Head Pose Estimation Using Fuzzy Approximator Augmented by Redundant Membership Functions

S.Kazem Shekofteh[1], Maryam Baradaran–K[2], Saeed Toosizadeh[2], Mohammad–R. Akbarzadeh–T[1], Mahdi Hashemi[3]
[1]Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
[2]Department of Artificial Intelligence, Islamic Azad University, Mashhad Branch, Mashhad, Iran
[3]Mechatronics Research Group, Shahid Aghasizadeh Research Center, Mashhad, Iran
kazem.shekofteh@stu-mail.um.ac.ir, {m.baradaran, s.toosi}@mshdiau.ac.ir, akbarzadeh@ieee.org, hashemi124@gmail.com

***Abstract*—Estimating the head pose of a human is a key task for human computer interaction, visual surveillance and face recognition applications hence an important problem in computer vision. Most of the works in this field suffer from lack of continuous estimating of the head pose. Fuzzy systems are known as universal approximator capable of approximating an unknown function by having just few limited information while gaining high accuracy. In this paper, a new approach is proposed for estimating the rotation angle of the head along horizontal axis based on a fuzzy approximator in which the membership functions are constructed in a way that there is no manual tuning needed since for each range in the input variable which is not covered completely, a new membership function is introduced. The proposed method is able to provide a continuous estimate of the head along horizontal axis with high accuracy, low computational cost while avoiding from getting involved into complex mathematical equations. Experiments on images from two standard well-known databases showed less than 7° of average absolute error in estimation which is a significant improvement over the method based on the simple membership functions.**

Keywords-component; Head pose estimation, Fuzzy approximator

I. INTRODUCTION

The ability of a computer to recognize the pose of a head is high desirable in computer vision. From an early age, people display the ability to quickly and effortlessly interpret the orientation and movement of a human head, thereby allowing one to infer the intentions of others who are nearby and to comprehend an important nonverbal form of communicationIn a computer vision context, head pose estimation is the process of inferring the orientation of a human head from digital imagery. Like other facial vision processing steps, an ideal head pose estimator must demonstrate invariance to a variety of variable factors.

At the coarsest level, head pose estimation applies to algorithms that identify a head in one of a few discrete orientations, e.g., a frontal versus left/right profile view. At the fine (i.e., granular), a head pose estimate might be a continuous angular measurement across multiple degrees of freedom (DOF). In the context of computer vision, head pose estimation is most commonly interpreted as the ability to infer the orientation of a person's head relative to the view of a camera. More rigorously, head pose estimation is the ability to infer the orientation of a head relative to a global coordinate system, but this subtle difference requires knowledge of the intrinsic camera parameters to undo the perceptual bias from perspective distortion.

It is often assumed that the human head can be modeled as a disembodied rigid object [1]. Under this assumption, the human head is limited to three degree of freedom (DOF) in pose, which can be characterized by pitch, roll, and yaw angles as pictured in Figure 1.

In prior work [2] we exploit the approximation properties of fuzzy systems to estimate the pose of a head. Since fuzzy systems are universal approximators [3], that is they can approximate any function on a compact set to arbitrary accuracy, we proposed a fuzzy estimator that can estimate the head pose with low computational cost, high accuracy and robustness and also with continuous output which is the major property of the proposed approach.

Contributions: This paper complements the work in [2] by addressing an issue: for each range in the input variable which is not covered completely by membership functions, a new membership function is introduced to improve the consistency of the fuzzy approximator.

The rest of this paper is organized as follows: in Section II we discuss related works. In Section III we describe the preliminaries of the idea and the role of fuzzy systems in our approach. The details of our algorithm are discussed in Section IV. In Section V, the experimental results are discussed. Section VI concludes the work and presents the scopes for future works.

II. RELATED WORKS

Many approaches have been proposed for head pose estimation which we categorized them in the prior work by the fundamental approach that underlies its implementation according to [1].

Appearance template methods use image-based comparison metrics to match a view of a person's head to a set of exemplars with corresponding pose labels. The proposed methods in [4, 5] are placed in this class.

- **Detector array methods** train a series of head detectors each attuned to a specific pose and assign a discrete pose to the detector with the greatest support. References [6-8] used detector arrays or Support Vector Machines (SVM) to estimate the head pose.

978-1-4244-8666-3/10/$26.00 © 2010 IEEE

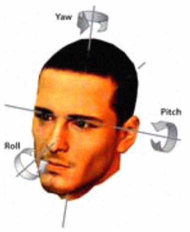

Figure 1. The three degrees of freedom of a human head can be described by the egocentric rotation angles pitch, roll, and yaw

- **Nonlinear regression methods** estimate pose by learning a nonlinear functional mapping from the image space to one or more pose directions such as works proposed in [9, 10].
- **Manifold embedding methods** seek low-dimensional manifolds that model the continuous variation in head pose. The proposed methods in [11, 12] used this approach.
- **Flexible models** fit a nonrigid model to the facial structure of each individual in the image plane. Ref. [13] is a sample of these methods.
- **Geometric methods** use the location of features such as the eyes, mouth, and nose tip to determine pose from their relative configuration. Authors of [14-16] used the inner and outer corners of each eye and the corners of the mouth with or without multiple cameras surrounding the head.
- **Tracking methods** operate by following the relative movement of the head between consecutive frames of a video. The works proposed in [17-19] tracked the head in a sequence of video frames.
- **Hybrid methods** combine one or more of these aforementioned methods to overcome the limitations inherent in any single approach such as [20].

Our proposed approach does not mainly fit into one of these categories. Furthermore it can be globally placed in the hybrid class. The main idea behind our method can be considered as a geometric one although without complex mathematical equations of geometric methods. The estimating process which will be done by a fuzzy estimator can be regarded as a manifold embedding method. This idea was never used before; hence the work is completely novel. Therefore the proposed algorithm is considered as a new class of hybrid methods.

III. BACKGROUND

A. Fuzzy Systems as Universal Approximators

We know that fuzzy systems are particular types of nonlinear functions, so no matter whether the fuzzy systems are used as controllers or decision makers or signal processors or any others, it is interesting to know the capability of the fuzzy systems from a function approximation point of view. It is proved that certain classes of fuzzy systems have this universal approximation capability [3].

To answer the question of how to approximate a function by a fuzzy system, we must first see what information is available for the nonlinear function $g(x) : U \subset R^n \rightarrow R$, which we are asked to approximate. Generally speaking, we may encounter the following situations:

- The analytic formula of $g(x)$ is known.
- The analytic formula of $g(x)$ is unknown, but for any $x \in U$ we can determine the corresponding $g(x)$. That is, $g(x)$ is a black box—we know the input-output behavior of $g(x)$ but not the details inside it.
- The analytic formula of $g(x)$ is unknown and we are provided only a limited number of input-output pairs $(x^j, g(x^j))$ where $x^j \in U$ cannot be arbitrary chosen.

The problem of head pose estimation is of the third type, which we are trying to develop a fuzzy system to estimate the pose of a head while having just few limited information.

B. Geometric Analysis of Facial Features

As it was mentioned in Section II, many researches have been done in the field of head pose estimation which we categorized into eight classes. Geometric approached are of interesting and practical classes. The methods in this class often use complex mathematical equations in order to find out the relations between facial features to estimate the angle of the head along one of three degrees of freedom.

In the prior work [2] we introduced a ratio named R_{LR} which was defined as the ratio of the width of one eye to the width of another eye:

$$R_{LR} = \frac{\textit{Observed Length of left eye}}{\textit{Observed Length of right eye}} \tag{1}$$

No one used this feature previously. Surprisingly we found out that this ratio is almost a constant value for each discrete position for any person placed in any angle of other axes such as pitch and roll, e.g. if a person's head is placed in -15° of yaw axis, despite his pose in other two axes, pitch and roll, the value of R_{LR} almost equals to the value of R_{LR} for another person in the same position. Hence we can obtain this ratio for some discrete known angles, and then estimate the angle of a head along yaw axis by just having this ratio which can be easily extracted from the image using one of the facial feature extraction methods. The major work should be done from now is how to construct a continuous estimate from several discrete known ratios.

Recall from Section II.A we could exploit a fuzzy system to approximate an unknown function for which we have just few limited information.

To solve this problem we used a fuzzy approximator to estimate the behavior of an unknown function for any input value of R_{LR} according to some limited known obtained relations between R_{LR} and head's angle along yaw axis. The details of the method were explained in [2]. The results of the statistical analysis are also shown in Table I.

The first column is the angle of the head along yaw axis. The other columns are statistical results over each angle. Columns min R_{LR} and max R_{LR} represent minimum and maximum value of R_{LR} for the corresponding angle respectively. Columns avg R_{LR} and stdev R_{LR} are the most interesting points which prove our claim.

TABLE I. STATISTICAL ANALYSIS FOR R_{LR} RATIO ON MORE THAN 7000 IMAGES

Angle \ Stats	min R_{LR}	max R_{LR}	avg R_{LR}	stdev R_{LR}
0°	0.87	1.16	**1**	0.05
-15°	0.64	1.01	**0.86**	0.1
-30°	0.6	0.9	**0.79**	0.11
-45°	0.32	0.65	**0.52**	0.1
-60°	0.28	0.49	**0.37**	0.08
-75°	0.12	0.24	**0.17**	0.05
-90°	0	0.1	**0.03**	0.03

These two columns are average and standard deviation of the ratio R_{LR} respectively. In the first row of the results, it is evident that the average ratio of the width of the left eye to the right eye is around 1.

The values 0.05 for *stdev* showed that 72% of data items are in range [$avg - stdev, avg + stdev$] and 95% of data items are in range [$avg - 2stdev, avg + 2stdev$]. This is consonant with properties of normal distributions. Furthermore since the value of *stdev* is very small compared to averages, the event of being 95% of data items in the range [$avg - 2stdev, avg + 2stdev$] also prove our claim that for a certain angle, the ratio R_{LR} is almost a constant value for any person positioned at any angle along two other axes, pitch and roll. Unfortunately most of the facial feature extraction methods do not essentially have the required accuracy to extract the coordinates of the eyes in the case of large rotations in the way axis i.e. angles -60°, -75° and -90°, hence the statistical analysis for these cases do not have as sufficient support as first rows. But we could fortunately achieve good approximation accuracy for them.

IV. CONSTRUCTING THE FUZZY APPROXIMATOR WITH REDUNDANT MEMBERSHIP FUNCTIONS

Here is the main different part of this work with our prior work in [2]. In order to construct the fuzzy system to approximate the head pose along yaw axis, we selected mamdani fuzzy system because of its intuitiveness and widespread acceptance and its applicability to this problem. We have 7 main known input-output pairs of the function for which we are to construct the fuzzy estimator. These 7 input-output pairs are listed in Table II. The input variable is a vector with just one element that is R_{LR}. Constructing the input membership functions based on these 7 pairs may lead in inconsistent membership functions, since there are some ranges in the input range that are not covered properly. For each of these ranges we defined a new membership function which can be significantly affect the consistency and accuracy of the new method. The output variable is also a single element vector that is the angle of the head along yaw axis. We considered 7 input and 7 output membership functions. The membership functions for both input and output variables are shown in Figures 2 and 3 respectively. The membership functions of input variable are constructed by considering values of averages and standard deviations for each average. The output membership functions are simple triangular one. As illustrated in Figure 2 we observed that such the aforementioned gaps exist between membership functions of in2 and in3 and between in4 and in5. Hence, we defined two new membership functions named in23 and in45 leading in total 9 membership functions for each input and output variable. These new membership functions are considered such that the support point of corresponding output membership function placed in the average point of support points of the two surrounding membership functions. Totally speaking, the approach starts from output membership functions e.g. to design the new membership function in23, we analyzed the R_{LR} value for faces that are positioned in -67.5° i.e. the average point of -60° and -75° equals to -67.5°. Another new membership function is constructed similarly. Table III shows these 9 input-output pairs. We also have 9 rules which are shown in Figure 4. Note that the rotation of the head along yaw axis is symmetric i.e. if value 0.86 for R_{LR} corresponds to -15°, then value $1/0.86$ will represent 15° rotation along yaw axis, hence for simplicity we consider only non-positive angles, but the work was completely implemented for both negative and non-negative angles.

V. EVALUATION OF THE PROPOSED METHOD

To evaluate our proposed method, we implemented our algorithm in Matlab using fuzzy toolbox. Two types of experiments were performed on this system. In the first experiment we fed the system with 1000 images of the database Pointing '04 [21] from which we extracted our statistical analysis described in Table I. In the second experiment, 1000 images from CAVE database were chosen to input the system. The results illustrated in Table IV were almost similar.

The first row demonstrates the results of the experiments performed on the first database Pointing '04. Average absolute error in this experiment was about 6.6° which is less than 7°. Minimum value for this experiment was completely expected too small.

TABLE II. SEVEN INPUT-OUTPUT PAIRS

input	output
1	0°
0.86	-15°
0.79	-30°
0.52	-45°
0.37	-60°
0.17	-75°
0.03	-90°

TABLE III. NINE FINAL INPUT-OUTPUT PAIRS

input	output
1	0°
0.86	-15°
0.79	-30°
0.71	-37.5°
0.52	-45°
0.37	-60°
0.28	-67.5°
0.17	-75°
0.03	-90°

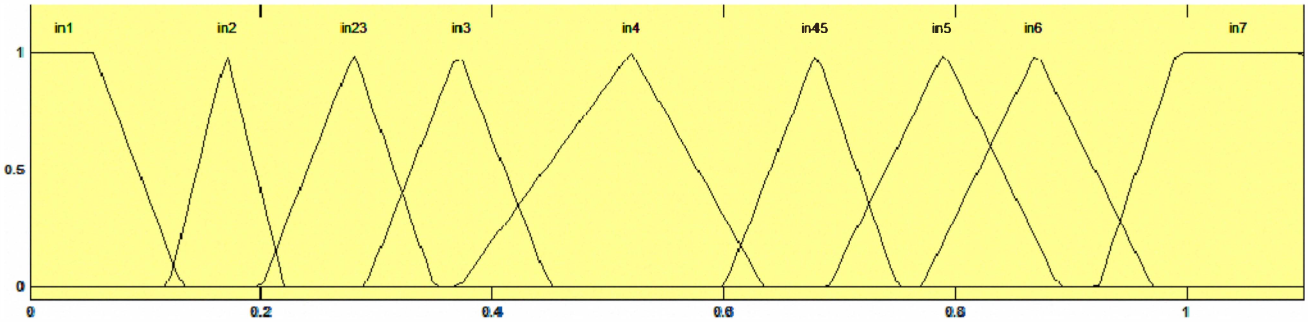

Figure 2. Membership functions for input variable 'ratio R_{LR}' named in1-in7 improved by in23 and in45

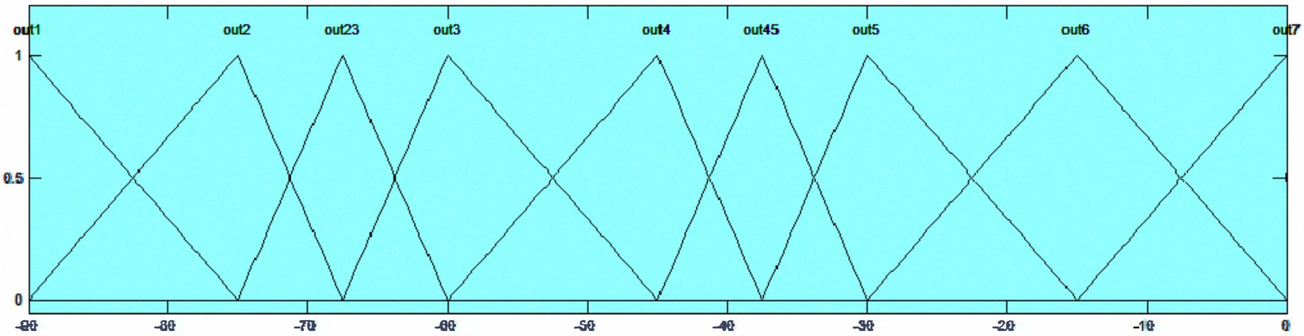

Figure 3. Membership functions for output variable 'angle' named out1-out7 with two new membership functions out23 and out45

IF ratio IS in1 THEN angle IS out1
IF ratio IS in2 THEN angle IS out2
IF ratio IS in23 THEN angle IS out23
IF ratio IS in3 THEN angle IS out3
IF ratio IS in4 THEN angle IS out4
IF ratio IS in45 THEN angle IS out45
IF ratio IS in5 THEN angle IS out5
IF ratio IS in6 THEN angle IS out6
IF ratio IS in7 THEN angle IS out7

Figure 4. Set of Rules

The second experiment on CAVE database had larger values than the first one. But the average absolute error value was still less than 7°. Compared to our previous work[2], this technique led to maximum 21.98% improvement in the accuracy of the estimator as described in Table V.

TABLE IV. EXPERIMENTAL RESULTS ON POINTING '04 AND CAVE DATABASE

DB	Num. Img	avg abs. err.
Pointing '04	1000	6.6°
CAVE	1000	7.83°

TABLE V. COMPARING THE RESULTS OF THE PREVIOUS WITH NEW METHOD ON POINTING '04 AND CAVE DATABASE

MAE / DB	previous method	new method	improvement
Pointing '04	8.46°	6.6°	21.98%
CAVE	8.7°	7.83°	10%

VI. CONCLUSION AND FURTHER WORK

In this paper we proposed an improvement over the prior work [2] in which we estimate the head pose of a person by means of a fuzzy approximator. The fuzzy approximator was formed of simple rules. In this work we defined new membership functions where some input ranges were not covered properly by membership functions.

For a head pose estimation system to be of general use, it should be invariant to identity, have sufficient range of allowed motion, require no manual intervention, and should be easily deployed on conventional hardware. Although some systems address all of these concerns, they often assume one or more conditions that simplify the pose estimation problem, at the expense of general applicability.

The proposed method inherits the benefits of the prior work which however because of the intuitive properties of fuzzy systems, in some cases will result in sub optimal estimation, but it has most of the features that a head pose estimation system should have to be general, i.e. it is invariant to identity, support sufficient range of motions and can be easily deployed on conventional hardware while it performs the estimation process with low computational time since it has low computational overhead. The major important property of our method is that it provides continuous estimation of head pose by just having few limited discrete prior known information.

We will further use an evolutionary method to optimize the components of fuzzy approximator such as membership functions.

REFERENCES

[1] E. Murphy-Chutorian and M. M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* vol. 31, pp. 607-626, April 2009.

[2] M. Baradaran–K, S. Toosizadeh, M. R. Akbarzadeh–T, and S. K. Shekofteh, "Exploiting Fuzzy Approximator to Head Pose Estimation," in *3rd International Symposium on Intelligent Ubiquitous Computing and Education.* Kunming, China, 2010.

[3] L.-X. Wang, "Approximation Propoerties of Fuzzy Systems I," in *A Course in Fuzzy Systems and Control*: Prentice-Hall International, Inc., 1997, pp. 128-129.

[4] S. Niyogi and W. Freeman, "Example-Based Head Tracking," in *IEEE International Conference of Automatic Face and Gesture Recognition*, 1996, pp. 374-378.

[5] J. Sherrah, S. Gong, and E.-J. Ong, "Face Distributions in Similarity Space under Varying Head Pose," *Image and Vision Computing,* vol. 19, pp. 807-819, 2001.

[6] J. Huang, X. Shao, and H. Wechsler, "Face Pose Discrimination Using Support Vector Machines (SVM)," in *14th International Conference of Pattern Recognition*, 1998, pp. 154-156.

[7] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," in *IEEE Conference of Computer Vision and Pattern Recognition*, 2001, pp. 511-518.

[8] Z. Zhang, Y. Hu, M. Liu, and T. Huang, "Head Pose Estimation in Seminar Room Using Multi View Face Detectors," in *International Workshop on Classification of Events, Activities and Relationships*, 2007, pp. 299-304.

[9] Y. Li, S. Gong, J. Sherrah, and H. Liddell, "Support Vector Machine Based Multi-View Face Detection and Recognition," *Image and Vision Computing,* vol. 22, 2004.

[10] E. Murphy-Chutorian and M. Trivedi, "Head Pose Estimation for Driver Assistance Systems: A Robust Algorithm and Experimental Evaluation," in *10th International IEEE Conference of Intelligent Transportation Systems*, 2007, pp. 709-714.

[11] B. Raytchev, I. Yoda, and K. Sakaue, "Head Pose Estimation by Nonlinear Manifold Learning," in *17th International Conference of Pattern Recognition*, 2004, pp. 462-466.

[12] J. Wu and M. Trivedi, "A Two-Stage Head Pose Estimation Framework and Evaluation," *Pattern Recognition,* vol. 41, pp. 1138-1158, 2008.

[13] Z. Gui and C. Zhang, "3D Head Pose Estimation Using Non-Rigid Structure-from-Motion and Point Correspondence," in *IEEE Region 10 Conference*, 2006, pp. 1-3.

[14] C. Canton-Ferrer, J. Casas, and M. Parda`s, "Head Pose Detection Based on Fusion of Multiple Viewpoint Information," in *International Workshop on Classification of Events, Activities and Relationships*, 2007, pp. 305-310.

[15] C. Canton-Ferrer, J. Casas, and M. Parda`s, "Head Orientation Estimation Using Particle Filtering in Multiview Scenarios," in *International Workshop on Classification of Events, Activities and Relationships*, 2007, pp. 305-310.

[16] J.-G. Wang and E. Sung, "EM Enhancement of 3D Head Pose Estimated by Point at Infinity," *Image and Vision Computing,* vol. 25, pp. 1864-1874, 2007.

[17] E. Murphy-Chutorian and M. Trivedi, "Hybrid Head Orientation and Position Estimation (HyHOPE): A System and Evaluation for Driver Support," in *IEEE Intelligent Vehicles Symposium*, 2008.

[18] S. Ohayon and E. Rivlin, "Robust 3D Head Tracking Using Camera Pose Estimation," in *18th International Conf. Pattern Recognition*, 2006, pp. 1063-1066.

[19] G. Zhao, L. Chen, J. Song, and G. Chen, "Large Head Movement Tracking Using SIFT-Based Registration," in *ACM International Conference of Multimedia*, 2007, pp. 807-810.

[20] Y. Hu, L. Chen, Y. Zhou, and H. Zhang, "Estimating Face Pose by Facial Asymmetry and Geometry," in *IEEE International Conference of Automatic Face and Gesture Recognition*, 2004, pp. 651-656.

[21] J. Tu, Y. Fu, Y. Hu, and T. Huang, "Evaluation of Head Pose Estimation for Studio Data," in *First Int'l Workshop Classification of Events, Activities and Relationships.* vol. 25, 2007, pp. 281-290.