Further Study on Camera Position Estimation from Image by ANFIS

Shao-Fan Lien [1], Kuo-Hsien Hsia [2], Chun-Chieh Wang [3], Ting-En Lee [4] and Juhng-Perng Su [5]

[1,4]Graduate School of Engineering Science & Technology (Doctoral Program)
National Yunlin University of Science & Technology
No. 123, University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.
(E-mail: g9510801@yuntech.edu.tw[1], E-mail: g9310808@yuntech.edu.tw[4])
[2]Department of Management Information System, Far East University
No. 49, Chung Hua Rd., Hsin-Shih. Tainan County 744, Taiwan, R.O.C.
(Tel : 886-6-597-9566#7008; Fax : 886-6-597-7010)
(E-mail: khhsia@cc.feu.edu.tw)
[3]Department of Electronic Engineering, Chienkuo Technology University
No. 1, Chieh Shou N. Rd., Changhua City, Taiwan, R.O.C.
(E-mail: jasonccw@ctu.edu.tw)
[5]Department of Electrical Engineering, National Yunlin University of Science & Technology
No. 123, University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.
[5]Department of Information Technology, Overseas Chinese University
No. 100, Ciao Kwang Rd., Taichung 40721, Taiwan, R.O.C.
(E-mail: sujp@yuntech.edu.tw)

Abstract: It is clear that different images will be obtained when one takes pictures with different camera positions. One can easily determine the characteristics of the captured image by projective geometry. However, it is hard to estimate the camera position only from an image. Machine learning methods are very useful for the nonlinear relation identifying. In this research, released situations of the images comparing to the earlier works are considered. An adaptive neuro-fuzzy inference system (ANFIS) network design is deployed and used for camera position estimating in this paper. From the experimental results, it is evidently that the proposed method can estimate the center of the camera correctly and effectively.

Keywords: ANFIS, Image feature extraction, Camera position estimation, Projective geometry.

I. INTRODUCTION

The 3D scenery and the pictured 2D image which could be related by projective geometry. However, the relationship is nonlinear and highly complicated. Adaptive neuro-fuzzy inference system (ANFIS) is provided for finding out this relationship, and that may be a possible way to building the relationship.

Consider a circular target in the 3D scenery, and the center of the circular target is projected to the centerline of a 2D image, the circle in the 3D scenery will be projected to a deformed circle in the 2D image. By projective geometry, the features of the deformed circle can be obtained from the diameter of the circular target, the distance between the camera center and the target center, and the variation angle of the camera. By this relation, it is possible to estimate the position of the camera from some features of the deformed circle and the circular target. However, the relationship between these features and the position of the camera is more complicated then previous research. In [1][2], an algorithm was developed for solving the problem under the condition that the target center must be at the center of image. However, it is not general enough for the applications of security systems [3] and aerial photography [4]. Thus we try to release the restriction and to solve this problem in this paper.

ANFIS is used for learning the mapping relation between the 3D world and the 2D image in this paper. The data pre-processing procedure [5] is provided to speed up the learning efficiency of the ANFIS. The experimental results show that the camera positions could be estimated quickly and accurately by the image features. Details are illustrated in the following contents.

II. PROJECTIVE TRANSFORMATION OF THE IMAGE

When we take pictures by camera, a point in the 3D space is projected to 2D image through the camera center, and different positions of camera center will result different projections of the point [6]. Consider the captured image. It is the projection of a circular target in the 3D space. A circular target will result in a deformed circle in the captured image. One can always trim the camera so that the crossover point of the axes of the deformed circle, called the 'center' of the deformed circle, is located on the vertical line crossing the center of the image, as shown in Figure 1. Obviously, this 'center' is the projection of the center of the circular target.

From Figures 1 and 2, one can choose five quantities to form the following features:

1. "$Ratio_{project} = h/k$" is dependent on the pitch angle of the camera.
2. "$Ratio_{height} = \bar{r}/d$" is dependent on the distance between the camera and the image center in the 3D world.
3. "$Ratio_{angle} = h/\bar{r}$" is dependent on both the pitch angle of the camera and the distance between the camera and the image center in the 3D world.
4. "$Ratio_{position} = k/Q$" is dependent on the target position.

The relationships of these parameters could be calculated by trigonometric relations and projective geometry. The relationships are shown in Figures 2 (a), (b) and (c). From above, the algorithm to find these ratios could be written as in Algorithm 1.

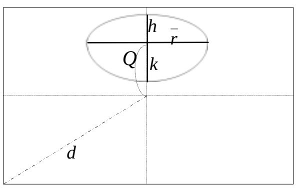

Figure 1. Projected image of a circular target.

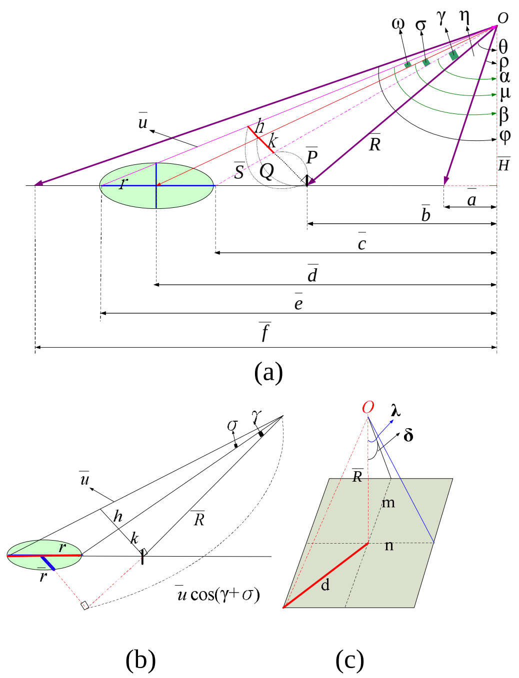

Figure 2. Relationship between the circular target and the projection: (a) Features of the circular target projection. (b) Feature $\bar{r}$ of the circular target projection. (c) Half visual angles.

Algorithm 1: Algorithm for finding the ratios.

Find the geometry projection of the **Perspective target**.

Procedure Perspective_target (***$Ratio_{project}$*, *$Ratio_{height}$*, *$Ratio_{angle}$*, *$Ratio_{position}$***)

"MinRtm: The lowest boundary of the target moving range in image."
"MaxRtm: The highest boundary of the target moving range in image."
"*r*: The radius of target."
"λ : The half vertical visual angle of camera."
"δ : The half horizontal visual angle of camera."
"R: The distance between the camera and the image center in the 3D world."
"θ : The pitch angle of the camera."
"The remainder definitions of parameters are shown in Figure 2. "

r=2.6, λ =27, δ =36
for R=20:1:300
for θ =-62:1:62
{

$\bar{H} = \sin\theta * \bar{R}$,

$\bar{a} = \sin\rho * \bar{H}$, $\bar{b} = \cos\theta * \bar{R}$,

$\rho = \theta - \eta$, $\varphi = \theta + \eta$

$\bar{f} = \tan\varphi * \bar{H}$, $\bar{d} = \bar{f} - 2r$,

MinRtm= $(\bar{b} - \bar{a} - r)$

MaxRtm= $(\bar{f} - r - \bar{b})$

for Range = MinRtm: 1: MaxRtm
{

$\bar{e} = \bar{b} + Range$,

$\mu = \tan^{-1}(\frac{\bar{e}}{\bar{H}})$, $\alpha = \tan^{-1}(\frac{\bar{d}}{\bar{H}})$

$\omega = \varphi - \mu$, $\sigma = \mu - \alpha$, $\gamma = \alpha - \theta$

$\bar{P} = \tan\gamma * \bar{R}$, $Q = \tan(\gamma + \sigma) * \bar{R}$,

$\bar{S} = \tan(\gamma + \sigma + \omega) * \bar{R}$

$k = Q - \bar{P}$, $h = \bar{S} - Q$, $\bar{u} = \sqrt{(\bar{e})^2 - (\bar{H})^2}$

$\bar{v} = \tan\lambda * \bar{u}$, $m = \tan\lambda * \bar{R}$,

$n = \tan\delta * \bar{R}$

$d = \sqrt{m^2 + n^2}$, $\bar{r} = \frac{\bar{R} * r}{\bar{u}\cos(\gamma + \sigma)}$

$Ratio_{project} = h/k$, $Ratio_{height} = \bar{r}/d$,

$Ratio_{angle} = h/\bar{r}$, $Ratio_{position} = k/Q$

}
}

III. FEATURE EXTRACTION AND IMAGE PROCESSING

In this paper, the features of image are $Ratio_{project}$, $Ratio_{height}$, $Ratio_{angle}$ and $Ratio_{position}$. The procedure for feature extraction and image processing is illustrated in Figure 3.

Thresholding and filtering are proposed for target segmentation in this procedure. The binary image could be get by thresholding, and that will be more easily for analyzing and computing [7] [8]. The range of the basic rectangle [7] is defined by the minimum rectangle containing the object. The definitions of the major axis and the minor axis are the longer edge and the shorter edge of the rectangle, respectively.

In this case, refer to Figure 1, the length of the major axis is equal to $\bar{r} + \bar{r}$, and the length of the minor axis is equal to h + k.

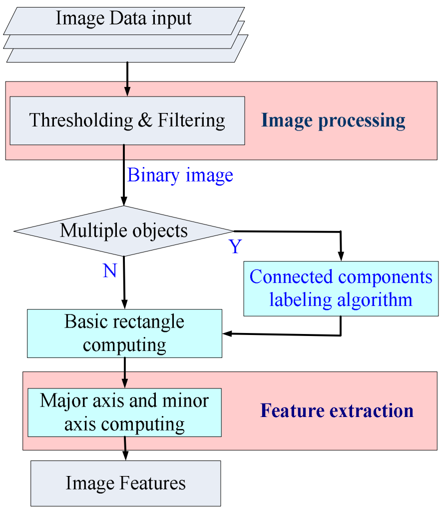

Figure 3. Procedure for feature extraction.

IV. DATA PRE-PROCESSING AND ANFIS TRAINING

In this paper, the ANFIS [5][9] is provided for replace the highly complicated relationship between a circular target in the 3D scenery and its projective transformation. The training data is constructed by Algorithm 1. The amount of the data is so huge that the training time is intolerably long. The data pre-processing procedure, illustrated in Figure 4, is proposed to speed up the training. There are two parts in this method, which are data segmentation and data classification. In the first part, the data are segmented into several stages for reducing the huge data base, are called "*Data stage_1~k*". In the second part, the "*Data stage_i*" is classified as four inputs related to Q, θ and $\bar{R}$, respectively. For further data classification, each class of data are sorted out the high correlated data and low correlated data for ANFIS training.

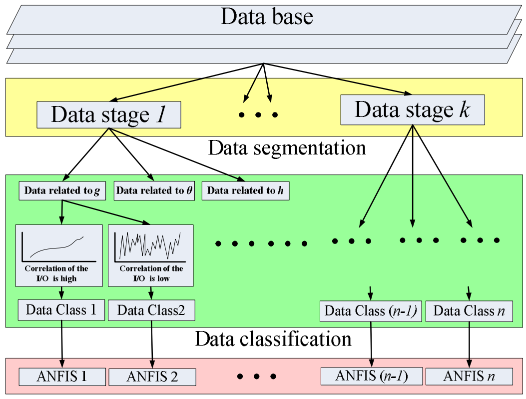

Figure 4. The data pre-processing procedure.

The ANFIS is trained with the highly correlated data, which will result in defining fewer fuzzy sets on the input data. This idea can save a lot of time for training, and the resulting ANFIS will have more precise estimation.

V. EXPERIMRNTAL RESULTS

Consider the case that the radius of the circular target in 3D world is 2.6 cm, the angle θ is varied from -62° to 62°, and the distance R is varied from 20 cm to 300 cm. The training data is generated by geometry computing. By the data pre-procession procedure, the data base is classed into the smaller and simpler stages for ANFIS training. Consequently, the numbers of the ANFIS are increased with the data stages. For this example, the data stage of "*height 126~130cm*" is selected for demonstrating the ANFIS training results.

There are three ANFIS's used for estimating Q, θ and R. The input data for these ANFIS's are $Ratio_{project}$, $Ratio_{height}$, $Ratio_{angle}$ and $Ratio_{position}$. The ANFIS training results are shown in Figures 5 (a), (b) and (c), and the resulting RMSE (root mean squared error) of Q, R and θ are 0.641, 0.124 and 0.365, respectively. The numbers of fuzzy sets on the inputs for each ANFIS are summarized in Table 1. Take the first row of Table 1 as an example, the ANFIS input $Ratio_{position}$ and ANFIS output Q has the highly correlation, therefore the number of fuzzy sets for $Ratio_{position}$ is 4; on the contrary, the number of fuzzy sets for other inputs are larger than 4. All the membership functions of the input fuzzy sets are in the shape of Gaussian.

Consider the image shown in Figure 6(a). The size of this image is 160*120 pixels. Figure 6(b) is the binary image of it, and Figure 6(c) shows the basic rectangle of the target. The length of the major axis is 29.6583 and that of the minor axis is 21.8644. In Figure 6(c), its features are extracted by image processing algorithm. The extracted features will be the input of ANFIS for the camera position estimation. For this case, the image features were extracted and that were shown in Table 2. The result of camera position estimation by ANFIS is shown in Table 3. From Table 3, it is shown that the camera position could be estimated from a real image by the proposed procedure.

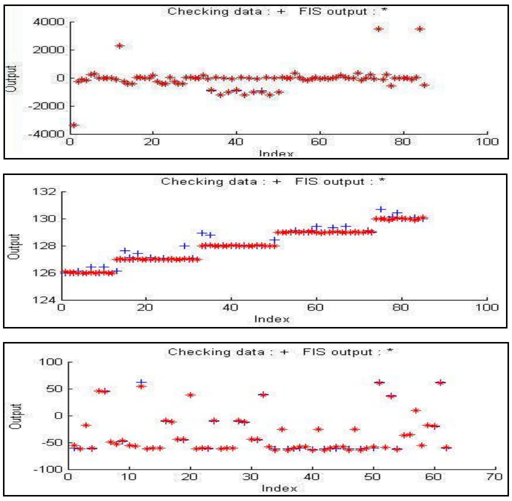

Figure 5. ANFIS checking results. (a) the distance Q (b) the distance R (c) the angle θ

Table 1. Numbers of fuzzy sets on inputs for each ANFIS

Input / ANFIS	$Ratio_{project}$	$Ratio_{height}$	$Ratio_{angle}$	$Ratio_{position}$
for Q	8	8	8	4
for R	8	4	8	8
for θ	4	8	8	8

(a) (b) (c)

Figure 6. Image of the reference object. (a) Image input (b) Binary image (c) Basic rectangle of image

Table 2. Image features of Figure 6(a).

Features	$Ratio_{project}$	$Ratio_{height}$	$Ratio_{angle}$	$Ratio_{position}$
Ratios	1.1092	1.0247	8.9016	0.0817

Table 3. Estimated results of Figure 6(a).

	Q (cm)	R (cm)	θ (°)
ANFIS Output	2.34	36.87	41.89
Camera position	2.5	37	42
Error	0.16	0.13	0.11

VI. CONCLUSION

In this paper, the restriction in [2] that the 'center' of the deformed circle, which is the projection of the circular target, has to be located at the center of the captured image has been released to be located at the vertical line crossing the center of the image. The proposed algorithm could estimate the camera position under this condition. The released condition caused the huge amount of training data. The data pre-processing procedure was introduced on the training process for ANFIS to speed up training. Experimental results show that this can simplify the ANFIS learning and reduce the learning time.

From the experimental results, we use the image feature extraction algorithm to find out the available features of the object as ANFIS inputs. The algorithm could extract the features of object precisely, and the obtained ANFIS also can estimate camera position effectively. It shows that the complicated mapping and highly nonlinear image projective transformations could be replaced by ANFIS.

REFERENCES

[1] K. H. Hsia, S. F. Lien, C. C. Wang, T. N. Lee and J. P. Su (2009), "Camera position estimation from image by neural network," ICIC Express Letters (EI). Vol.3, No.4, pp. 1161-1166.
[2] K.H. Hsia, S.F. Lien, C.C. Wang and J. P. Su (2009), "Camera position estimation from image by ANFIS," Fourth International Conference on Innovative Computing, Information and Control (ICICIC 2009).
[3] K. Sage and S. Young (1999), "Security applications of computer vision," Aerospace and Electronic Systems Magazine, IEEE, pp. 19-29.
[4] J.A. Franco, M. Moctezuma, and F. Parmiggiani (2002), "A fusion-based segmentation algorithm for high-resolution panchromatic aerial photography," Proc. International Geoscience and Remote Sensing Symposium, IEEE, vol. 6, pp. 3396-3398.
[5] M. T. Hagan, H.B. Demuth and M. Beale (2004), Neural Network Design, Thomson.
[6] R. Hartley and A. Zisserman (2003), Multiple View Geometry in computer vision, Cambridge University Press.
[7] L.G. Shapiro and G. C. Stockman (2001), Computer Vision, Prentice Hill.
[8] M. Nixon and A. Aguado (2008), Feature Extraction & Image Processing, Amsterdam Academic, Boston.
[9] A.D. KulKarni (2001), Computer Vision and Fuzzy-Neural systems, Upper Saddle River, Prentice Hall.