A fuzzy approximator with Gaussian membership functions to estimate a human's head pose

Baradaran-K, M.a, Shekofteh, S.K.b, Toosizadeh, S.a, Akbarzadeh-T, M.-R.b

a Department of Artificial Intelligence, Islamic Azad University, Mashhad Branch, Mashhad, Iran
b Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract

Estimating the head pose plays an important role in computer vision and also as a key tool for visual surveillance and face recognition applications hence a prominent problem in computer vision. Most of the works in this field suffer from lack of continuous estimating of the head pose and high accuracy. We know fuzzy systems as universal approximator capable of approximating an unknown function by having just few limited information while attaining high accuracy. In this paper, an improved 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 of Gaussian type. 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 6° of average absolute error in estimation which is a significant improvement over the approximator with simple triangular membership functions. © 2010 IEEE.

Language of original document

English

Author keywords

Component; Fuzzy approximator; Gaussian membership function; Head pose estimation

References (23) View in table layout


A Fuzzy Approximator with Gaussian Membership Functions to Estimate A Human’s Head Pose

Maryam Baradaran–K m.baradaran@mshdiau.ac.ir, S.Kazem Shekofteh2, Saeed Toosizadeh3, Mohammad–R. Akbarzadeh–T akbarzadeh@ieee.org
1Department of Artificial Intelligence, Islamic Azad University, Mashhad Branch, Mashhad, Iran
2Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
m.baradaran@mshdiau.ac.ir, kazem.shekofteh@stu-mail.um.ac.ir, s.toosi@mshdiau.ac.ir, akbarzadeh@ieee.org

Abstract— Estimating the head pose plays an important role in computer vision and also as a key task for visual surveillance and face recognition applications hence a prominent problem in computer vision. Most of the works in this field suffer from lack of continuous estimating of the head pose and high accuracy. We know fuzzy systems as universal approximator capable of approximating an unknown function by having just few limited information while attaining high accuracy. In this paper, an improved 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 of Gaussian type. 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 6° of average absolute error in estimation which is a significant improvement over the approximator with simple triangular membership functions.

Keywords-component; Head pose estimation, Fuzzy approximator, Gaussian Membership Function

I. INTRODUCTION

Head pose estimation is an interesting research topic that can be used to indicate the subject’s focus of attention in 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.

Although it might seem like an explicit specification of a vision task, head pose estimation has a variety of interpretations. 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 our prior works [2, 3] we exploited the approximation properties of fuzzy systems to estimate the pose of a head. Since fuzzy systems are universal approximators [4], 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 works in [2, 3] by using Gaussian membership function instead of simple triangular one. We will then show that Gaussian membership functions are nearer to real data because of the instinctive properties of real data which is their normal distribution.

The remainder of the 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 [5, 6] 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 [7-9] used detector arrays or Support Vector Machines (SVM) to estimate the head pose.
- 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 [10, 11].
Manifold embedding methods seek low-dimensional manifolds that model the continuous variation in head pose. The proposed methods in [12, 13] used this approach.

Flexible models fit a nonrigid model to the facial structure of each individual in the image plane. Ref. [14] 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 [15-17] 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 [18-20] 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 [21].

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 [4].

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 \( g(x) : U \subset \mathbb{R} \rightarrow \mathbb{R} \), 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 \( (x', g(x')) \) where \( x' \in U \) 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 approaches 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:

\[
R_{LR} = \frac{\text{Observed Length of left eye}}{\text{Observed Length of right eye}}
\]  

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 use 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 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 value 0.05 for stdev shows that 72% of data items are in range \([\text{avg} - 2 \text{stdev}, \text{avg} + 2 \text{stdev}]\) and 95% of data items are in range \([\text{avg} - \text{stdev}, \text{avg} + \text{stdev}]\).
This is consonant with properties of normal distributions. Furthermore since the value of \textit{stdev} is very small compared to averages, the event of being 95\% of data items in the range \([\text{avg} - 2 \times \text{stdev}, \text{avg} + 2 \times \text{stdev}]\) 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.

In this paper we claim that the rotation of the head along horizontal axis is a linear process, i.e. the ratio \(R_{LR}\) will decrease (or increase) linearly when the head is rotated. For further explanation consider the value of ratio \(R_{LR}\) when the head is positioned at 0° along horizontal axis which equals to 1. This is obvious when the face is in a complete frontal view, so the length of both left and right eye are the same. When the face is positioned at -90° along horizontal axis the value of ratio \(R_{LR}\) equals to 0, because the left eye is completely invisible. Any other rotations of the head between 0° and -90° will be linear because this process itself is linear. Hence the average value for each angle is a linear function of the corresponding angle value. We will call these statistics as \textit{Theoretic statistics} and statistics in Table I as \textit{Experimental statistics}.

The above mentioned \textit{Theoretic statistics} are shown in Table II. Comparing these two tables showed that the experimental statistics are very close to theoretic ones, i.e. the difference between \(
\text{avg } R_{LR}\) in Theoretic information and \(
\text{avg } R_{LR}\) in Experimental information is very small. This also endorses the statistics extracted from our experiments.

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

<table>
<thead>
<tr>
<th>Angle</th>
<th>stat s</th>
<th>(\min R_{LR})</th>
<th>(\max R_{LR})</th>
<th>(\text{avg } R_{LR})</th>
<th>(\text{stdev } R_{LR})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0°</td>
<td>0.87</td>
<td>1.16</td>
<td>1</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>-15°</td>
<td>0.64</td>
<td>1.01</td>
<td>0.86</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>-30°</td>
<td>0.32</td>
<td>0.65</td>
<td>0.52</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>-45°</td>
<td>0.28</td>
<td>0.49</td>
<td>0.37</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>-60°</td>
<td>0.12</td>
<td>0.24</td>
<td>0.17</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>-75°</td>
<td>0</td>
<td>0.1</td>
<td>0.03</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>-90°</td>
<td>0</td>
<td>0.1</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\text{This is consonant with properties of normal distributions. Furthermore since the value of } \textit{stdev} \text{ is very small compared to averages, the event of being 95\% of data items in the range } \[\text{avg} - 2 \times \text{stdev}, \text{avg} + 2 \times \text{stdev} \] \text{ also prove our claim that for a certain angle, the ratio } R_{LR} \text{ 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.}

### TABLE II. \text{THEORETIC VS. EXPERIMENTAL STATISTICAL ANALYSIS FOR } R_{LR} \text{ RATIO}

<table>
<thead>
<tr>
<th>Angle</th>
<th>\text{Theoretic}</th>
<th>\text{Experimental}</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{avg } R_{LR}</td>
<td>\text{stdev } R_{LR}</td>
<td>\text{avg } R_{LR}</td>
</tr>
<tr>
<td>0°</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15°</td>
<td>0.8335</td>
<td>0.05</td>
</tr>
<tr>
<td>30°</td>
<td>0.6668</td>
<td>0.05</td>
</tr>
<tr>
<td>45°</td>
<td>0.5001</td>
<td>0.05</td>
</tr>
<tr>
<td>60°</td>
<td>0.3334</td>
<td>0.05</td>
</tr>
<tr>
<td>75°</td>
<td>0.1667</td>
<td>0.05</td>
</tr>
<tr>
<td>90°</td>
<td>0</td>
<td>0.05</td>
</tr>
</tbody>
</table>

\text{This is consonant with properties of normal distributions. Furthermore since the value of } \textit{stdev} \text{ is very small compared to averages, the event of being 95\% of data items in the range } \[\text{avg} - 2 \times \text{stdev}, \text{avg} + 2 \times \text{stdev} \] \text{ also prove our claim that for a certain angle, the ratio } R_{LR} \text{ 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. \text{CONSTRUCTING THE FUZZY APPROXIMATOR WITH GAUSSIAN MEMBERSHIP FUNCTIONS}

Here is the main different part of this work with our prior works in \cite{2,3}. 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 extracted from Theoretic statistics 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 \(R_{LR}\). We construct the input membership functions based on these 7 pairs. These membership functions are of Gaussian type which can be closer to the data extracted from real environment. The mean value for each Gaussian membership function is equal to the corresponding theoretic average showed in Table II and the value of standard deviation is considered as one third of the distance to the next average point. The reason for selecting this fraction of the distance is that we expect membership functions to cover about 99\% of data within its coverage range. This issue is illustrated in Figure 2. The output variable is also a single element vector that is the angle of the head along yaw axis. The output membership functions are simple triangular one. The membership functions for both input and output variables are shown in Figures 3 and 4 respectively. Note that since 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, for simplicity we consider only non-positive angles, but it is completely implemented for both negative and non-negative angles.

![Figure 2: Defining standard deviation for each membership function. mean = 0.8335, stdev = 0.05](image)

| Stat s | \text{avg } R_{LR}| \text{stdev } R_{LR}| \text{avg } R_{LR}|
|-------|------------------|--------------------|
| input | output |
| 0°    | 0                |
| 15°   | 0.8335           |
| 30°   | 0.6668           |
| 45°   | 0.5001           |
| 60°   | 0.3334           |
| 75°   | 0.1667           |
| 90°   | 0                |

\text{This is consonant with properties of normal distributions. Furthermore since the value of } \textit{stdev} \text{ is very small compared to averages, the event of being 95\% of data items in the range } \[\text{avg} - 2 \times \text{stdev}, \text{avg} + 2 \times \text{stdev} \] \text{ also prove our claim that for a certain angle, the ratio } R_{LR} \text{ 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.}
V. EVALUATION OF THE PROPOSED METHOD

A. Basic Fuzzy Approximator with Theoretic Statistics-based Triangular Membership Functions

Here we define a basic fuzzy approximator which has simple triangular membership functions extracted from theoretic statistics explained in Table II. The membership functions of this approximator are the same as the proposed approximator but in triangular shape instead of Gaussian.

B. Evaluation

To evaluate our proposed method, we implemented our algorithm in Matlab using fuzzy toolbox. We compare the proposed method with previous work in [2] and basic approximator discussed in Section V.A. More than 2000 images selected from two datasets Pointing '04 [22] and CAVE [23]. The result of experiments on these test images are shown in Table IV. We can see that mean absolute error in the new method is less than 6° while this value for our previous work [2] was less than 9° and for Linear Triangular method was less than 8°. Therefore the proposed method can decrease the mean absolute error of the previous work [2] by about 31% and the Linear Triangular method by 26% that is showed in Table V.

VI. CONCLUSION AND FURTHER WORK

In this paper we have proposed a new fuzzy approximator constructed by Gaussian input membership functions that performed much better than prior work [2] in which we estimate the head pose of a person by means of a fuzzy approximator constructed with simple triangular membership functions. The fuzzy approximator was formed based on the hypothesis of Theoretic Statistics.

For a general purpose head pose estimation system, it should have sufficient range of allowed motion, be invariant to identity, 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.

<table>
<thead>
<tr>
<th>Table IV. Comparing Experimental Results over Three Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Previous work</td>
</tr>
<tr>
<td>Linear Triangular</td>
</tr>
<tr>
<td>Linear Gaussian</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table V. Improvement Ratio over Simple Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Previous work</td>
</tr>
<tr>
<td>Linear Triangular</td>
</tr>
</tbody>
</table>
The proposed method 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 and 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. 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 with high accuracy such that the mean absolute error of this method is less than 6\degree.

We are now working on optimizing the components of fuzzy approximator such as membership functions by an evolutionary method such as Genetic Algorithm.

REFERENCES


