step-factor* is less or equal to 4°. Hence for speed up purposes, we can divide the number of rotations of the 3D head by four while maintaining a good accuracy.

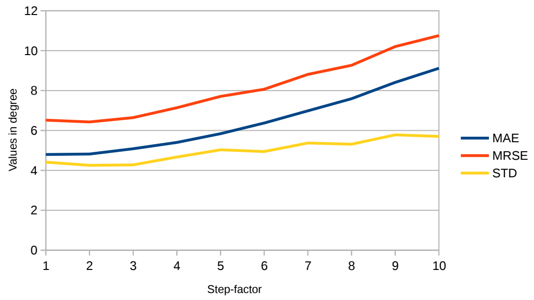

Figure 10: FacePix : Effect of *step-factor*(see section 3.2) on MAE, MRSE and STD.

4.2.2 Experiments on Pointing04

In this section, we consider discrete yaw classification. Hence, the results correspond to the number of well estimated yaw over the total number of head yaw to estimate. In order to cluster the estimated head yaw into discrete classes, it is associated to the nearest yaw class. There are 7 head yaw classes : $-45°$, $-30°$, $-15°$, $0°$, $15°$, $30°$, $45°$.

In Table 2, results on Pointing04 dataset are shown in terms of MAE and recognition rates. It is clear that our method outperforms most of the approaches of the state of the art. Again, other methods require a learning step for estimating head yaw while our method does not. Concerning this dataset, the HYE1 method outperform the HYE2 one, so the following charts show the results computed with this solution. This is due to many false detections which appears in several detection results for this dataset. As we said before, HYE1 is more robust to that problem.

Table 2: Pointing04 : Comparison with the state of the art.

Method	MAE	Reco. rates
(Stiefelhagen, 2004)	9.5°	52%
(Gourier et al., 2007)		
Human Performance	11.8°	40.7%
Associative Memories	10.1°	50.0%
(Tu et al., 2007)		
High-order SVD	12.9°	49.25%
PCA	14.11°	55.20%
LEA	15.88°	45.16%
(Ji et al., 2011)		
Regression	8.6°	-
(Narayanan et al., 2014)		
CE	7.2°	-
Center CE	6.82°	-
Boundary CE	6.9°	-
Proposed Method		
HYE1	**6.96°**	**63.81%**
HYE2	7.15°	59.05%

As it is shown in Figure 11, the worst recognition rates are obtained with angles $-45°$ and $45°$. Again, this is due to the properties of the VJ detector we used in this paper. The results for other angles are better or equal 60%.

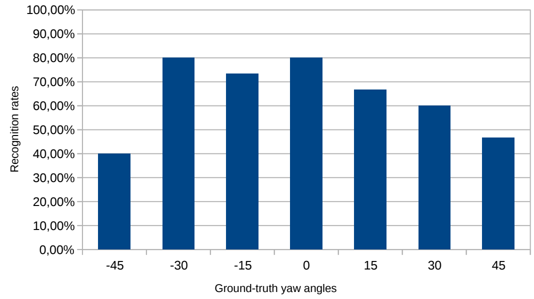

Figure 11: Pointing04 : Histogram of recognition rates for ground-truth yaw angles.

In Figure 12, one can note that for most persons, the results are good ($> 50\%$). Worst results are obtained for persons 2,7,8,9. The reason is a bad segmentation of the head of these persons due, among other things, to their hair which cover their ears.

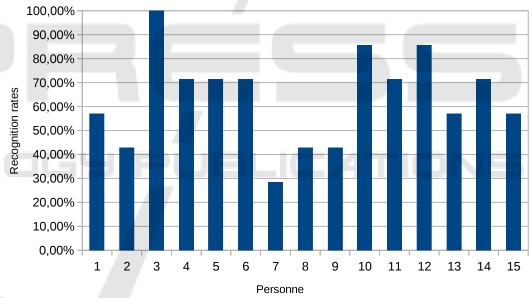

Figure 12: Pointing04 : Histogram of recognition rates for Personne1-30.

5 CONCLUSION

A new approach to estimate head yaw angle is presented in this paper. Face images are projected on 3D ellipsoid and artificially turned about the vertical z-axis. The goal is to determine the angle the head must be turned to be frontal. Frontal faces are detected using a well-known frontal face detector. Advantages are that it is a person-free, model-free and unsupervised approach. It is not a black box, so every parameters can be set easily. Experiments on well-known datasets have shown that this method gives comparable or better results than the state of the art. In future works, other methods to segment heads and detect frontal faces can be explored. Also, the 3D modeling of the 3D face could be improved using more accurate 3D shape than a simple ellipsoid. This method could also be extended to pitch angle estimation. In this paper, we propose two methods to estimate head yaw angles (HYE1 and HYE2). A hybrid method which uses HYE1 or HYE2 depending on the size of the connected component could be defined to take advantages of both approaches.

ACKNOWLEDGEMENTS

This work was supported by "Empathic Products", ITEA2 1105 and the authors thank IRCICA USR 3380 for financial support.

REFERENCES

Aissaoui, A., Martinet, J., and Djeraba, C. (2014). Rapid and accurate face depth estimation in passive stereo systems. *Multimedia Tools and Applications*, 72(3):2413–2438.

Auguste, R. (2014). Racv library can be downloaded from. https://github.com/auguster/libRacv/.

Balasubramanian, V., Ye, J., and Panchanathan, S. (2007). Biased manifold embedding: A framework for person-independent head pose estimation. In *CVPR, 2007. IEEE Conference on*, pages 1–7.

Black, J., Gargesha, M., Kahol, K., Kuchi, P., and Panchanathan, S. (2002). A framework for performance evaluation of face recognition algorithms. In *ITCOM, Internet Multimedia Systems II*, Boston.

Blanz, V. and Vetter, T. (1999). A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, pages 187–194, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Dahmane, A., Larabi, S., Bilasco, I. M., and Djeraba, C. (2014). Head Pose Estimation Based on Face Symmetry Analysis. *Signal, Image and Video Processing*.

Danisman, T. and Bilasco, I. M. (2015). In-plane face orientation estimation in still images. *Multimedia Tools and Applications*, pages 1–31.

Gourier, N., Hall, D., and Crowley, J. L. (2004). Estimating Face Orientation from Robust Detection of Salient Facial Features. In *Proceedings of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures*.

Gourier, N., Maisonnasse, J., Hall, D., and Crowley, J. (2007). Head pose estimation on low resolution images. In Stiefelhagen, R. and Garofolo, J., editors, *Multimodal Technologies for Perception of Humans*, volume 4122 of *LNCS*, pages 270–280. Springer Berlin Heidelberg.

Ji, H., Liu, R., Su, F., Su, Z., and Tian, Y. (2011). Robust head pose estimation via convex regularized sparse regression. In *ICIP, 2011*, pages 3617–3620.

Jung, S.-U. and Nixon, M. (2010). On using gait biometrics to enhance face pose estimation. In *Biometrics: Theory Applications and Systems, 2010 Fourth IEEE International Conference on*, pages 1–6.

Kwon, O., Chun, J., and Park, P. (2006). Cylindrical model-based head tracking and 3d pose recovery from sequential face images. In *Proceedings of the 2006 International Conference on Hybrid Information Technology - Volume 01*, pages 135–139, Washington, DC, USA. IEEE Computer Society.

Liu, X., Lu, H., and Li, W. (2010). Multi-manifold modeling for head pose estimation. In *ICIP, 2010*, pages 3277–3280.

Murphy-Chutorian, E. and Trivedi, M. (2009). Head pose estimation in computer vision: A survey. *PAMI, IEEE Transactions on*, 31(4):607–626.

Narayanan, A., Kaimal, R., and Bijlani, K. (2014). Yaw estimation using cylindrical and ellipsoidal face models. *Intelligent Transportation Systems, IEEE Transactions on*, 15(5):2308–2320.

Rother, C., Kolmogorov, V., and Blake, A. (2004). "grabcut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314.

Stiefelhagen, R. (2004). Estimating head pose with neural nnetwork - results on the pointing04 icpr workshop evaluation data. In *Pointing 2004 Workshop: Visual Observation of Deictic Gestures*.

Tu, J., Fu, Y., Hu, Y., and Huang, T. (2007). Evaluation of head pose estimation for studio data. In Stiefelhagen, R. and Garofolo, J., editors, *Multimodal Technologies for Perception of Humans*, volume 4122 of *LNCS*, pages 281–290. Springer Berlin Heidelberg.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. pages 511–518.

Zaidan, A., Ahmad, N., Karim, H. A., Larbani, M., Zaidan, B., and Sali, A. (2014). Image skin segmentation based on multi-agent learning bayesian and neural network. *Engineering Applications of Artificial Intelligence*, 32:136 – 150.