

Head Pose Estimation from Depth Map

Abstract- In this paper we propose a new method for head pose estimation using the depth sensor Kinect. Our approach infers the head pose based on the symmetry or asymmetry computed on the depth map of the face. This approach does not require the location of the nose or any other feature on the face such as has been done in many works, but uses only the depth map of the face. Two features are proposed for characterizing the pan, roll and tilt rotation of the head. The first one, measures the area of nearest region of the face relatively to the face area. The second one, it concerns the symmetry on the depth map of the face. Experiments are conducted on our acquired data. The obtained results are promising and demonstrate the usefulness of the proposed features.

Keywords-Depth Map; Head pose estimation; Symmetry

I. INTRODUCTION

Depth estimation for faces or body is an important problem that has been largely studied in many computer vision applications such as face recognition, face and gesture recognition, face animation and analysis. This interest is due to the invariance of the depth map to illumination changes in the scene and thus the accuracy of head or body pose estimation is guaranteed.

In this paper we propose a new method for head pose estimation using the depth sensor Kinect. Our approach is based on the symmetry or asymmetry computed on the depth map of the face. Two features are proposed for characterizing the pan, roll and tilt rotation of the head. The first one concerns the relative area of nearest region of the face to the Kinect. The second one concerns the asymmetry on the depth map of the face. We will see that relatively to the computed axis located along the head, the asymmetry is in relationship with the pan angle. Our approach does not need the location of the nose or any other feature on the face such as it has been done in many works, but uses only the depth map of the face. In addition, we assume that face is located using the image and applying Viola and Jones detector.

Experiments are conducted on our acquired data. The obtained results demonstrate the usefulness of the proposed features. In the next section, a review of relevant works to this topic is given. Section 3 is devoted to our approach. First, we explain our basic principle; the algorithm is given. The final section contains the conducted tests and the obtained results.

II. PREVIOUS WORKS

Head pose estimation has been well studied but remains one of the open problems of computer vision applications due to the external environment which influences the image quality. Many states of the arts were proposed where we can see the difficulty of this problem [2].

To improve the performances of head pose estimation, the depth data that may be acquired by stereovision or the depth camera (Kinect) has been added and some approaches are proposed [3, 4, 5, 6, 7].

Due to the availability of depth-sensing technologies, many works were proposed by using the depth for solving the problem of head pose estimation.

Many assumptions are made: the nose is visible [8], the head is the only object present in the field of view [2], and the 2D image data is combined with the depth [9], [10].

In the recent work [1], authors deal with depth images where other parts of the body might be visible and therefore need to discriminate which image patches belong to the head and which don't using Discriminative Random Regression Forests. In other recent work [11], a reference depth image of a human subject is obtained. The method searches the 6-dimensional pose space to find a pose from which the head appears identical to the reference view. This search is formulated as an optimization problem whose objective function quantifies the discrepancy of the depth measurements between the hypothesized views to the reference view.

Instead of other methods, we estimate head pose directly using some features extracted from the depth map without any constraint on the scene. In addition, our method deals with any head pose and there is no constraint for pan, roll or tilt rotations such as it has been assumed in the literature.

III. OUR APPROACH

A. Basic principle

Once the face is located on the image acquired by the Kinect by using the Viola and Jones detector [12], the depth map associated to head is divided into a set of plans Π_i of depths. Each plan Π_i is defined as the set of pixels in the depth map having a value of depth in a specific interval $I_i = [Val_i^{\min}; Val_i^{\max}]$.

Whatever the position of the head relatively to the Kinect, the distance between minimal and maximal values of depths is

divided into six equidistant intervals. Figure 1 illustrates the depth map with six plans colored, the closest plan is colored in red and the far one in grey.

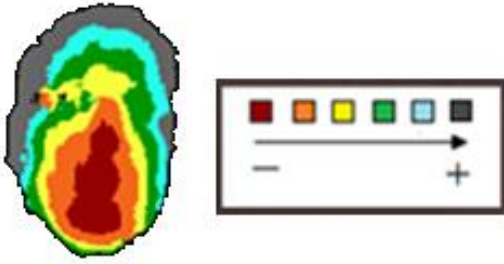


Figure 1: Locating the plans on the depth map.

A set of features are extracted from the depth map associated to the face. **The first feature** concerns the relative area N_r of the nearest region of the face to the device (Kinect) which characterizes the pan angle. The more the pan angle increases, the more the area of the nearest region increases compared to the face area. Indeed, due to the geometry of face, if the head is in front to the Kinect, the region of nose and lips is the nearest and occupies less pixels in map depth. However, when the head performs pan motion, the nearest area will concerns the cheek part and becomes largest. Figure 2 illustrates an example of three motions of pan where the nearest region colored by red color changes and follows the motion.

Let S_i be the area of the regions of the same depth (Plan Π_i). Let S^* be the region of minimum area. The ratio N_r is defined by equation 1:

$$N_r = \frac{S^*}{\sum_{i=1}^{i=n} S_i} \dots (1)$$

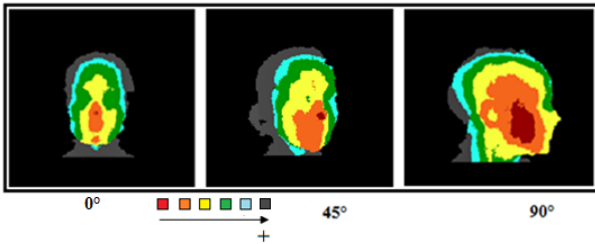


Figure 2: Depth maps of head which performs three Pan motion, where the color red (resp. blue) indicates the nearest (resp. far) region.

This first feature characterizes also the pan combined to the tilt motion. Indeed, the tilt motion allows the displacement of the nearest region relatively to the bounding box encompassing the face from the center (tilt equal zero) to the high part in case of inclination of the head towards the low, inversely in case of inclination of the head towards the high (see figure 3).

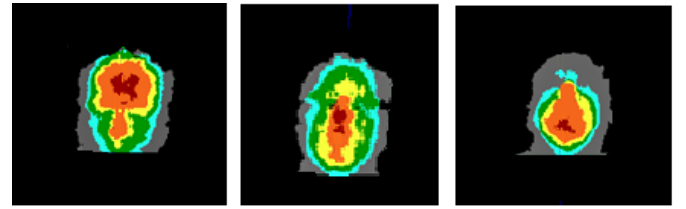


Figure 3: Depth maps head performing tilt motion.

In order to find the correspondence between the ratios N_r and the values of pan angle, a set of depth maps were acquired by using different values of pan motion.

The second feature concerns the symmetry on the depth map of the face. Firstly, the minimum rectangle encompassing the face is located and the line which passes by the center of this rectangle parallel to its length (respectively width) will be noted the symmetry axis (L), (respectively (W)). When head is in front of the camera (pan motion is equal to zero), there is symmetry between regions of the same depths relatively to the axis (L). This property is also valid for front pose even if head performs a roll motion (see figure 4).

Let (S_i^L, S_i^R) be the areas of the pair of regions (Left and Right) of the same depth (of the plan Π_i). The asymmetrical ratio A_r is defined by equation 2:

$$A_r = \sum_{i=1}^{i=n} \delta_i \frac{|S_i^L - S_i^R|}{S_i^L + S_i^R} \dots (2)$$

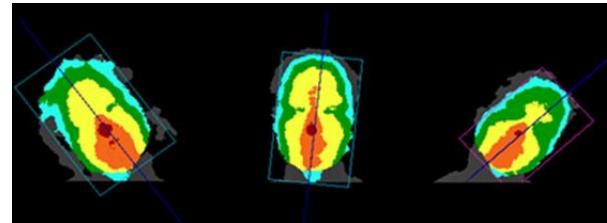


Figure 4: For frontal poses, the axis (L) divides the head area into symmetrical regions whatever the roll rotation.

Where $n = 6$ is the number of regions of different depths and δ_i is a parameter associated to the plan of depth Π_i . The values of δ_i are determined empirically in the training step by giving more importance (more weight) to the nearest plans in the computation of A_r .

The value of the ratio A_r is low when head is in front to the Kinect. However, more the pan rotation increases, more is the value of A_r is high. The variation of this value when the head is performing a pan rotation for different subjects is illustrated by the graph of figure 5.

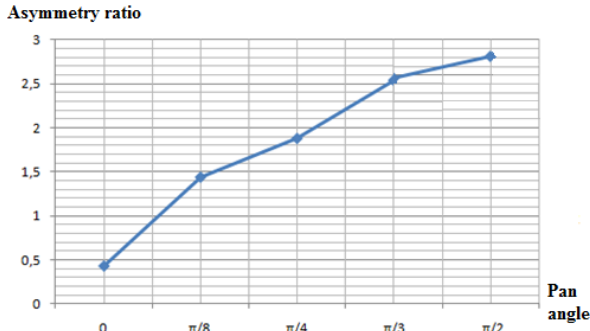


Figure 5: Asymmetry ratio variation for pan rotation.

The value of A_r characterizes different motions. For frontal poses (pan angle is equal to zero), whatever the roll angle this ratio is very low (see figure 4). For non-frontal poses, this value increases and is related to pan angle.

The **third feature** concerns the position of nearest region relatively to the axis (W). This feature characterizes the tilt motion. For frontal poses, it is easy to infer the motion tilt (see figure 4). However, for non-frontal poses, the orientation of the axis (L) and the position of nearest region are used conjointly to estimate the tilt motion (see figure 6).

The **fourth feature** concerns the orientation of the axis (L). For frontal poses, the orientation of this axis indicates the roll motion. Figure 4 illustrates this case where symmetry is verified relatively to the axis (L).

B. Algorithm

The following algorithm resume the main steps to be performed for head pose estimation.

Algorithm

Begin

1- Step learning

For a subset of poses of the data set chosen randomly, where all poses are acquired at least one time, estimate the parameters δ_i and the values of the fourth features related the poses.

2- Step testing

-Locate the head in the image and the associated region in the depth map and compute the bounding box encompassing the head and the two axes (L), (W).

-The pan and tilt angles are estimated by analyzing the values of computed features. Depending on the value of N_r , the frontal pose or non-frontal pose is distinguished.

If the pose is frontal, depending on the value of position of nearest region relatively to the axis (L), the tilt motion is estimated.

-For non-frontal poses, the asymmetry ratio A_r allows the estimation of pan motion. The position of nearest region relatively to the axis (W) and the orientation of (L) allow the estimation of tilt motion.

-The roll angle is computed as the orientation of the symmetry axis (L) in case of frontal pose.

End.

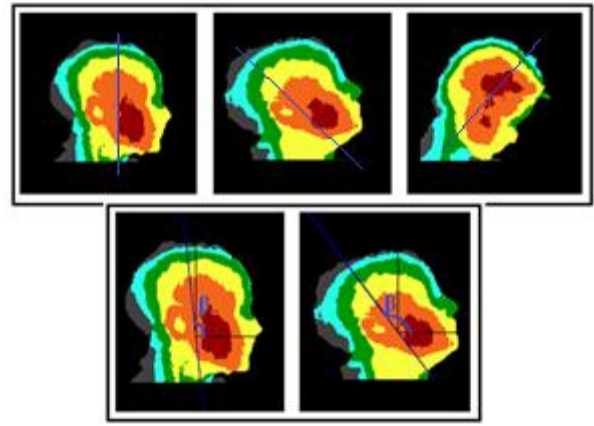


Figure 6: Head performing pan and tilt rotation.

IV. EXPERIMENTAL RESULTS

A. Data set used

In order to compare our results and the results of existing methods, we used our acquired data acquired with a Kinect sensor. People were recorded while turning their heads, sitting in front of the sensor, at roughly one meter of distance.

Once the face is located, we apply the masque to the depth map. The map of depths associated to heads are extracted and the pixels are colored with 6 colors depending on their depths as explained in section 3 (see figure 7).

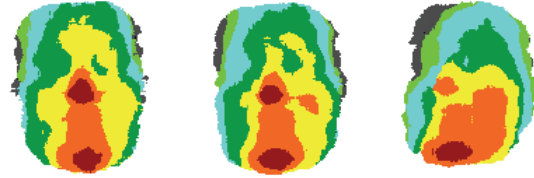


Figure 7: The depth maps associated to the heads partitioned into 6 plans.

B. Conducted tests

Based on proposed features, experiments are conducted on acquired data in our laboratory. As first tests, some acquired poses of different persons are used in order to determine the correspondence between the different values of the two features N_r , A_r .

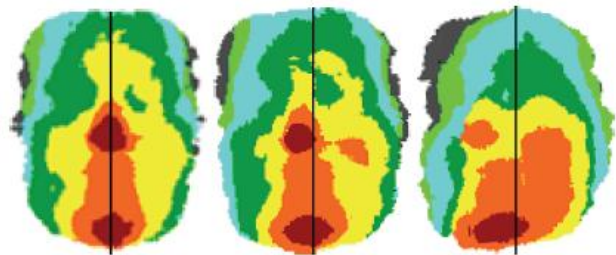


Figure 8: The computed axes of symmetry (L).

After this, some images were acquired and the poses were estimated using the two features. Figures 9, 10 illustrate the images of the person performing combined motion and the located regions in the depth maps. The estimated poses coincide with the ground truth data. The estimated error does not exceed 2 degree.

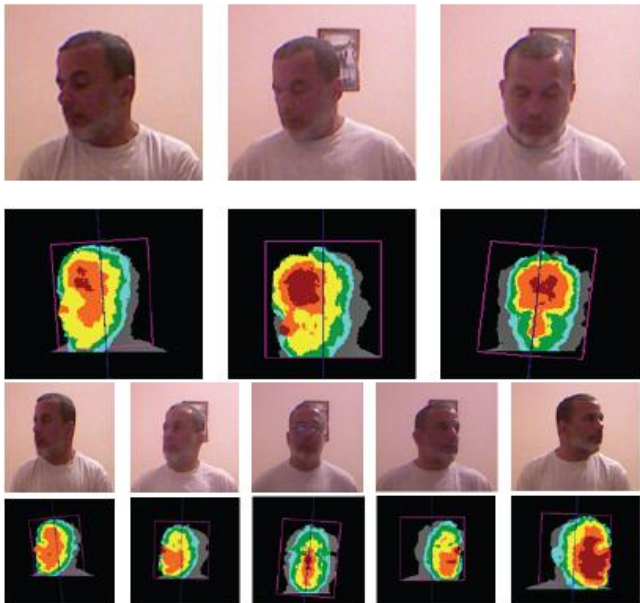


Figure 9: Some obtained intermediate results (depth maps).

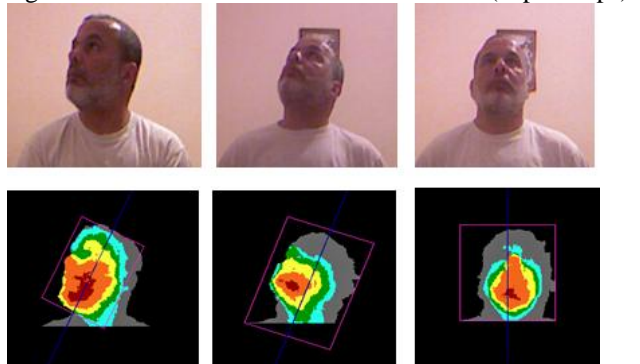


Figure 10: Some obtained intermediate results (depth maps).

The obtained results demonstrated that symmetry is a good feature for head pose estimation. The next step of this work is the comparison with the ground truth data of Biwi Kinect Head Pose Database [1] in order to study its accuracy.

IIV CONCLUSION

We presented a new method for head pose estimation based on the depth map. We exploited the symmetry/asymmetry on the depth map to deal with pan and tilt motions. The orientation of the symmetry axis of the bounding box indicates the roll angle of the head. Two features are extracted: the relative area of nearest region and the asymmetry ratio. These features are used to infer the motion angle (pan and tilt). The advantage of this method is the low cost of computation and the absence of

constraints concerning the limits of head motion of the presence of special cues on face.

We applied our approach of data set acquired in our laboratory. The obtained results are promising.

REFERENCES

- [1] G. Fanelli, M. Dantone, J. Gall, A. Fossati, L. Gool, Random Forests for Real Time 3D Face Analysis, *Int. J. Comput. Vision*, 101(3), 2013, pp. 437-458.
- [2] E. Chutorian, M. Trivedi, Head pose estimation in computer vision: A survey, *PAMI*, 31(4), 2009, pp. 607-626.
- [3] M. D. Breitenstein, D. Kuettel, T. Weise, L. Van Gool, H. Pfister, Real-time face pose estimation from single range images, *CVPR*, pp. 1-8, 2008.
- [4] S. Malassiotis, M. Srinivas, Robust real-time 3D head pose estimation from range data, *Pattern Recognition*, vol. 38, 2005, pp. 1153-1165.
- [5] R. Yang and Z. Zhang, Model-based head pose tracking with stereovision, *Aut. Face and Gesture Rec.*, pp. 255-260, 2002
- [6] N. Pears, T. Heseltine, M. Romero, From 3D point clouds to pose-normalised depth maps, *Int. J. Comput. Vision*, 89 (2-3), 2010, pp. 152-176.
- [7] F. Kondori, S. Yousefi, H. Li, S. Sonning, S. Sonning, 3D head pose estimation using the kinect, *WCSP*, pp. 1-4, 2011.
- [8] G. Fanelli, J. Gall, L. Van Gool, Real time head pose estimation with random regression forests, *CVPR*, pp. 617-624, 2010.
- [9] Q. Cai, D. Gallup, C. Zhang, Z. Zhang, 3d deformable face tracking with a commodity depth camera, *ECCV 2010*, pp. 229-242.
- [11] P. Paderis, X. Zabulis, A. A. Argyros, Head pose estimation on depth data based on Particle Swarm Optimization, *Workshop on Human Activity Understanding from 3D Data (HAU3D'2012)*, pp. 1-4, 2012.
- [10] E. Seemann, K. Nickel, R. Stiefelhagen, Head pose estimation using stereo vision for human-robot interaction, *Aut. Face and Gesture Rec.*, pp. 626-631, 2004.
- [12] P. Viola, M. Jones, Robust real-time face detection, *Int. J. Comput. Vision*, 57(2), 2004, pp. 137-154.