

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Calibration of the Camera Used in a Questionnaire Input System by Computer Vision

¹Yong-Ren Pu, ¹Su-Hsing Lee and ²Chao-Lin Kuo

¹Department of Occupational Safety and Health, Chang Jung Christian University,
No. 396, Sec. 1, Changrong Road, Gueiren District, Tainan City 71101, Taiwan, Republic of China

²Institute of Maritime Information and Technology, National Kaohsiung Marine University,
No. 482, Zhongzhou 3rd Road, Cijin District, Kaohsiung City 80543, Taiwan, Republic of China

Abstract: Quantitative information collection is important for scholars using questionnaires, since the number of administered questionnaires in a major survey is usually quite large. To develop an automatic and vast input method for questionnaires by computer vision, the authors adopt the perspective projection model in the image processing. This study presents a method to calibrate the camera used in the framework of this questionnaire input system. The distortions of the questionnaire image are derived therein so that the system is able to locate all checkboxes accurately. The computer software is coded with a visual programming language and is developed to extract the interiors of all checkboxes by seeking their edges. The experiment is designed, by reducing the errors produced by the recognized centers of some designated checkboxes, to determine the optimal values of various focal lengths to the image plane.

Key words: Image processing, calibration, perspective projection, questionnaire, computer vision

INTRODUCTION

Questionnaire survey is a very important tool for scholars in many research fields, especially in the areas of public health and social science, to collect quantitative information. Its application follows strict procedures which include design, printing, pre-test, administering, collection, data input and analysis. Hoping that the results can achieve some extent of validity and representation, researchers often administer a considerable quantities of questionnaires that may reach thousands or, in some cases, tens of thousands (Oppenheim, 1992). Therefore, in a major survey, data input plays a key role in the whole procedure. It demands not only accuracy but also speed, so that the researchers can advance to the next stage promptly and smoothly.

In the authors' experiences, most questionnaires are designed to have many checkboxes to be answered according to the printed questions. The so-called Scantron cards which are made of study from some special materials must be read by Optical Mark Read (OMR) scanners. Other questionnaires made of regular study should be recorded manually. Some commercial businesses contract to input the large amount of questionnaires by their key punch operators trained to key in those checkboxes one by one in a speedy manner.

No matter how it is done by machines or man powers, the data input costs much and takes time.

Along with the advance in computer capabilities, the technologies of image processing have been improving for almost 50 years. The early applications took place in the fields of aerospace engineering and medical science (Gonzales and Woods, 2002). Nowadays these technologies are used in a broad range of areas. Some studies have been done or progressing in the medical applications such as: the white blood cell nucleus segmentation under the microscope (Madhloom *et al.*, 2010), the breast cancer detection from the mammogram images (Alhadidi *et al.*, 2007) and 3D mesh reconstruction of lung nodules in the CT scan (Raghava and Kumar, 2006). Other studies focus on the industrial applications such as: The identification of sugarcane nodes for locating the cutting points (Moshashai *et al.*, 2008) and the evaluation of dry powder mixing homogeneity in industrial processes (Koc *et al.*, 2007). As to the field of space technology, Ahmadi *et al.* (2008) worked on the road extraction from the satellite images.

In recent years, there are some developments for the method of information input different from using OMR scanners. A US-patented system utilizes a capture device on which a questionnaire is placed (Tunney, 2006). The device is a flat panel which is able to digitally capture

respondents' pen strokes that are processed and recorded into data in real time. Although, this system can identify multiple of questionnaire pages, it needs to operate on site and is not possible to administer a large amount of questionnaires at one time. Tseng (2005) have developed an automatic collection support system by integrating some primary modules which are actually functions of the existing commercial software, to collect data of checkboxes from numerous questionnaires. By OMR technology the system can process either electronic files or scanning images of answered questionnaires. However, this system does not perform well in positioning electronic questionnaire drafts and still uses a scanner or a fax machine to acquire scanning image files that occupy certain storage memories.

This research is a pilot study of the development of an automatic questionnaire input system. The ultimate goal is to execute a vast data input by computer vision in a speedy manner for all kinds of questionnaires answered in checkboxes. To achieve this, a graphical user interface which is coded with a visual programming language is developed and integrated with a CCD (Charge-Coupled Device) camera. The software is functioned with image capturing, positioning, checkbox labeling/locating, check/cancel recognizing and text file output.

SYSTEM FRAMEWORK

The hardware framework of the developing automatic questionnaire input system is shown in Fig. 1. It integrates various devices including 1394 CCD, image interface, computer, controller and study feeder. A computer with CPU of DualCore Intel Core 2 Duo E8300, 2833 MHz is utilized and equipped with NI PCI-8252

interface to acquire IEEE 1394 image signals. An IEEE 1394 CCD of AVT Guppy F146B is seated on a copystand and hooked to the interface. The height of CCD above the questionnaire is set to be 30 cm to include the whole page in the CCD image. The software is coded with LabVIEW 7.1.

When operating in the initial locating mode, the operator places one blank questionnaire beneath the camera which captures its image and sends it to the developed graphical user interface installed in the computer. The operator uses the cursor to label the checkboxes in sequence to store their coordinates. When in automatic operating mode, a cycle of vast data input by computer vision would be as follows. The software commands the controller through RS232 to send an output signal to the study feeder. One page from the top of a stack of questionnaires is then fed forward by the study feeder. The CCD overhead continuously grabs images and sends them back to the computer via the image interface. The software subsequently snaps a frozen picture of that particular questionnaire. This picture is composed of pixels in a two-dimensional array that forms a still image. To speed up the process, the image is often segmented into small Regions of Interest (ROIs), each of which contains a checkbox whose answer is processed and recognized therein. Once the image processing for all checkboxes is completed and stored, it goes to the next cycle.

Due to the properties of the camera and its lens, the image of a questionnaire is inevitably skewed and distorted, no matter how the questionnaire is placed either by hand or the study feeder. The consequence makes it impossible for the software to locate most of the checkboxes accurately. To overcome this, it is necessary to consider the geometric camera model described by

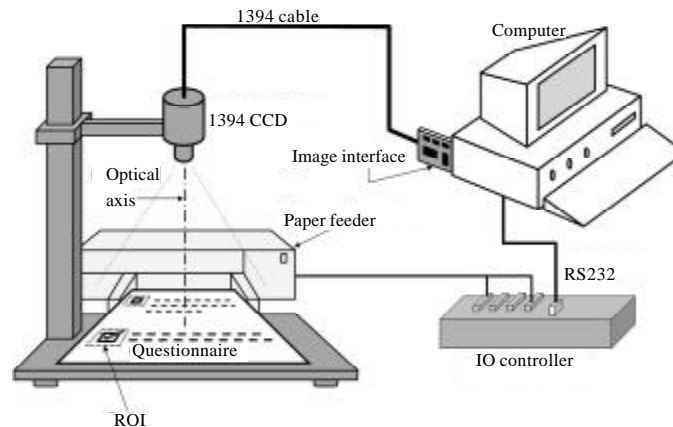


Fig. 1: The schematic diagram of automatic questionnaire input system

Forsyth and Ponce (2003) and calibrate it to accommodate those aberrations appeared in this specific application.

PERSPECTIVE PROJECTION MODEL

Consider a fronto-parallel plane Π_r where the questionnaire is placed. h is the distance from the pinhole camera to the plane as shown in Fig. 2. According to the perspective projection model, the image plane Π_i is located at a distance called the focal length f ahead of the camera. Any point P on Π_r has an image p on Π_i viewed from the camera. Usually it is convenient to describe an object, for example a questionnaire, in Π_r with another set of origin O' and coordinate axes (i_q, j_q) . This means that for the notations denoted by wP and qP of the same point P on Π_r in the world frame (W) and the questionnaire frame (Q), respectively, there is a rigid transformation including the rotation of axes at an angle ϕ along with a translation $\overline{OO'}$. In general, the origin of the image plane is at the upper left corner c and the principal point $o = [u_0, v_0]$ is the center where the optical axis penetrates Π_i .

In the calibration procedure, there are intrinsic and extrinsic parameters needed to be considered to realize the spatial relationship of objects through the camera lens. The homogeneous perspective projection equation is:

$$p = M^qP \tag{1}$$

where, $M = K [R \ t]$ is the perspective projection matrix which consists of two parts: the matrix of intrinsic parameters K and the extrinsic rigid transformation $[R \ t]$.

With the intrinsic parameters of camera, every point on Π_r mapped to Π_i suffers some aberrations. Without loss of generality, neglecting the component along the optical axis, an image point $p = [u \ v \ 1]^T$ on Π_i is mapped from $P = [x \ y \ 1]^T$ on Π_r by the homogeneous equation $p = KP$, where the matrix of intrinsic parameters:

$$K = \begin{bmatrix} \alpha & -\alpha \cot\theta & u_0 \\ 0 & \beta/\sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

also known as a planar affine translation. α and β are the magnifications along i_p and j_p . The angle between i_p and j_p is θ which is not exact 90 degrees if there exist some manufacturing errors on the lens. One can simplify the model for the system in Fig. 1 such that $\alpha = \beta$ since the cell size of the CCD camera is squared ($4.65 \mu\text{m} \times 4.65 \mu\text{m}$). Furthermore, the center of the CCD matrix coincides with the principal point o , so that both u_0 and v_0 are one half of the resolution of the image along their respective axes. θ

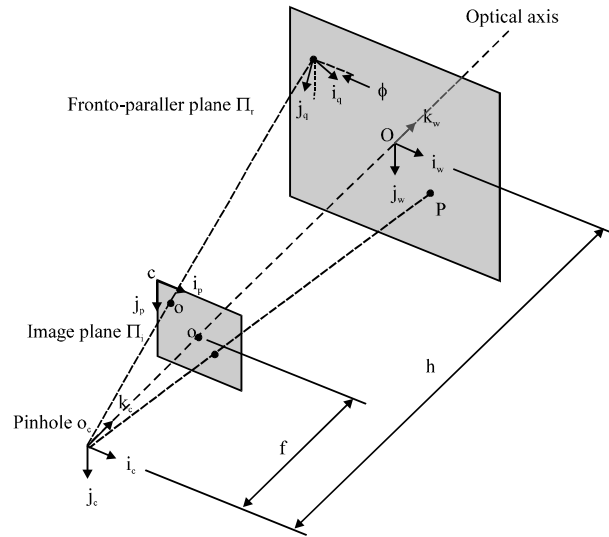


Fig. 2: Perspective projection model of the system

is taken to be 90 degrees provided that the qualities of our CCD camera and lens are actually at the industrial level. Therefore, Eq. 2 reduces to:

$$K = \begin{bmatrix} \alpha & 0 & u_0 \\ 0 & \alpha & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

For the extrinsic parameters, consider the coordinate changes of rigid transformation which includes rotation and translation. Since the component along optical axis is omitted, the transformation from the frame (Q) to the world frame (W) due to the angle ϕ is simplified as:

$$[R \ t] = \begin{bmatrix} r & {}^wO_q \\ 0^T & 1 \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi & {}^wO_q \\ \sin\phi & \cos\phi & \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where, r is the 2×2 rotation matrix and ${}^wO_q = \overline{{}^wOO'}$ is on the two dimensional plane. It is easy to see that homogeneous equation ${}^wP = [R \ t]^q P$ transforms the coordinates of a point on Π_r from the frame (Q) to the frame (W).

Besides the aforementioned aberrations, there is another one called radial distortion which one needs to account for in the general homogeneous perspective projection equation:

$$p' = DM^qP \tag{5}$$

where:

$$D = \begin{bmatrix} d(p') & 0 & 0 \\ 0 & d(p') & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

is the radial distortion matrix. Note that the point p' lies on the CCD image which is not exactly the image plane Π_i . The shrinking coefficient $d(p')$ depends on the position of p' or, specifically speaking, the length of $\overline{op'}$ with the relation:

$$d(p') = \frac{|\overline{op'}|/f}{\tan(|\overline{op'}|/f)}. \quad (7)$$

This shrinking distortion is referred to as the fisheye effect.

In order to locate the positions of checkboxes on a questionnaire page through the CCD image, it needs to realize all of the intrinsic and extrinsic parameters, or to calibrate the camera, before the system can be functioning accurately. If all the parameters are given or easily measured, the job is straightforward. Usually, it is not the case.

METHODS

It is noted that each checkbox appearing in the CCD image undergoes translation, rotation, aberrations and fisheye distortion. During image processing, certain corrections have to be made so that all of the stored coordinates of the checkbox centers, i.e., p_1, p_2, \dots , on the

initial page can map to the local coordinates of the following pages.

Positioning of the locating marks: Each questionnaire is a printed page of rectangular size. For the purpose of proper affine translation, two L-shaped locating marks are printed on both upper-left and upper-right corners of the page along with the contents of questions, as shown in Fig. 3. Through image processing of the edges finding as discussed in Pu *et al.* (2009), the inner corners of both locating marks can be determined. The left one is set to be the origin of the local coordinates where u -axis is horizontal to the right and v -axis is vertical downward. These two points will later be mapped to the corresponding positions on Π_i by the fisheye correction.

By the edges findings, the system can locate a checkbox with a cursor in the initial locating mode as follows. One point inside a checkbox is assigned by hand. Starting from that point, the corresponding ROI is defined larger than and enclosing the checkbox. Take projections of the foreground pixels onto both horizontal and vertical axis. The dramatic changes of foreground projections designate where those four edges are. Once extracting an interior area of a checkbox, the system can determine where the center is and label it with p_i , where $i = 1, 2, \dots$, number of checkboxes.

Fisheye correction: Every camera has certain range of view angle that causes an effect of radial distortion. To accurately locate the checkboxes on each questionnaire, it is necessary to correct this effect. The questionnaire in CCD image looks distorted in Fig. 4. The actual page

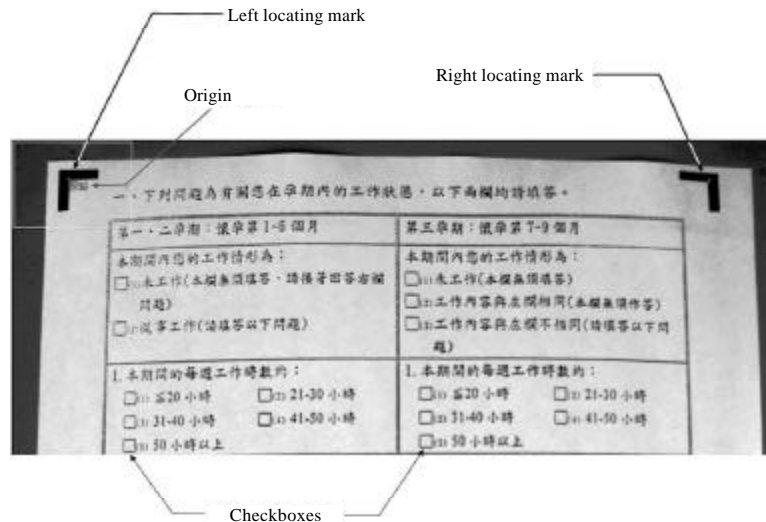


Fig. 3: The positioning of left and right locating marks

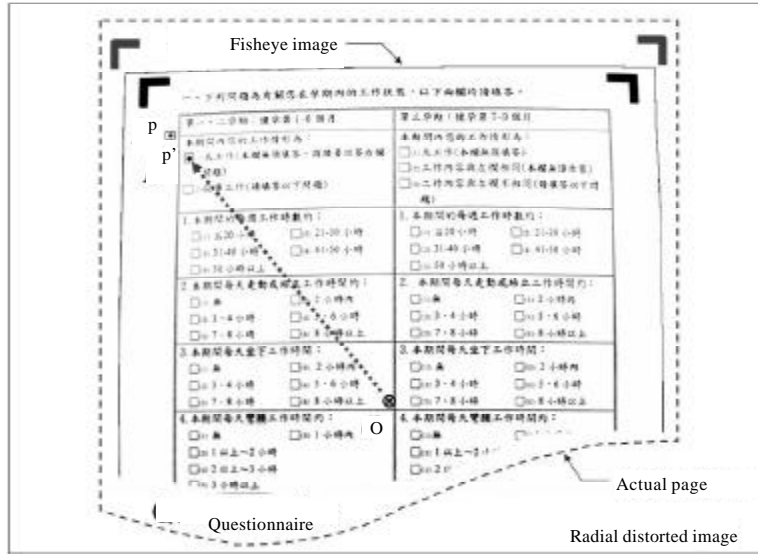


Fig. 4: The shrinking effect of a fisheye image when viewing a questionnaire through CCD

tends to shrink toward the center o where the optical axis penetrates. It's because the fisheye image is a projected image from the actual page onto the visual sphere of radius f , also known as the focal length. Let a checkbox center p' and the line segment op' be on the CCD image. The actual location of p on Π_i can be found by the ratio $op'/op = d(p')$ depicted in Eq. 6 and 7. Note that all of the lengths above, including f , are measured in pixels.

Through different lenses, the CCD image shows objects differently in size. Especially, the focal length varying in the view will cause the distance f on Π_i in Fig. 1 to change its value. Since the focal parameters of multiple lenses are usually coupled and unknown, it is possible to conduct a calibration experiment to determine f . Instead of using a reference pattern or a grid in the image for calibration, the authors directly use a piece of questionnaire page to derive the optimal f in pixels for a fixed height h by trial and error. Once it is done, the radial distortion matrix D caused by the fisheye effect can be determined.

Rigid transformation of coordinates: When each questionnaire is placed and ready for image grabbing, it goes skew inevitably. There are a translation and a rotation between the world frame (W) and the questionnaire frame (Q). This transformation not only takes place on Π_r but also on Π_i . As matter of fact, our goal is to locate the corresponding positions of all checkboxes for each questionnaire sent to the computer screen which is actually the distorted Π_i . For simplicity, let

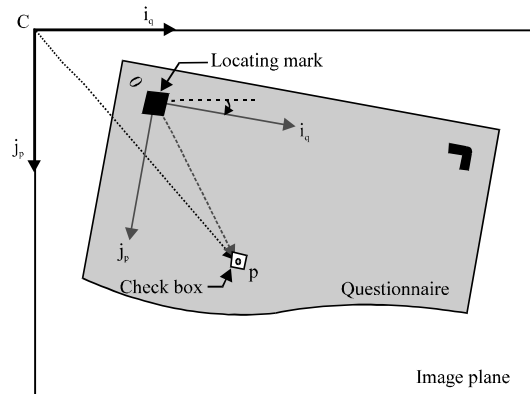


Fig. 5: The schematic diagram of a questionnaire under translation and rotation

the coordinate axes i_q and j_q map to the image plane (P). Instead of doing it back and forth between Π_i and Π_r , one can only perform the transformation on Π_i , which still follows the relation in Eq. 4 except that the translation needs some modifications. Moreover, the matrix of intrinsic parameters K in Eq. 3 does not need to be considered if the transformation takes place in Π_i only.

A questionnaire image on Π_i after fisheye correction is schematically shown in Fig. 5. The previously stored positions of both locating marks can determine the rotation angle ϕ . Since ${}^q\overline{op} = {}^q\overline{cp} + {}^q\overline{oc}$, it follows that:

$${}^q\overline{p} = [R \ t]^p p, \quad (8)$$

where,

$$[R \ t] = \begin{bmatrix} R^{-1} & {}^a o_p \\ 0^T & 1 \end{bmatrix}$$

is defined in Eq. 4. Note that ${}^a o_p = {}^a \overline{oc}$ represents the feature of translation. The positions of both locating marks after fisheye correction are then used to determine ${}^a \overline{oc}$ and the rotation angle ϕ . The centers of all checkboxes in the initial questionnaire page correspond to the reference coordinates which will be provided later for automatically locating the checkboxes in the coming questionnaires shown on the CCD image. This process needs to do the reverse of all the mapping mentioned above.

EXPERIMENTS SETUP

In this section the calibration experiments are designed to determine the optimal focal length f in pixels at different h by trial and error. h is the height, also called the shooting range, from the overhead CCD lens to the questionnaire along the optical axis as shown in Fig. 1. For each lens at the fixed focal parameters, different shooting range corresponds to different f . If f has a certain inaccuracy, this error will propagate through Eq. 7 and let the fisheye correction lose accuracy as well. To some extent, this will cause a failure in locating a checkbox by mapping from the reference point of the initial page

which is unacceptable during the automatic operating mode.

To calibrate the optimal focal length f , it is designed to perform the location of four corner checkboxes depicted in Fig. 6 at a fixed height. By varying f one records the sums of the deviations, denoted by S in pixels, between the recognized centers (locating points) and the actual centers of those corner checkboxes. It is noted that the specified checkboxes are chosen to increase the sensitivity of S to f . The height is set firstly to be 28 cm to include whole questionnaire in the CCD image. In this case, according to our experience, the range of f is initially set between 1,000 and 5,000 pixels. For each increment of the thousands digit in f the software locates the corner checkboxes and computes S . The smallest value of S determines the likely value of f . Next, the range of f is narrowed down to be 2,000 and centered at the likely value. The series of S is computed again for the increment of the hundred digit in f and, accordingly, indicates a new likely value of f . The same procedure repeats until the likely values of f have the significant Figure down to the unit digit. Their median is called the optimal value of f . Once this is done, the same calibrations start over for the heights at 26, 27, 29 and 30 cm to accommodate the fact that the shooting range h may vary and is to decrease gradually as each page sent by the feeder is piled on top of the stack during the automatic operating mode.

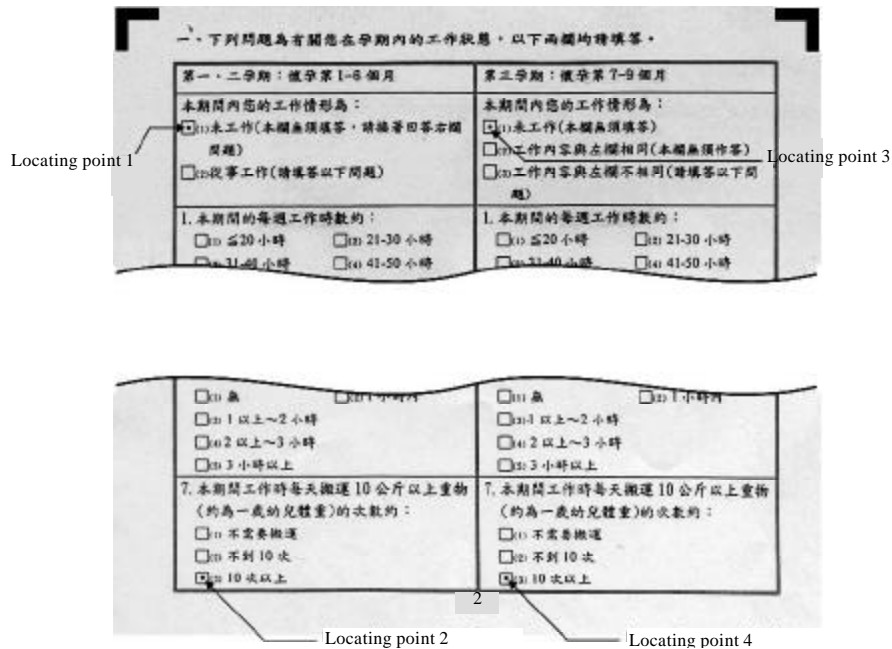


Fig. 6: Four corner checkboxes used in the calibration

RESULTS AND DISCUSSION

As f varied from 1,000 to 5,000 in the increment of 1,000, the image processes are performed to locate the centers of those four corner checkboxes. The sums of their deviations were computed and are shown in Fig. 7 where S starts at the most erroneous value when f is set to be 1,000, decreases very quickly as f increases and even down to zero when f is equal to 3,000. As f goes higher, S increases gradually and approaches to 50 pixels. It is obvious that $f = 3,000$ is the likely value and falls into the optimal range which lies in the smaller interval from 2,000 to 4,000. Next, the same procedure is conducted for f varying on that interval in the increment of 100. The results are shown in Fig. 8 which demonstrates that the optimal range of f is narrowed down to the interval between 2,900 and 3,100. Fig. 9 shows that within that new interval in the increment of 10 the optimal range of f is between 2,990 and 3,030 with the significant figure up to the tens digit. Finally, the diagram depicted in Fig. 10 shows that the optimal range of f is found to be the exact interval from 2,997 to 3,021 where the optimal value of f is then chosen to be 3,009 which concludes the calibration for $h = 28$ cm.

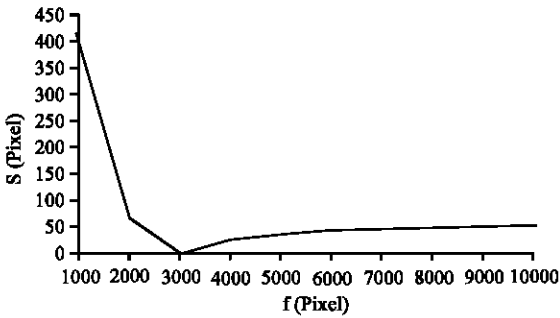


Fig. 7: The sum of deviations of four locating points for f varying from 1,000 to 10,000 in the increment of 1,000

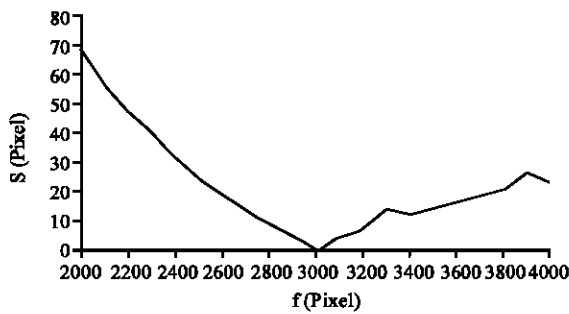


Fig. 8: The sum of deviations of four locating points for f varying from 2,000 to 4,000 in the increment of 100

Recall that $S = 0$ indicates that the corresponding f makes the locating point coincide with the actual center of a checkbox which is very easy to extract the interior out of that checkbox. Other nearby values of f , however, may also succeed in extracting the interior as long as S is small enough so that the recognized center falls inside the checkbox and departs at least a couple pixels from the four edges. In other words, there is a buffer for f to fall off its optimal range and to, still, be able to locate a checkboxes. In the case above f with a value within the interval from 2,800 to 3,200 will make the locating point deviate only 4 pixels at most from the actual center. There is no problem in processing a checkbox with size larger than 12×12 pixels.

Furthermore, the calibrations are repeated one by one for the heights of 26, 27, 29 and 30 cm. Their results are shown in Table 1. It shows that as the shooting range h

Table 1: Results of the calibrations for different shooting ranges

Shooting range h (cm)	Optimal value of f (pixel)	Optimal range of f (pixel)	Minimum of S (pixel)
28	2792	[2778, 2817]	1
29	2893	[2874, 2911]	1
28	3009	[2997, 3021]	0
29	3097	[3085, 3109]	0
30	3195	[3190, 3199]	1

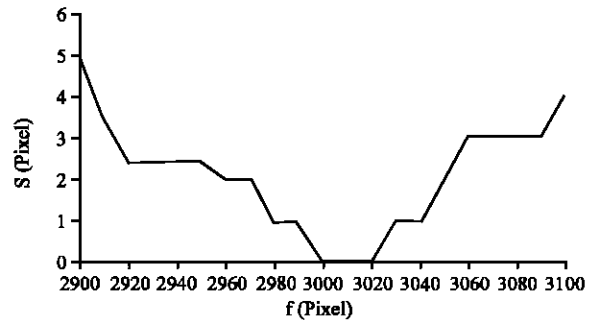


Fig. 9: The sum of deviations of four locating points for f varying from 2,900 to 3,100 in the increment of 10

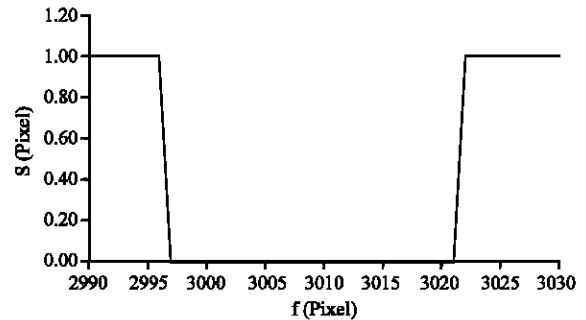


Fig. 10: The sum of deviations of four locating points for f varying from 2,990 to 3,030 in the increment of 1

goes higher the optimal focal length f gets larger as well. The corresponding optimal range of f , however, gets smaller. These results provide us with important information when tuning the parameters of the system. On the other hand, the sum of the deviations S may not reach zero in every case even when f falls into the optimal range. It is because most properties in the calibration have unit in pixel which must be discrete. Any error will inevitably propagate through the equations mentioned above.

CONCLUSION

This study, using the perspective projection model, presented a method to calibrate the camera in the developing questionnaire input system. The skews and aberrations in the image, especially the fisheye effect, were amended and implemented in the software in order to locate each checkbox accurately. Most of the intrinsic and extrinsic parameters of the camera could be either deduced or omitted; except that the focal lengths at various shooting ranges were determined by experiments. The results show that when setting the focal length at its optimal range the software is able to locate the center of a checkbox with the error no more than a pixel.

REFERENCES

- Ahmadi, F.F., M.J. Valadan Zoej, H. Ebadi and M. Mokhtarzade, 2008. Road extraction from high resolution satellite images using image processing algorithms and cad-based environments facilities. *J. Applied Sci.*, 8: 2975-2982.
- Alhadidi, B., M.H. Zu'bi and H.N. Suleiman, 2007. Mammogram breast cancer image detection using image processing functions. *Inform. Technol. J.*, 6: 217-221.
- Forsyth, D.A. and J. Ponce, 2003. *Computer Vision: A Modern Approach*. Prentice Hall, New Jersey, USA., ISBN-10: 0130851, pp: 693.
- Gonzales, R.C. and R.E. Woods, 2002. *Digital Image Processing*. 2nd Edn., Prentice-Hall, New Jersey, ISBN-10: 0201180758.
- Koc, A.B., H. Silleli, C. Koc and M.A. Dayioglu, 2007. Monitoring of dry powder mixing with real-time image processing. *J. Applied Sci.*, 7: 1218-1223.
- Madhloom, H.T., S.A. Kareem, H. Ariffin, A.A. Zaidan, H.O. Alanazi and B.B. Zaidan, 2010. An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold. *J. Applied Sci.*, 10: 959-966.
- Moshashai, K., M. Almasi, S. Minaei and A.M. Borghei, 2008. Identification of sugarcane nodes using image processing and machine vision technology. *Int. J. Agric. Res.*, 3: 357-364.
- Oppenheim, A.N., 1992. *Questionnaire Design, Interviewing and Attitude Measurement*. Pinter Publishers, London, New York.
- Pu, Y.R., Y.C. Chang and S.H. Lee, 2009. Development of a questionnaire input software by machine vision. *Proceedings of 2nd International Symposium on Knowledge Acquisition and Modeling*, 30 Nov. 1 Dec, Wuhan, China, pp: 225-228.
- Raghava, N.S. and G.S. Kumar, 2006. Image processing for synthesizing lung nodules: A experimental study. *J. Applied Sci.*, 6: 1238-1242.
- Tseng, W.H., 2005. Building an integrated OMR-and-paper-based data automatic collection support system. Master Thesis, Institute of Health Information and Decision Making, National Yang Ming University.
- Tunney, W.P., 2006. Method and system for identifying multiple questionnaire pages. United States Patent 7,031,520 <http://www.freepatentsonline.com/7031520.html>.