Learning symmetrical model for head pose estimation

Afifa Dahmane, Slimane Larabi, Chaabane Djeraba, Ioan Marius Bilasco

Learning Symmetrical Model for Head Pose Estimation

Afifa Dahmane[1,2], Slimane Larabi[1], Chabane Djeraba[2], Ioan Marius Bilasco[2]
[1] *USTHB University, BP 32 EL ALIA, Algiers, Algeria*
[2] *LIFL, UMR CNRS 8022, USTL University, 59658 Villeneuve dAscq, Lille, France*
{*afifa.dahmane,chabane.djeraba,marius.bilasco*}*@lifl.fr, slarabi@usthb.dz*

Abstract

This paper tackles the problem of head pose estimation which has been considered an important research task for decades. The proposed approach selects a set of features from the symmetrical parts of the face. The size of bilateral symmetrical area of the face is a good indicator of the Yaw head pose. We train a Decision Tree model in order to recognize head pose with regard to the areas of symmetry. The approach does not need the location of interest points on face and is robust to partial occlusion. Tests were performed on a different dataset from that used for training the model and the results demonstrate that the change in the size of the regions that contain a bilateral symmetry provides accurate pose estimation.

1. Introduction

In this paper, we address the problem of head pose estimation from digital images which consists of locating a person's head and estimating the orientation of its three degrees of freedom (*Yaw*, *Pitch* and *Roll*). Robust and accurate head pose estimation is a classic problem in computer vision. It has been widely used in many applications such as video conferencing or driver monitoring. It is often linked with visual gaze estimation and provides a coarse indication of the gaze in situations where the eyes of a person are not visible. The estimation of the head pose relies on the pose similarity assumption (different people at the same pose look more similar than the same person at different poses). Many approaches based on facial features are proposed to deal with head pose estimation. However the obvious difficulty lies in detecting outlying or missing features in situations where facial landmarks are obscured. Also, low resolution imagery make it difficult to precisely determine the feature locations.

Over the years, many techniques have been proposed for head pose estimation from a monocular camera. They can be categorized in three different classes:

Model-based approaches; These approaches may be geometric [14], where a set of specific facial features such as eyes, mouth and nose are used to estimate the head pose. Also, model-based methods can be a flexible model [23], where a non-rigid model is fit to the image such that it conforms to the facial structure.

Appearance-based approaches; Instead of concentrating on the specific facial features, the appearance of the entire head image is modelled and learned from the training data. These approaches formulate the head pose estimation as a pattern classification problem. Several works have used a range of classifiers such as SVM [9] or randomized ferns [1]. These approaches can also estimate the head pose using a template [12] or by directly comparing facial images with a set of template images [2]. Also, several regressors are possible such as Convex Regularized Sparse Regression (CRSR) [7] and Gaussian Progress Regression (GPR) [15].

Hybrid approaches; The above two approaches may be combined [20]. The temporal information could also be introduced to improve the head pose estimation using the results of the head tracking [19].

Each category approaches has specific limits. Appearance-based approaches suffer from information about identity and lighting which are contained in the face appearance and although model-based methods are fast and simple, they are sensitive to occlusion and usually require high resolution images which may be not available in many applications such as driver monitoring or video surveillance.

The discrimination between head orientations is based upon two cues: the deviation of the head shape from bilateral symmetry, and the deviation of the nose orientation from the vertical [22]. Therefore, we assume that head pose is more related to the geometry of the face images and the symmetry of the face is a good indicator about the geometric configuration and therefore about the pose of the head. The symmetry property of the head have been used as a visual intent indicator in [11] for people with disabilities. Symmetry based illumination model proposed in [6] is based on a three features (the two eyes and the nose tip). For every combination of two eyes and a nose, head pose is computed using a weak geometry projection and internally calibrated camera. The face pose (roll and yaw angles) are estimated from a single uncalibrated view in [21] where the symmetric structure of human face is exploited taking the mirror image of a test face image as a virtual second view and based on the extraction of facial feature points of the test and its mirror image and their matching. A reliable and recent survey in head pose estimation can be found in [13].

This paper presents a geometrical approach based on the symmetrical properties of the face to achieve head pose estimation. The proposed approach uses features that may be extracted even if the head is relatively far from the camera and do not need the location of interest points on face since the facial symmetry appears easily.

The paper is organized as follows. First, we provide the methodology used for the estimation of the head pose using the symmetrical parts of the face in Section 2 and 3. In Section 4, the proposed approach is evaluated using the FacePix database [3] and the results of the head pose estimation are discussed. Also, tests with video sequences from the Boston University head pose dataset [18], are performed. Finally, we conclude and discuss the potential future work in Section 5.

2. Features

Face symmetry on images provides knowledge about the geometrical configuration of the face. As the low resolution images prevent the use of approaches which need detection of facial features, we introduce the use of symmetrical model of the face to deal with head pose estimation. We demonstrated in [5] that the amount (lengths and widths) of symmetrical parts on face are good and robust features for head pose estimation. In the present work, we will learn a model from the features that we extract from the bilateral symmetry of the face in order to estimate the Yaw orientation of the head.

2.1 Properties of symmetrical parts on face

We will use the bilateral symmetry of the face to deal with the head pose estimation problem [5]. When the face is in front of camera, the symmetry between its two parts (left, right) appears clearly and the line which passes between the two eyes and nose tip defines the symmetry axis. However, when the head performs a motion, for example, a yaw motion, this symmetry evolves.

We obtain the highest ratio when the head is facing the camera. The greater the movement angle, the lesser the number of symmetrical pixels.

2.2 Extraction of symmetries from head images

Many approaches have been proposed to solve the reflectional symmetry detection problem in images [10, 4, 17]. These methods find the symmetrical parts using some constraints, initial assumptions or are based on features that are difficult to extract when the face is far from the camera. However, the method proposed by Stentiford [16] permits the extraction of symmetries from 2D facial images without manual intervention or prior specification of the features that are associated to these symmetries. So, we detect symmetry basing on this method.

We implemented the idea proposed by Stentiford [16] with some improvements in order to reduce the time processing and to get a higher accuracy regarding the position of the vertical symmetry axis within an image. The correct location of the symmetry axis depends on the location of the head. To build the model, we manually locate the head for good accuracy in automatically detecting the symmetry axis. We start by locating a tight bounding box around the head. We draw an ellipse inside and then we locate all symmetrical points in the region of the eyes. The location of the eyes is estimated as follows: 1/2 of the face height and a few pixels from the top of the face. However, for testing video sequences, the process is fully automatic (what we see in the experiments section).

3. Head pose estimation

In order to determine the amount of yaw motion, a Decision Tree classifier is trained using the relative features extracted from the symmetrical parts according to the amount of motion. To increase performance of prediction, we use the Alternating Decision Tree that is based on boosting [8]. The tree alternates between prediction nodes and decision nodes. The root node is a prediction node and contain a value. The sum of the prediction values crossed when following all paths for which all decision node are true, is used to classify a given instance.

We have used the FacePix database [3] for our learning and evaluation. The set of head pose images represents the angles for which the symmetry axis is properly detected. The poses vary from -45° (left) to +45° (right).

We start by extracting features from the region of interest such as the ratio of the symmetrical pixels related to the total number of facial pixels, as well as the width of the hull that includes the pixels. These features will constitute the feature vector. Finally, we construct the model from the vectors derived from images of several people caught in different poses. These images represent seven poses associated with the following Yaw angles: -45°, -30°, -15°, 0°, 15°, 30° and 45°. The classifier estimating the head orientation has four discrete classes which represent seven poses: class 1: -45° and 45°, class 2: -30° and 30°, class 3: -15° and 15° and class 4: 0°. The right and left poses are gathered in the same class as they contain the same symmetry and, therefore, the same information. Figure 1 shows the seven poses with their features.

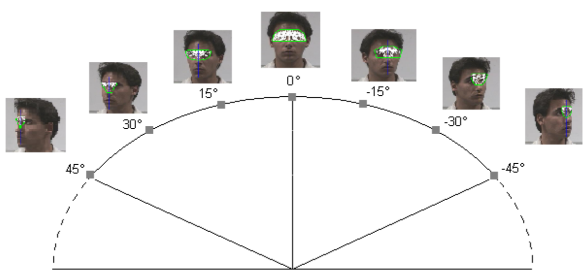

Figure 1. The seven Yaw poses

4. Experimental results and discussion

We evaluate the obtained model in order to validate the features extracted from symmetry. Also, we use a different dataset to perform tests. The results are presented below.

4.1 Learning

We use the FacePix database [3] to build the head pose model and to evaluate it. The FacePix database consists of three sets of face images. The set with pose angle variations is composed of 181 face images, representing angles from -90° to +90° at 1 degree increments, of 30 different subjects. Among the 181 poses, we use 7 poses (poses varying from -45° to +45° with 15° step) because when exceeding this interval, the bilateral symmetry disappears from the image plane.

We split the data into 6 equal subsets and performed 6-fold crossvalidation. In each run, 5 subsets are used as the training set and the other subset is used as a test set. The subjects in the training and test set are completely distinct since each subject is taken once.

In order to classify any input image into one of the four discrete head poses, we learn an Alternating Decision Tree. We build a model and evaluate it. Figures 2 shows the classification rates using Alternating Decision Tree for the 4 pose classes. We obtain an weighted average of correct classification equal to 81.4%. The classification rate for class 3 witch corresponds to the two poses : -15° and 15°, is the lowest, because often these poses are closer to that of class 2 and class 4 The other poses are better classified, their classification rate exceeds 80%.

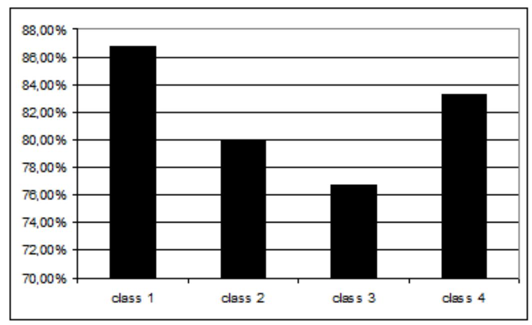

Figure 2. Classification rates for the 4 pose classes

4.2 Testing

In addition to the validation experiments, we test the decision tree on a different dataset. The model was used to estimate the head pose in videos including sequences from the Boston University head pose dataset [18]. Since we do not have information about the localization of the face, we use the Viola Jones detector, which provides the area for features extraction. To infer a continuous head pose estimation, we interpolated the three neighbouring poses obtained from the neighbouring frames) using a 3-order B-spline. Sample frames from a video are shown in figure 3. The symmetry area is superposed on the original frames and the estimated poses are compared with ground truth. The mean absolute angular error for this sequence is 4.47°.

The main advantage of the method is that the calculation can start at any pose, without any initialisation, since the head and the symmetry axis are automatically detected for poses between -45° and +45°. The symmetry is properly extracted for this interval even if the image is noisy or low resolution. Thus, results from video sequences provide accurate estimation of the motion. The mean absolute angular error for the Boston University head pose dataset (Yaw motion sequences) is 6.49°.

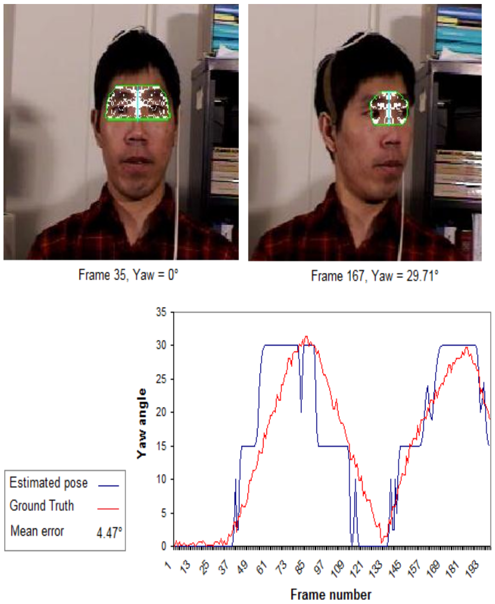

Figure 3. Sample frames from a video sequence (Boston University head pose dataset)

5. Conclusion

We have presented a new approach to perform head pose Yaw estimation. We use a set of features extracted from the symmetry of the face to learn a model for Yaw angles. These features may be extracted even if the head is far from the camera and do not need the detection of special facial landmarks. The results obtained by our approach have been evaluated using the FacePix head pose database where classification accuracy reached 81.4% and tested using video sequences from the Boston University head pose dataset. We now have a valuable model that we can deploy on data in the real world. In our future work, we will explore temporal correlation obtained from the head tracking to extend the range of motion and we will add features to classify both Yaw and Pitch angles.

References

[1] B. Benfold and I. Reid. Colour invariant head pose classification in low resolution video. *BMVC*, 2008.

[2] D. Beymer. Face recognition under varying pose. *CVPR*, 1994.

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

[4] H. Cornelius and G. Loy. Detecting rotational symmetry under affine projection. In *ICPR*, pages 292–295, 2006.

[5] A. Dahmane, S. Larabi, and C. Djeraba. Detection and analysis of symmetrical parts on face for head pose estimation. In *ICIP*, 2010.

[6] M. Gruendig and O. Hellwich. 3d head pose estimation with symmetry based illumination model in low resolution video. In *Lecture Notes in Computer Science*, volume 3175, pages 45–53. Springer, 2004.

[7] F. S. Z. S. Y. T. Hao Ji1, Risheng Liu. Robust head pose estimation via convex regularized sparse regression. *ICIP*, 2011.

[8] G. Holmes, B. Pfahringer, R. Kirkby, E. Frank, and M. Hall. Multiclass alternating decision trees. In *ECML*, pages 161–172. Springer, 2001.

[9] J. Huang, X. Shao, and H. Wechsler. Face pose discrimination using support vector machines (svm). In *ICPR*, volume 1, pages 154–156, 1998.

[10] G. Loy and J. Eklundh. Detecting symmetry and symmetric constellations of features. In *ECCV*, 2006.

[11] T. Luhandjula, E. Monacelli, Y. Hamam, B. van Wyk, and Q. Williams. Visual intention detection for wheelchair motion. In *International Symposium on Visual Computing (ISVC)*, pages 407–416, 2009.

[12] J. J. C. T. A. Meltem Demirkus, Boris Oreshkin. Spatial and probabilistic codebook template based head pose estimation from unconstrained environments. *ICIP*, 2011.

[13] E. Murphy-Chutorian and M. M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 31(4):607–626, 2009.

[14] Y. Pan, H. Zhu, and R. Ji. *3-D Head Pose Estimation for Monocular Image*. Fuzzy Systems and Knowledge Discovery. Springer, 2005.

[15] A. Ranganathan and M.-H. Yang. Online sparse matrix gaussian process regression and vision applications. *ECCV*, 2008.

[16] F. Stentiford. Attention based facial symmetry detection. In *International Conference on Advances in Pattern Recognition (ICAPR)*, 2005.

[17] C. Sun and D. Si. Fast reflectional symmetry detection using orientation histograms. *Real-Time Imaging*, 5, 1999.

[18] R. Valenti and T. Gevers. Robustifying eye center localization by head pose cues. In *IEEE conference on Computer Vision and Pattern Recognition*, 2009.

[19] R. Valenti, Z. Yucel, and T. Gevers. Robustifying eye center localization by head pose cues. In *CVPR*, 2009.

[20] T. Vatahska, M. Bennewitz, and S. Behnke. Feature-based head pose estimation from images. In *7th International Conference on Humanoid Robots*, 2007.

[21] S. D. Vinod Pathangay and T. Greiner. Symmetry-based face pose estimation from a single uncalibrated view. *8th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2008)*, pages 1–8, 2008.

[22] H. R. Wilson, F.Wilkinson, L. Lin, and M. Castillo. Perception of head orientation. *Vision Research*, 40(5):459–472, 2000.

[23] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models. In *CVPR*, volume 2, pages 535–542, 2004.