



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

A Real-Time Vision Tracking System Using Human Emulation Device

¹Ko-Chun Chen, ²Kuei-Shu Hsu and ¹Ming-Guo Her

¹Department of Mechanical Engineering, Tatung University,
40 Chung-Shan North Rd. 3rd Sec., Taipei, Taiwan 104, Republic of China

²Department of Applied Geoinformatics, Chia Nan University of Pharmacy and Science,
Taiwan 717, Republic of China

Abstract: This study presents a real time effective human emulation vision tracking system, which is built up by two webcams and can detect objects in a 3-D space. Because of our programming in this software program, it becomes suitable for teaching and research. In this paper, we set up two webcams by an optical parallel method and deliver images to PC by a USB (universal serial bus) interface. The computer carrying out the image processing using an effective tracking rule to monitor and fast lock the object. Besides, this close-loop system contains two servo motors that control the webcams platform to pitch and yaw target centered in the monitoring image. It can measure the depth of the objects from two webcams by trigonometric function relationship. Finally, we make a virtual environment to verify the capability of this tracking system.

Key words: Real-time, visual tracking, parallel method, image processing, trigonometric

INTRODUCTION

In recent years, the visual technology of computer has great advancement. Because of the progressed ability of CPU (central processing unit), it is enable to deal with a large amount of data within a short time. Because the visual data was too huge to handle in the past time, we have to raise the speed of image process by upgrading the ability of CPU or developing more powerful visual tracking system.

By using the forms of image processing such as an image transform, image an enhancement, a threshold, a filtering, an edge detection and a thinning (Starck *et al.*, 2003; Olson and Huttenlocher, 1997; Henstock and David, 1996), we also can simplify the image data directly and decrease the workload of CPU.

The purpose of this paper is to build a suitable, effective visual tracking system. We use the inexpensive web camera (webcam) for an image input in the USB interface to build and construct a whole set of image tracing system economically (John *et al.*, 2003; Murray and Jennings, 1997). The major function of this visual tracking system is to track a target promptly and demarcate the plane coordinates of the target (Bertozzi and Broggi, 1998; Yau and Wang, 1999; Ohya *et al.*, 1998). This system measures the 3-D coordinates of the objects upon the relationships between two webcams and the target (Barnard and Thompson, 1980). Besides, the DC servo motor produces pulse

signal in accordance with the system coordinate to maintain target in the center of images (Mohan and Nevatia, 1989; Winters and Santos-Victor, 2002). In order to produce a high effective, suitable and real-time visual-tracking-system (Kuo and Wu, 2002; Yue *et al.*, 2006), this study goes with several rules:

- Picking up the image input hardware, program language software and the calculating host
- Storing the real-time image and making sure that the image caught by webcam can be stored in the computer (Bjorkman and Kragic, 2004; Kim *et al.*, 2003)
- Processing the image data, such as a gray, image subtraction and image edge detection. Doing Research of the motionless objects tracking methods and motion objects tracking methods (Seara and Schmidt, 2004; Kidono *et al.*, 2002; Lobo *et al.*, 2003; Black and Ellis, 2006)

The visual tracking equipment with the webcam delivers the continuous motion image data to the USB interface computer shown in Fig. 1. The benefit of the USB interface computer is plugged in to simplify the procedures.

The image process uses a threshold to transform the motion analog data to a digital image data and to remove the unnecessary image data, so that we can get the useful image data.

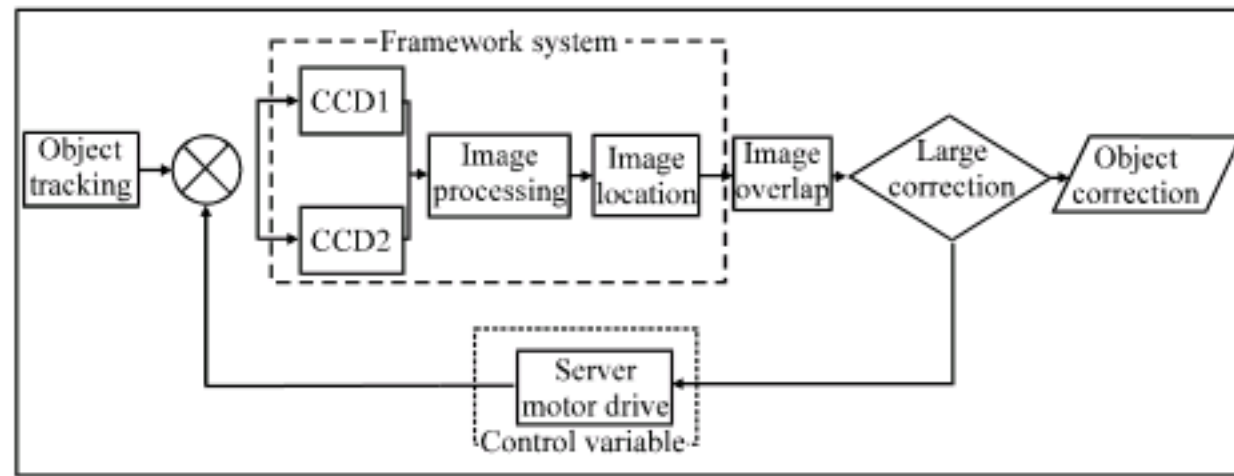


Fig. 1: System process flow

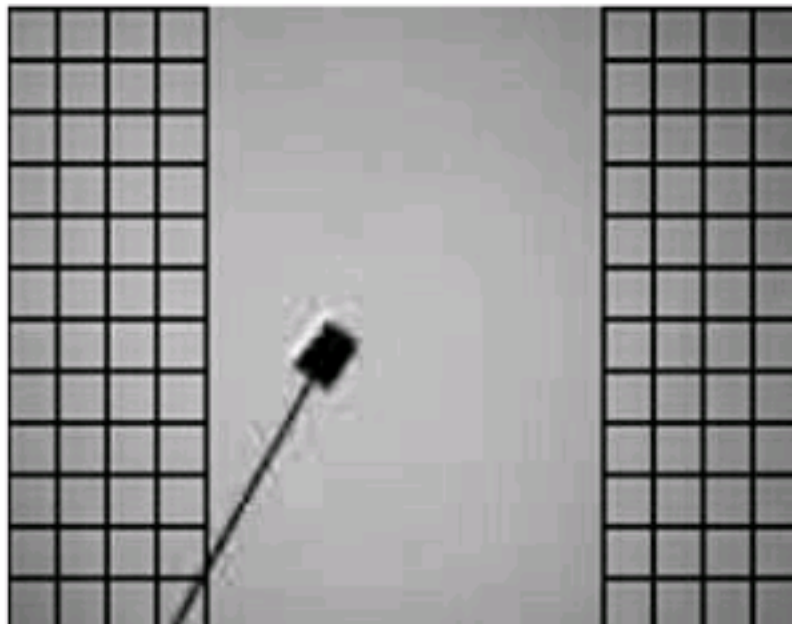


Fig. 2: The boundary of working area non-working area

The processed useful data can get the position of the target through the tracking judge rules. We can have the real-time tracking images when the target moves and system keep tracking it continuously. In addition, the DC servo motor controls the movement of Yaw axis of the webcams system platform. The system has a working area and not-working area. When the target moves into the working area, the system will have a controller of the computer-based to produce pulses signal to activate the servo motor, so that the system platform yaws. When the webcam yaws and changes the image position of the target, the target will get into the not-working area. As indicated in Fig. 2 that the area with the meshes is a working area, in which the image's pixel is 80×240. The area without meshes is a non-working area, in which the image's pixel is 160×240.

This system has two webcams, in which the distance of two webcams will not change by the movement of the axis. When the target moves to the webcams, there will be a relation of triangle between the webcams and the target. By using the formula of trigonometric, we can get the depth value between the webcams to the target, z .

The tracking platform can rotate with 80 degrees. When the target moves out of the area, the tracking platform will stop moving. Therefore, this system can not

keep tracking and analyzing to the target. When the target moves to the webcams again, the system will also activate the servo motor in accordance with the working area and non-working area.

IMAGE TRACKING SYSTEM AND TRACKING JUDGE RULES

This section will explain how to transform the analog image to a digital image. Besides, an image processing with hardware, an environment of the programming software, an image-software and a system calculating host will also be depicted in this paragraph.

As shown in Fig. 3, the visual tracking system has two webcams. The construction of webcams is built by optical parallel method. The images are transferred by USB interface. Moreover, there are three servo motors to control the webcams system platform. The precision of servo controller is around 0.1 degree. System program environment is built in the program of Visual Basic 6.0 in which the designer can correct the program easily. By using the latest ActiveX of VB6 software, the convenient interface can be constructed easily. The most effective way to process the image is to simplify the color data of each image. It is necessary for the analysis of the target position by filtering the unnecessary color data by the gray and threshold technology until the color data is left with a binary data. The binary is an important part for the image processing. The Binary data is easy in saving, handling and making a recognition in a specific characteristics (2,3). The important part of the binary processing is set with a threshold value of 0 to 255 around. We can define a value of 0~255. We define the numerical value of dark as 0 if the image value is bigger than this value. Likewise, we define numerical value white as 1 if the getting is smaller than the value. According to the aim of the position with two webcams, we can define three kinds of situations including an optical parallel, an optical being crossed and an optical not being crossed. This study will use the optical parallel to convert the

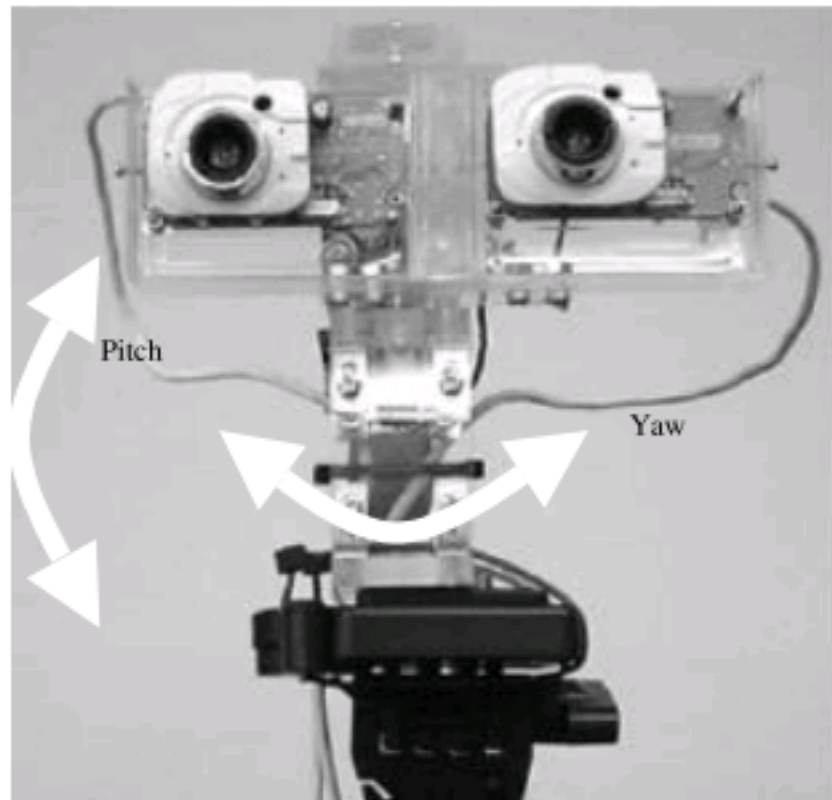


Fig. 3: A low cost tracking vision system platform

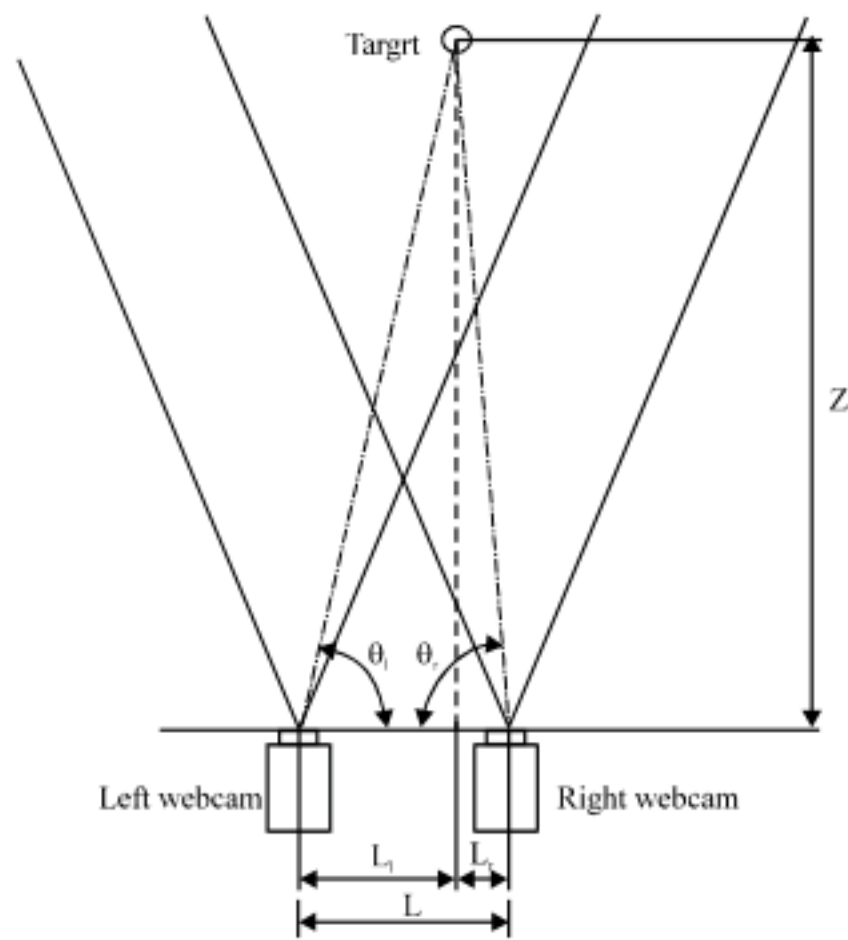


Fig. 4: The relationship between the camera and the object

relationship of image pixel and the distance. Therefore, we can get the distance from the camera to the object quickly.

Using the Fig. 4, we obtain:

$$Z = L_1 \tan \theta_l = L_r \tan \theta_r \quad (1)$$

and

$$L = L_1 + L_r \quad (2)$$

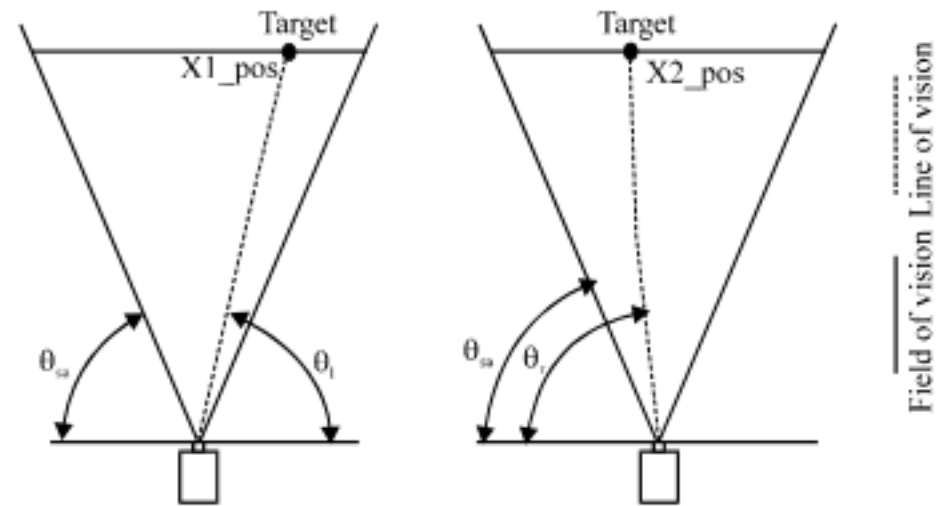


Fig. 5: Left target in the left webcam and right target in the right webcam

By using Eq. 1 and 2, it has:

$$Z = L \frac{\tan \theta_l \tan \theta_r}{\tan \theta_l + \tan \theta_r} \quad (3)$$

The distance L of the left webcam and the right webcam is measured as 7.6 cm. The value of a vertical distance z between the base line and the target is necessary for us. First, as indicated in Fig. 5 that there is a relationship between the left webcam, the right webcam and the target. Here, the true line is the view angle of each webcam, the dotted line is the extend line of the target from the webcam and the vertical dotted line is the distance of the target. After analyzing these two webcams individually in Fig. 4, the geometry of the target can be obtained shown in Fig. 5. As indicated in Fig. 5 that the position target in the left webcam is X1_pos and the position target in the right webcam is X2_pos. We can get X1_pos and X2_pos through the image tracking rules. When the target is in X1_pos, the real position will be located on the bold line in Fig. 5. It will have the view angle with respect to the left webcam. Equally, when the target is in X2_pos, the real position will be located on the bold line in Fig. 5 again. It will also have the view angle with respect to the right webcam. Finally, we can get the formula of θ_l, θ_r as below:

$$\alpha = 90^\circ - \theta_{sa}$$

Webcam Vision angle = 2α :

$$\text{Res2Ang} = \frac{2\alpha}{Xp}$$

$$\theta_l = 180^\circ - [\theta_{sa} + (X1_pos \times \text{Res2Ang})] \quad (4)$$

$$\theta_r = \theta_{sa} + (X2_pos \times \text{Res2Ang}) \quad (5)$$

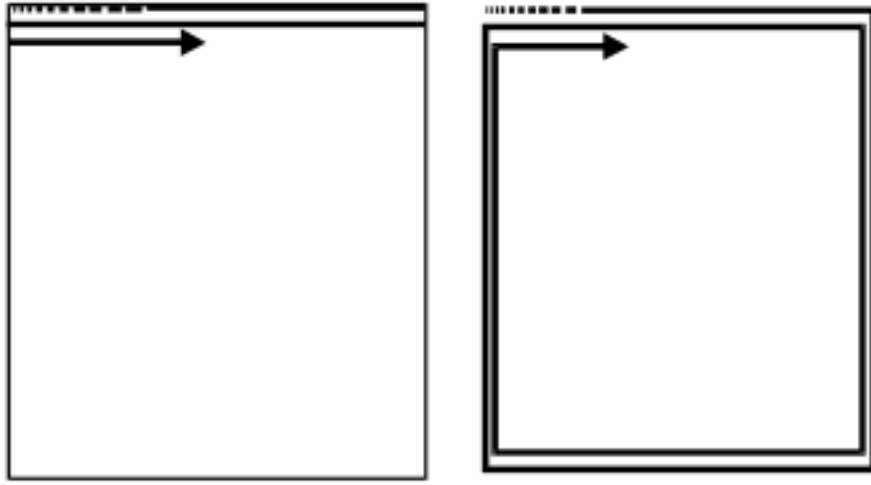


Fig. 6: Left scanning from left to right and a right scanning into the center

$$Z = L \frac{\tan \theta_i \times \tan \theta_r}{\tan \theta_i + \tan \theta_r} \quad (6)$$

In this research, the visual image is used to evaluate the object's distance. All the experimental data, accuracy and measuring speed are related to the calculation of the visual image. How to achieve a fast measurement, reduce the error in experimental work and efficiently promote the quality of the research becomes the main issue. As indicated in Fig. 6, the image judging begins by comparing colors with respect to the individual image pixel from the left to the right. After the image processing, the color of a single pixel is reduced from 16,770,000 colors to 2 colors. For an image with 320×240 pixels, the total color-comparison is decreased 77361 times.

The image-judgment-method is used to recognize the object's location inside the image. An efficient image-judgment-method will not only speed up the object's measurement but also reduce the measurement error. The image-judgment-method is essential for the distance-measuring system.

In practical usage, the image-judgment-method used to seek the area projected by the light source is too accurate; therefore, it is inefficient in the image judgment. For a working piece with a given error range of 50±0.5 mm, it would be meaningless if the required accuracy of manufacturing reaches 50±0.05 mm. Therefore, to facilitate the imaging judgment, a fundamental judgment method needs to be planned, again. That is, the compound-image-measurement method becomes the modified image-judgment method. The fundamental judgment method maintains a fixed judging region and scans the image. However, the compound judgment method will adjust the judging zone in accordance with the various locations of a moving object. Therefore, the scanning zone shown in Fig. 7 will be adjusted automatically.

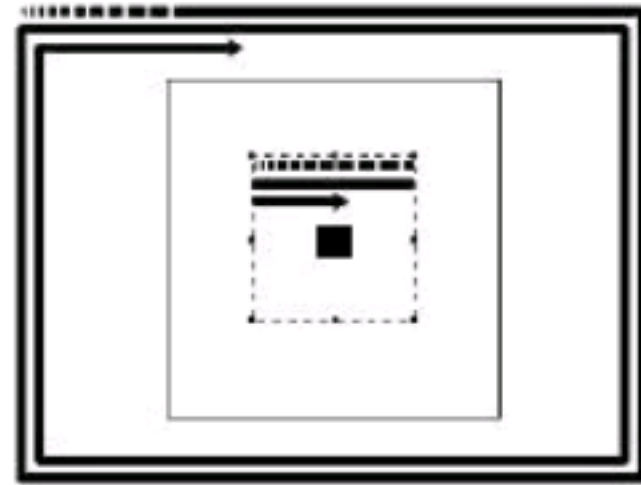


Fig. 7: Compound image judgment

FRAMEWORK ANALYSIS AND EXPERIMENT

This study uses 3D vision to judge the distance and the movement of motors. As indicated in Fig. 8, a 2D system is extended to a 3D system. We calculate the Z position of the target by trigonometry relationship. The plat servo motor will also keep the target in the center of frame.

The input image is 320×240 pixels. We calculate the z position of target by trigonometry relationship. The related experimental data is summarized in Table 1.

In this real-time tracking experiment, there are two computers A and B. The computer A will display a moving ball by a sine wave. There are 63 points in this wave line as shown in Fig. 9. It takes 2.03 sec in displaying these balls. Therefore, the frequency of balls is about 31 Hz and the data of moving track will be saved in a text file to compare with another experimental data.

As in dictated in Fig. 10, the original sine wave data will be saved in the file Sample A. Computer B will keep tracking the monitor of computer A. The final tracking data will be saved in the file of Sample B.

Finally, we combine the data of Sample A and B. The result in Fig. 10 will be saved. The points of Sample B are much more than these in Sample A. It is because the tracking of Sample A is focused on the center of the ball only. On the other hand, the Sample B displays the whole shapes of the balls.

The image distortion will have errors during the data sampling. To decrease the error, the promotion of the quality of image and the adoption of a better webcam are required. Because the frame of tracking system is based on points, the sine wave is composed of points. We divide the X-axis into 62 parts. In each part, we compare the tracking points of the Y-axis and the sine wave points to obtain an error value. The point of the D sine wave is at the center of the ball. If the tracking points are D₁ and D₂ that in dictated in Fig. 11, the average of D₁ and D₂ in the Y-axis will be calculated to evaluate how to close to the point of the sine wave.

Table 1: 3D vision experiment

Test No.	Actual distance (cm)	Webcam distance (cm)	Error (cm)
1	25	23.259	1.7
2	50	48.748	1.3
3	80	76.252	3.8
4	100	93.870	6.2
5	150	152.651	2.6
6	200	201.538	1.5

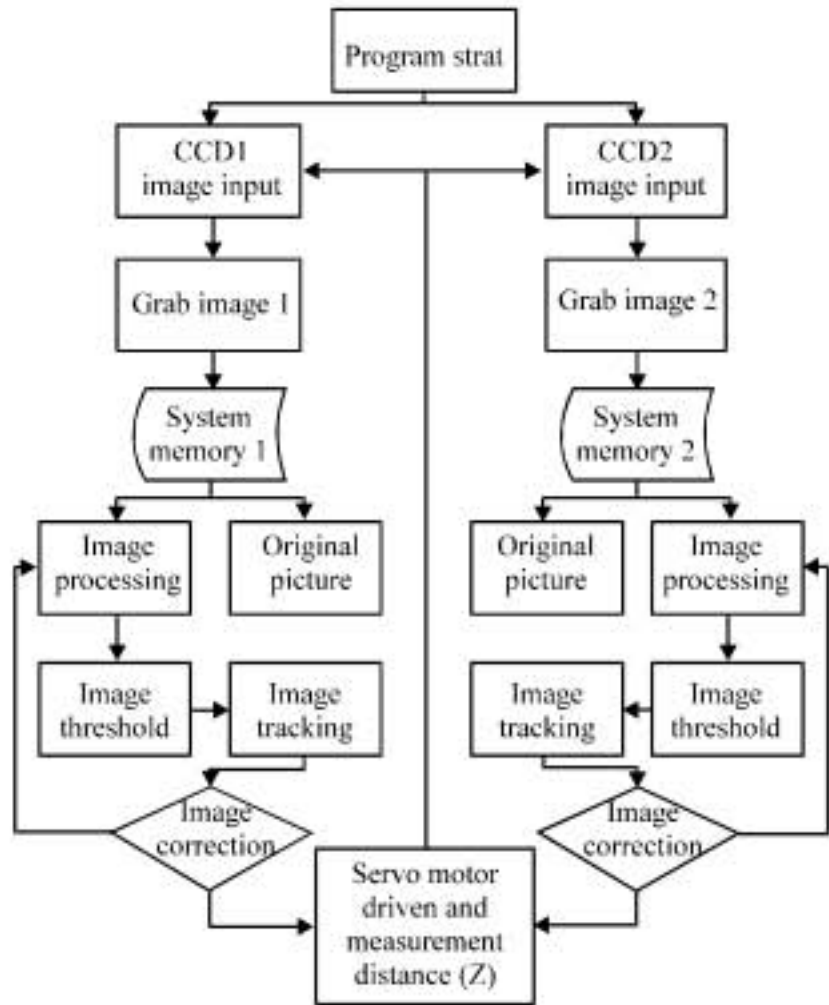


Fig. 8: A 3D-vision-program framework



Fig. 9: A moving ball by sine wave

Therefore, the original sine wave in the Y-axis is D_c . Besides, there are n points in the X-axis. The range of the sine wave is R and the average of tracking data in the Y-axis is \bar{D}_c . The error is:

$$Err(error) = \frac{1}{n-1} \sum_{i=1}^n \frac{|\bar{D}_c - D_i|}{R} \quad (7)$$

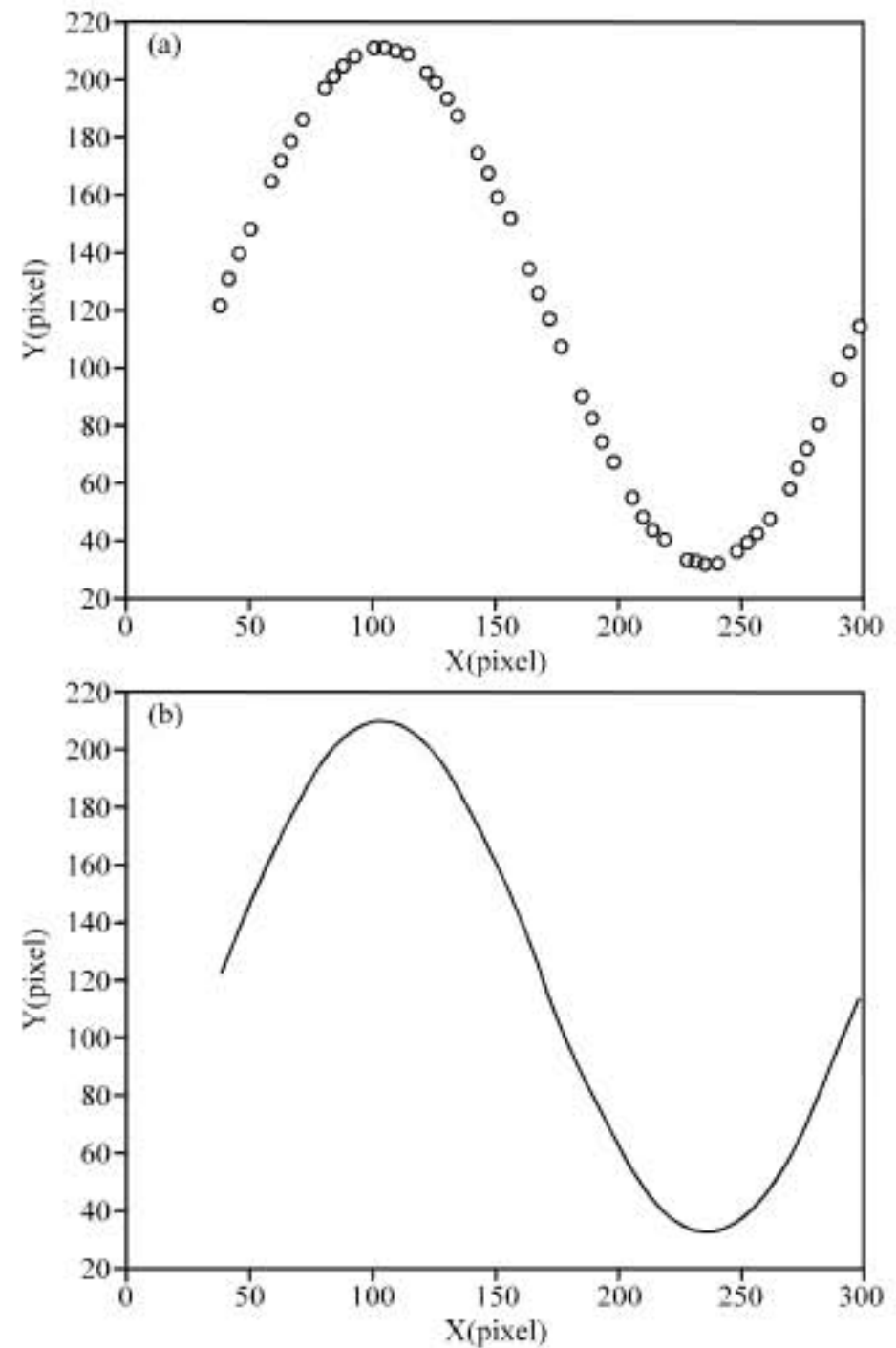


Fig. 10: (a) The data of sample and (b) the data of sample

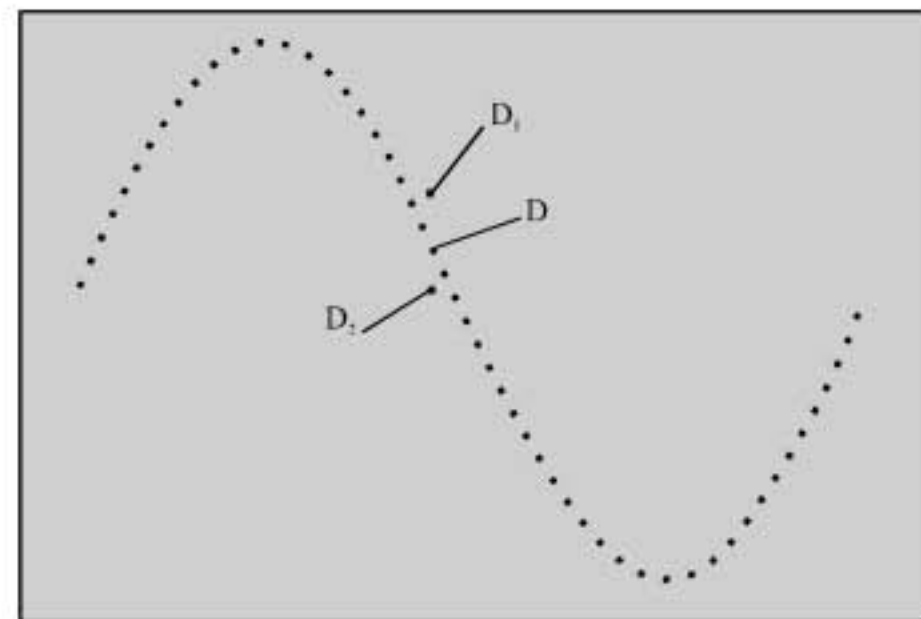


Fig. 11: The points of the sine wave

If the sample B has 162 points, it will distribute within 62 sections. In Fig. 12, the cross point in section $n-3$ is $S(65,157)$ and the rhombus points of this section are $A(66,150)$, $B(66,157)$, $C(66,157)$, $D(66,164)$ and $E(66,171)$.

We set S as a based point with $Y = 157$. The errors with respect to A, B, C, D, E by $(D_c - D_i)$ are:

$A = (150-157) = -7$; $B = (157-157) = 0$; $C = (157-157) = 0$;
 $D = (164-157) = 7$; $E = (171-157) = 14$; The total error is D_c ,

$$\bar{D}_c = 14/5 = 2.8$$

Where, $D_c = A+B+C+D+E = 14$

The average error is:

In Fig. 13, the range R is 146. The section error in the section n-3 is:

$$\text{Section error} = \text{Average error}/\text{Range}(R) = 0.019$$

The experimental results are shown in Fig. 14-18. The sphericity track and the grid line track is the sample,

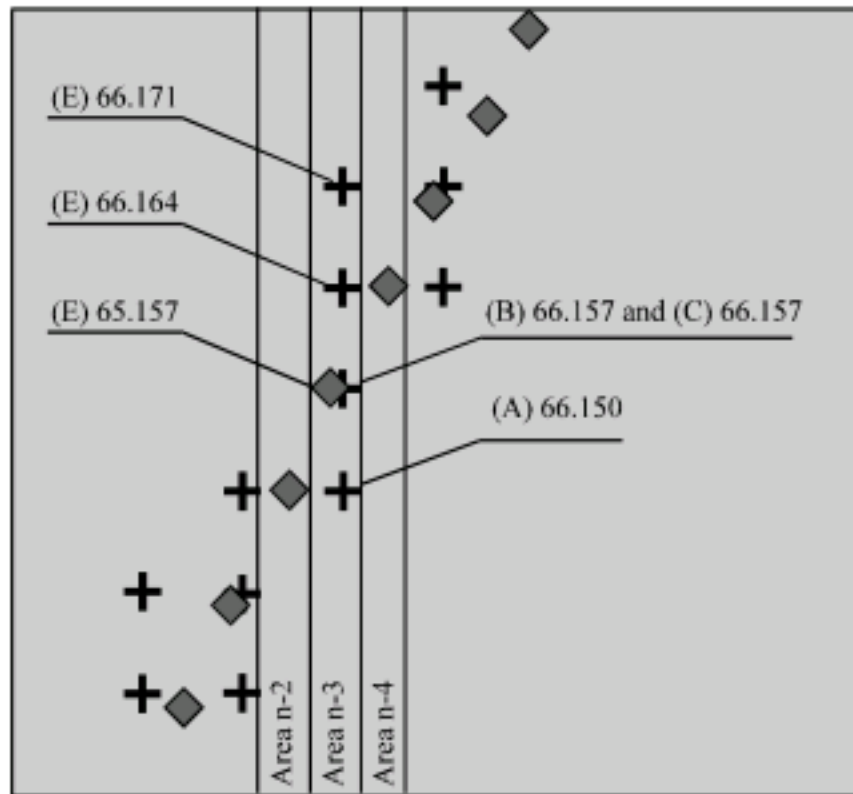


Fig. 12: The distribution of points

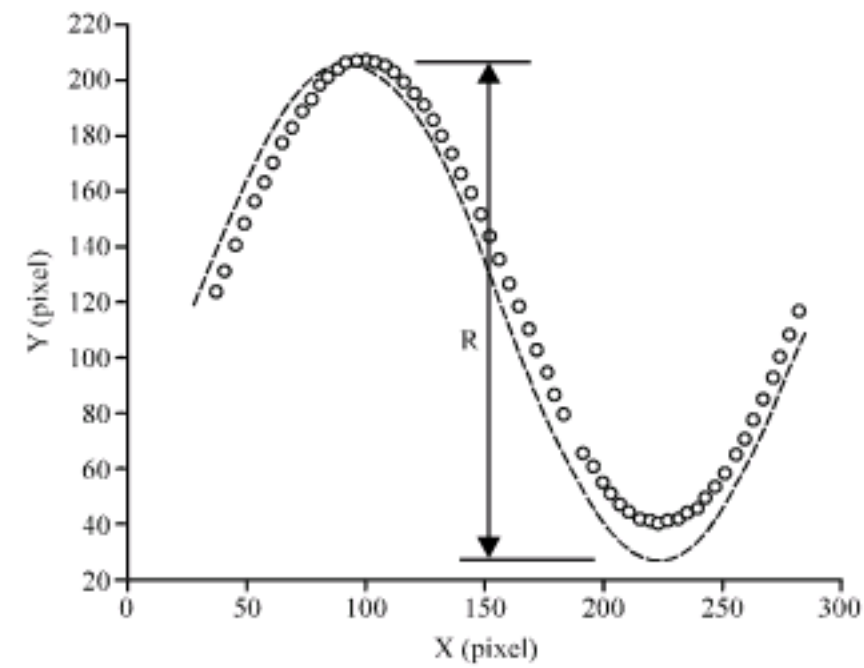


Fig. 13: The range (R)

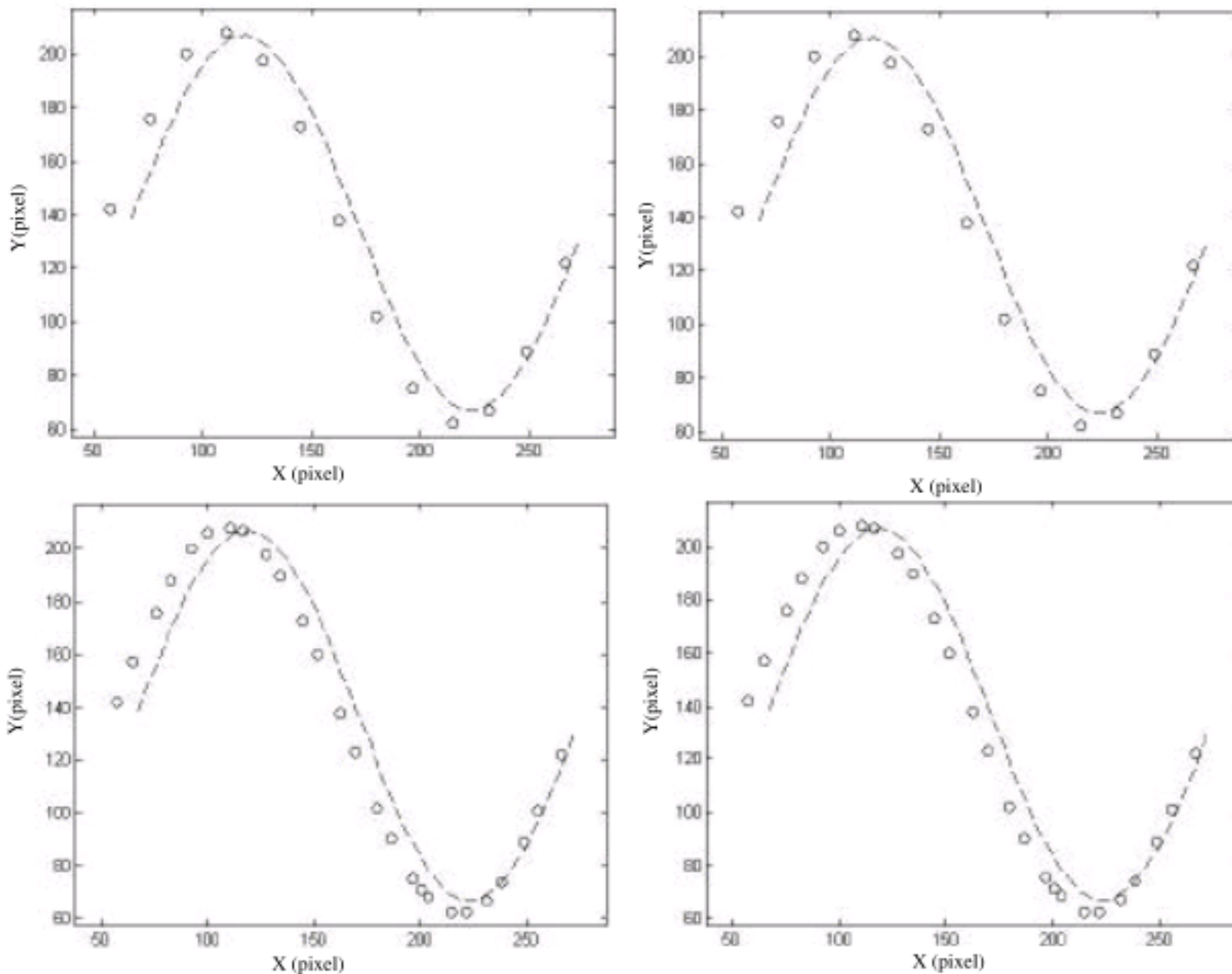


Fig. 14: The experimental results, when sampling density = 0, (a) result 1(3.73%); (b) result 2(3.88%), when sampling density = 2, (c) result 1 (2.55%) and (d) result 2(2.59%). The deisive least points = 2, (The percentage data is error value)

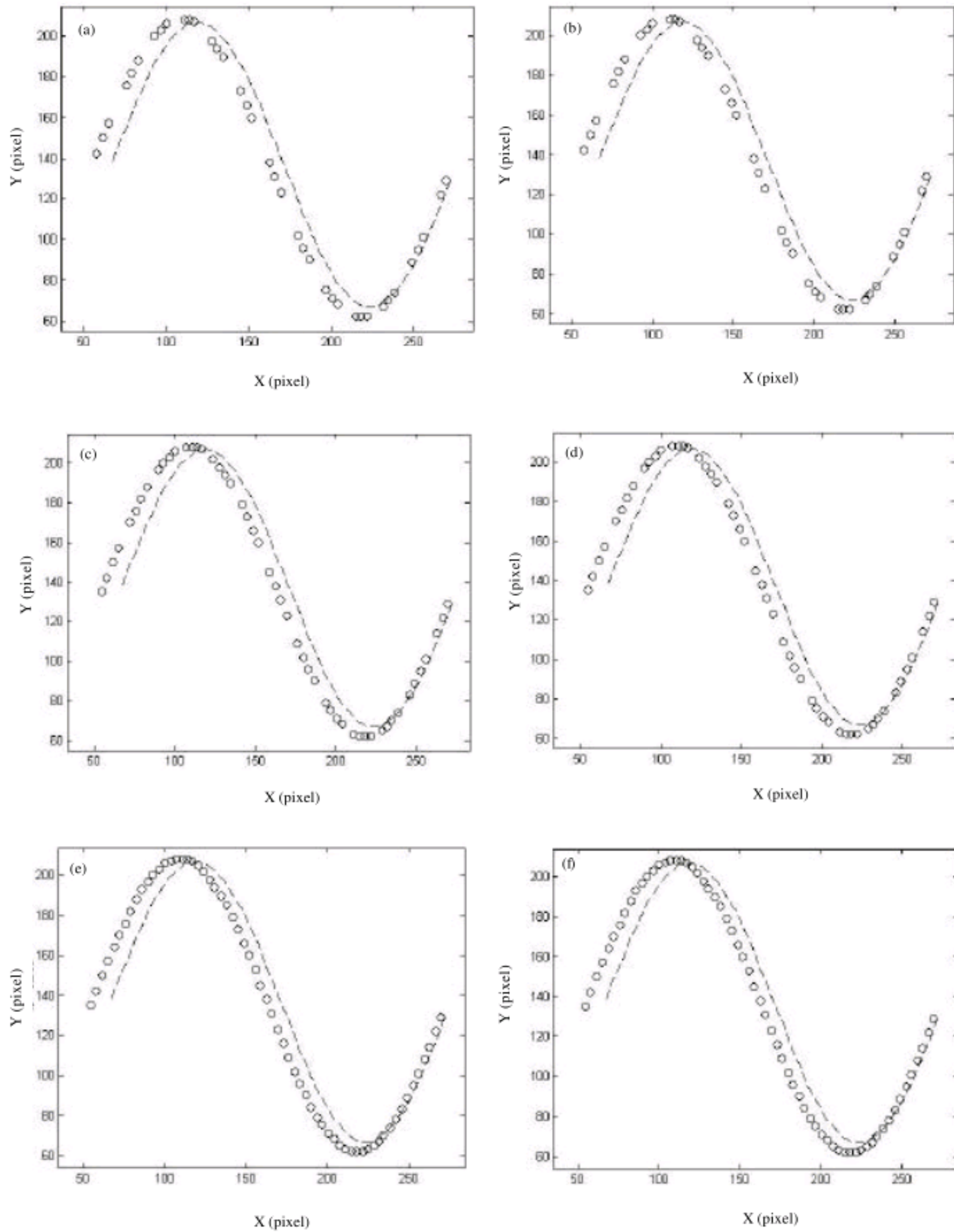


Fig. 15: The experimental results, when sampling density = 4, (a) result 1(2.63%); (b) result 2(2.58%), when sampling density = 2; (c) result 1 (2.46%); (d) result 2(2.40%) and sampling density = 8 (e) results 1(2.32%) and (f) result 2(2.15%)

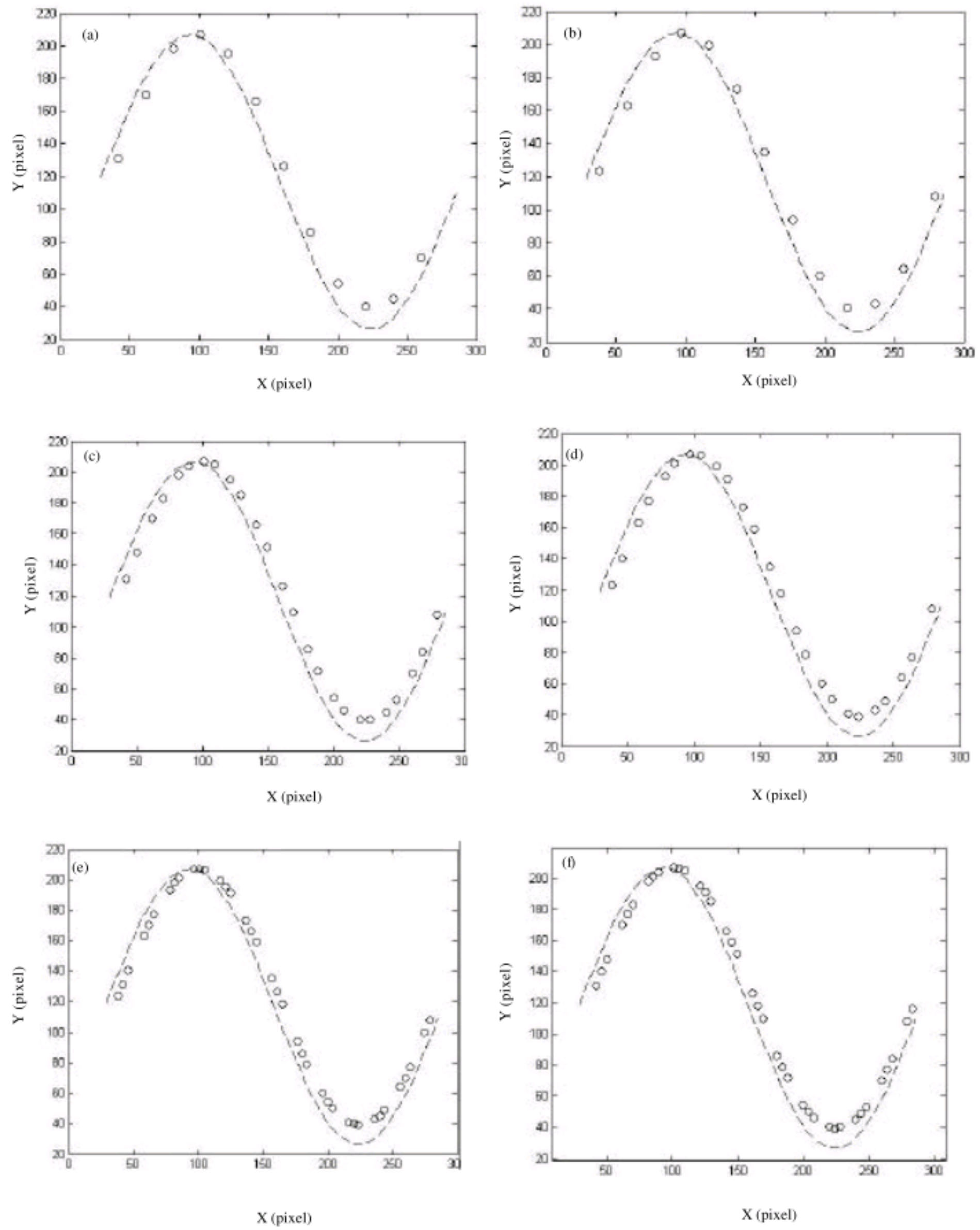


Fig. 16: Continued

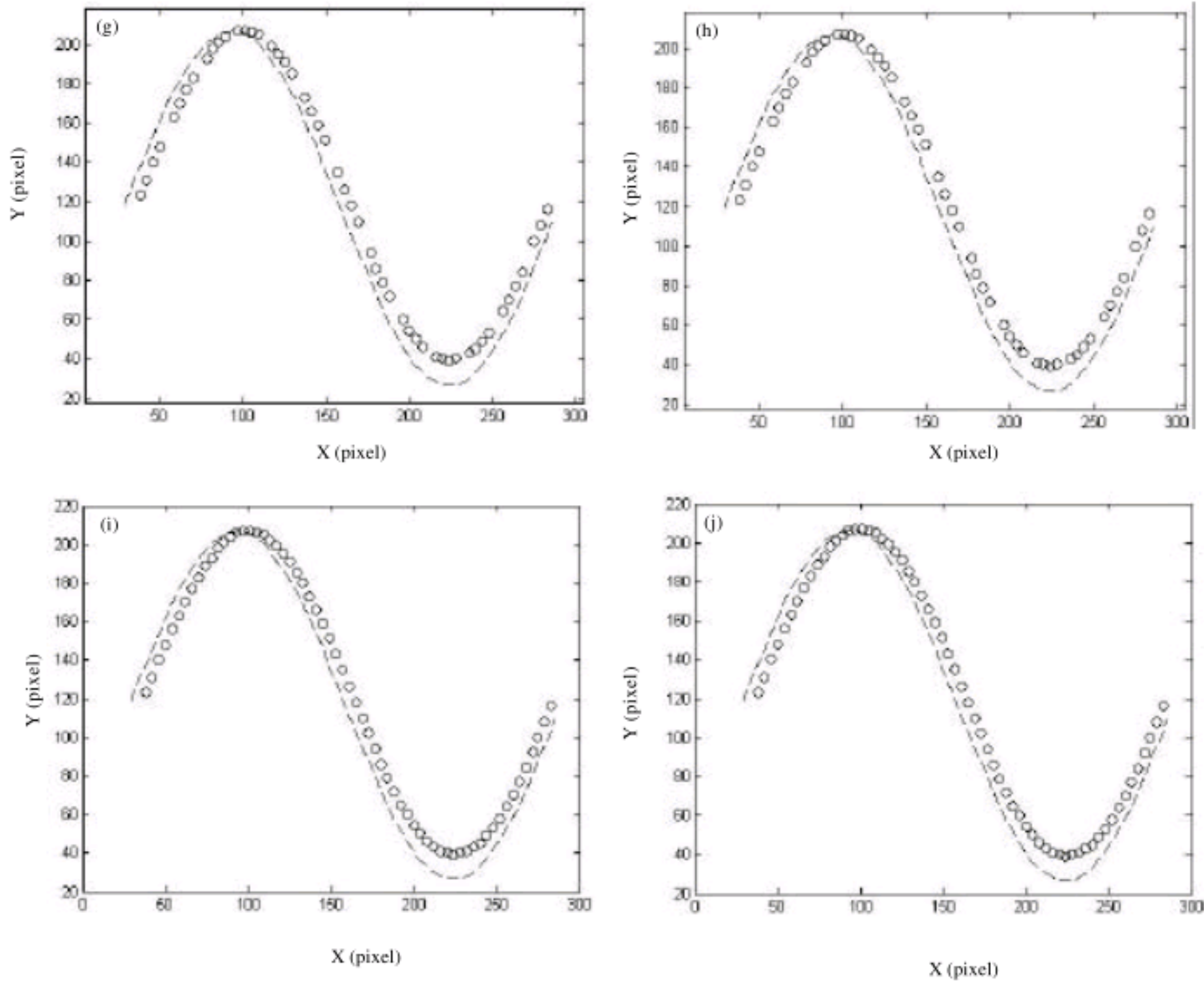


Fig. 16: The experimental results, when sampling density = 0, (a) result 1(5.62%); (b) result 2(4.92%), when sampling density = 2; (c) result 1(4.80%); (d) result 2(5.05%) and sampling density = 4 (e) results 1(4.68%) and result 2(4.73%), when sampling density = 6; (f) result 2(4.73%); (g) result 1(4.66%); (h) result 2(4.60%) and when sampling Density = 8; (i) result 1(4.80%); (j) result 2(4.50%), The deisive least points = 2, (The percentage data is error value)

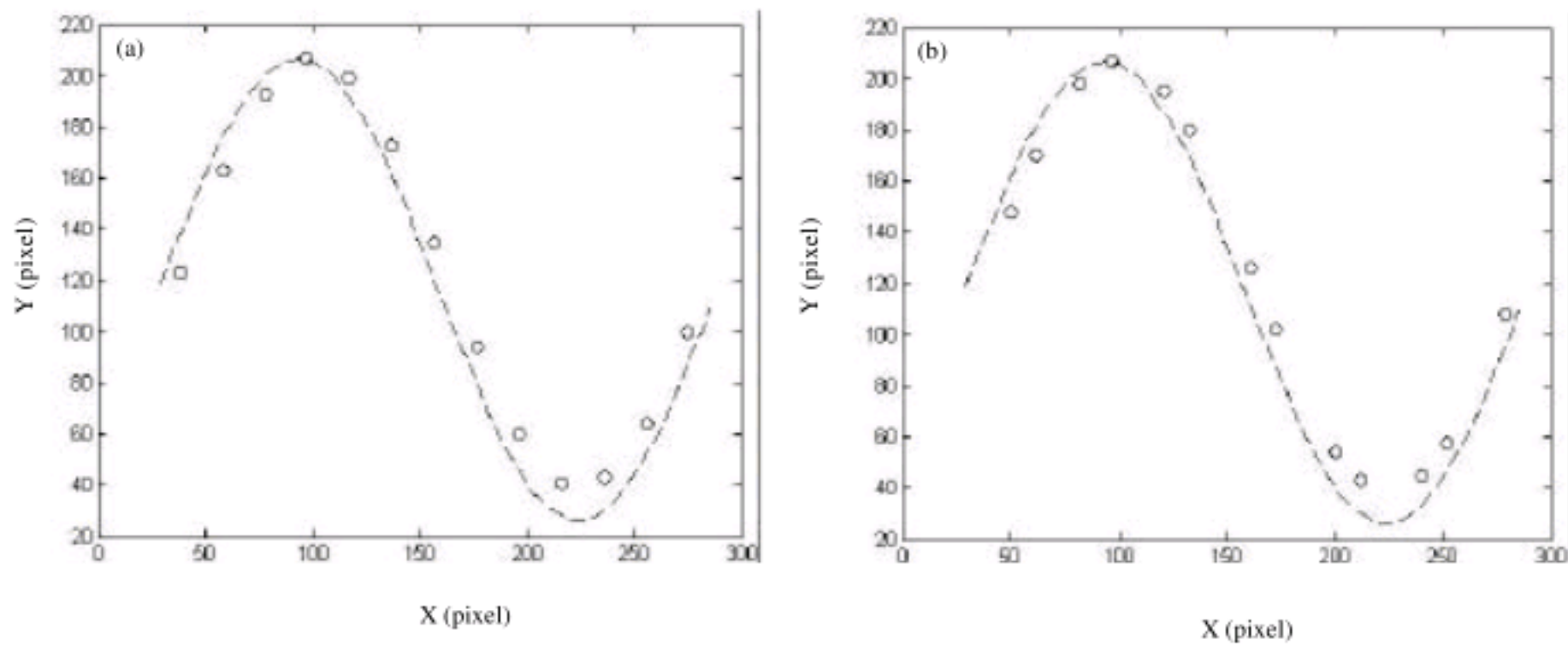


Fig. 17: Continued

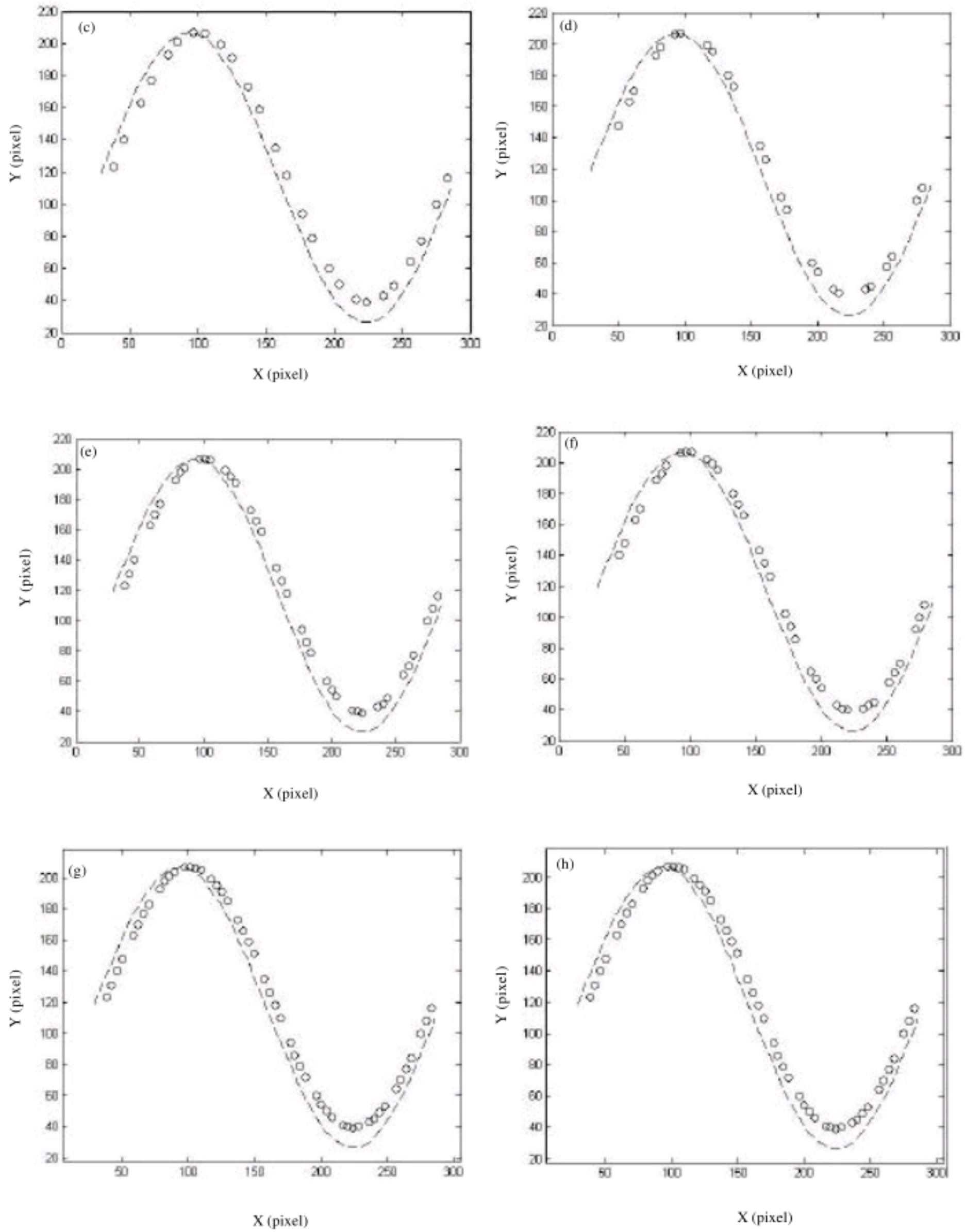


Fig. 17: Continued

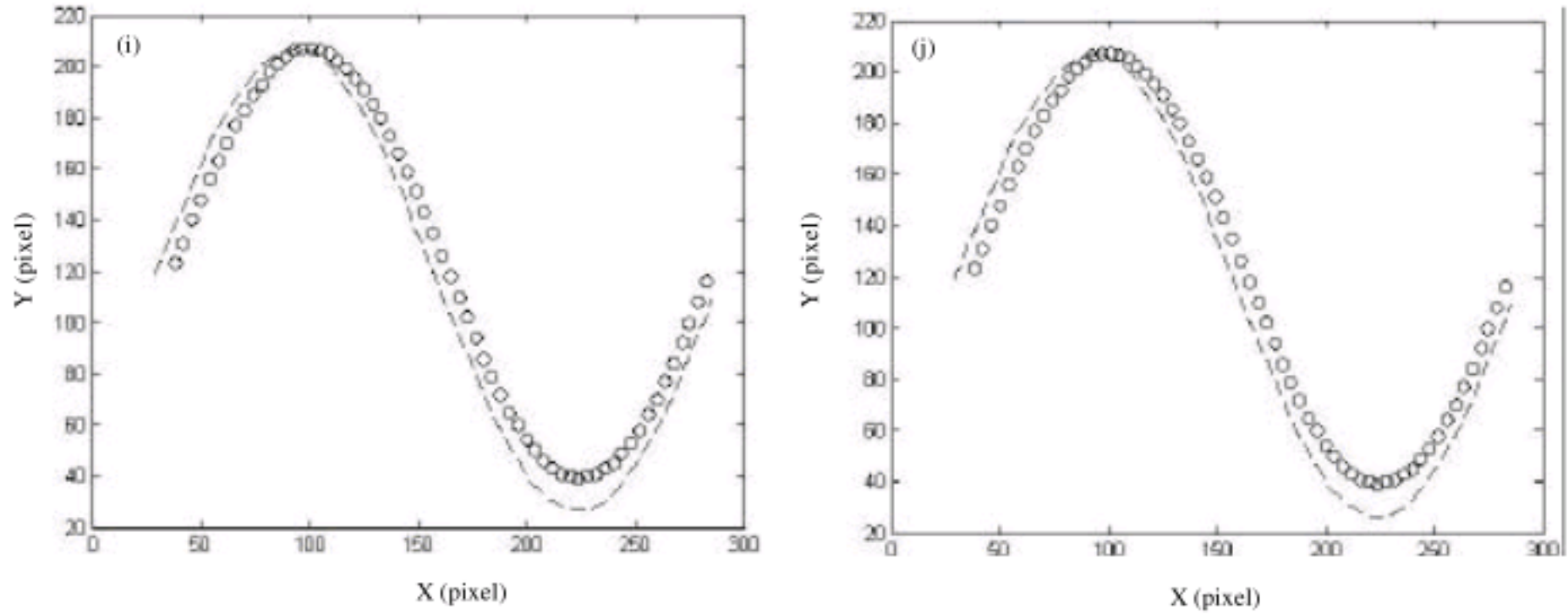


Fig. 17: The experimental results, when sampling density = 0, (a) result 1(5.41%); (b) result 2(5.37%), when sampling density = 2; (c) result 1(4.97%); (d) result 2(4.77%) and when sampling density = 4 (e) results 1(4.83%) (f) result 2(4.71%), when sampling density = 6; (g) result 1(4.76%); (h) result 2(4.95%) and when sampling Density = 8; (i) result 1(4.56%); (j) result 2(4.58%), The deisive least points = 8

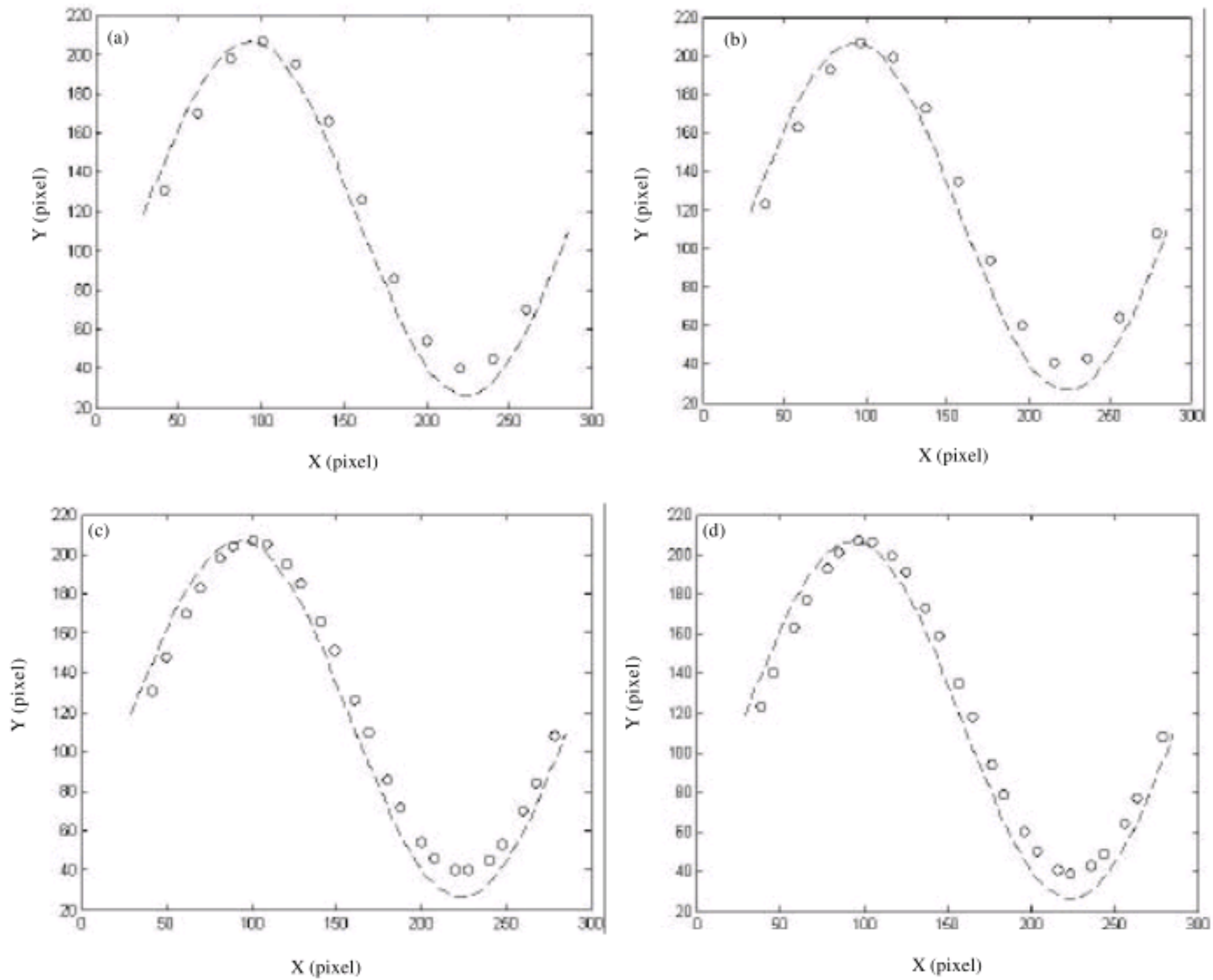


Fig. 18: Continued

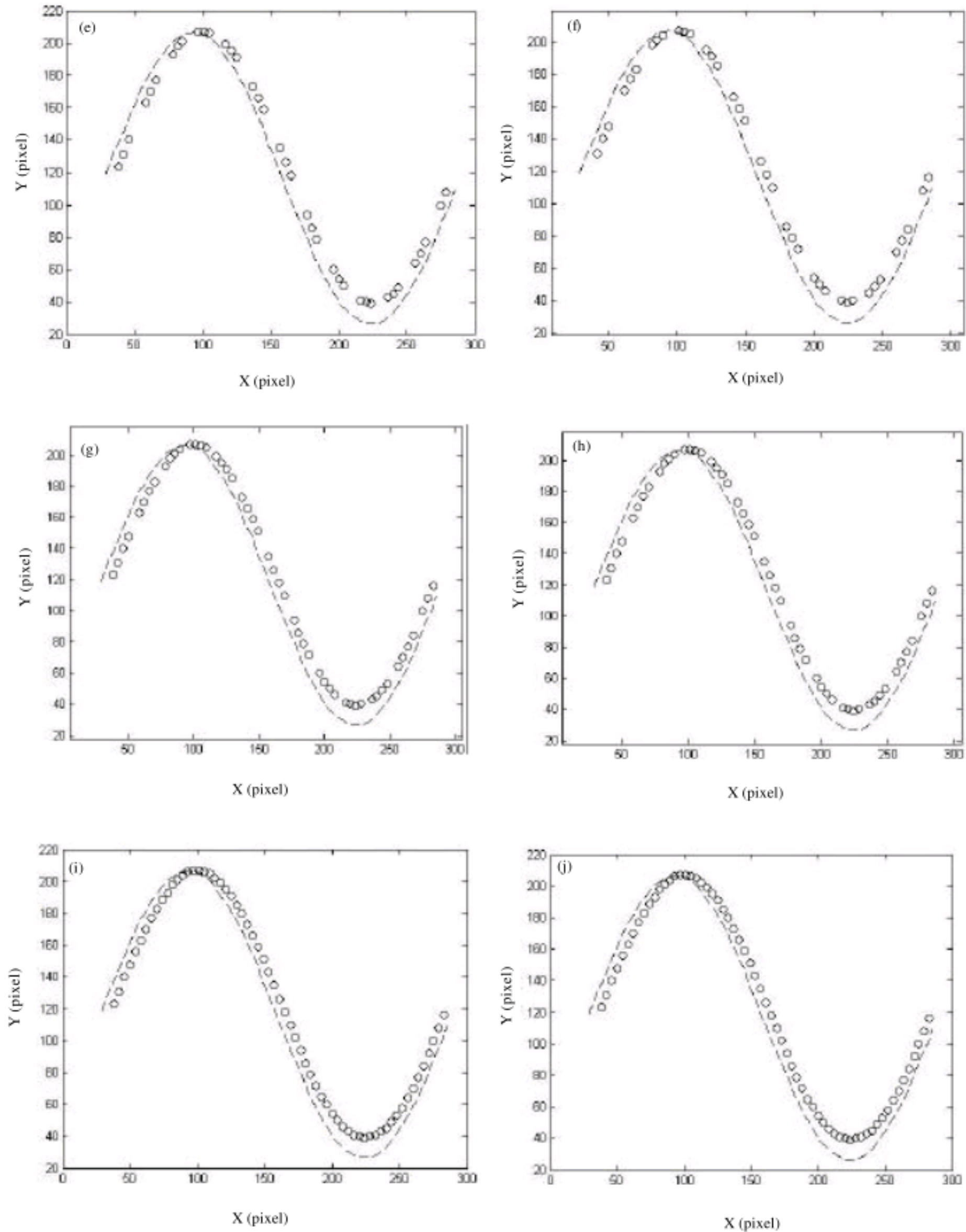


Fig. 18: The experimental results, when sampling density = 0, (a) result 1(5.01%); (b) result 2(4.87%), when sampling density = 2; (c) result 1(4.62%); (d) result 2(4.65%) and when sampling density = 4 (e) results 1(4.47%) (f) result 2(4.77%), when sampling density = 6; (g) result 1(4.62%); (h) result 2(4.76%) and when sampling Density = 8; (i) result 1(4.50%); (j) result 2(4.17%), The deisive least points = 8

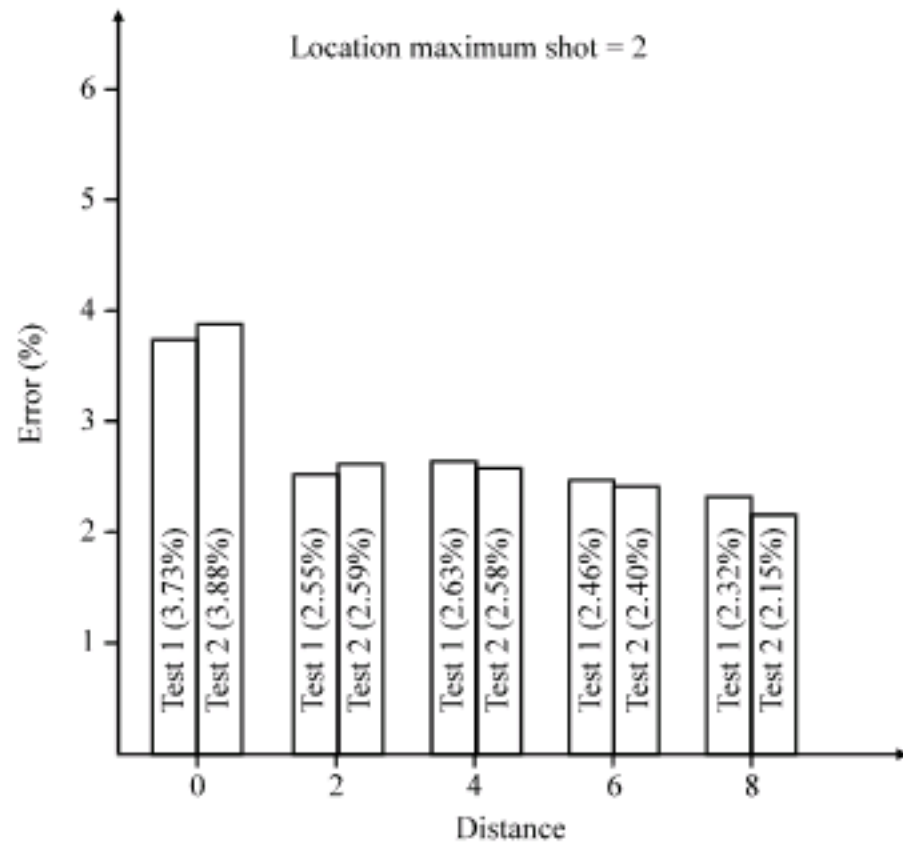


Fig. 19: Location maximum No. of aim is 2

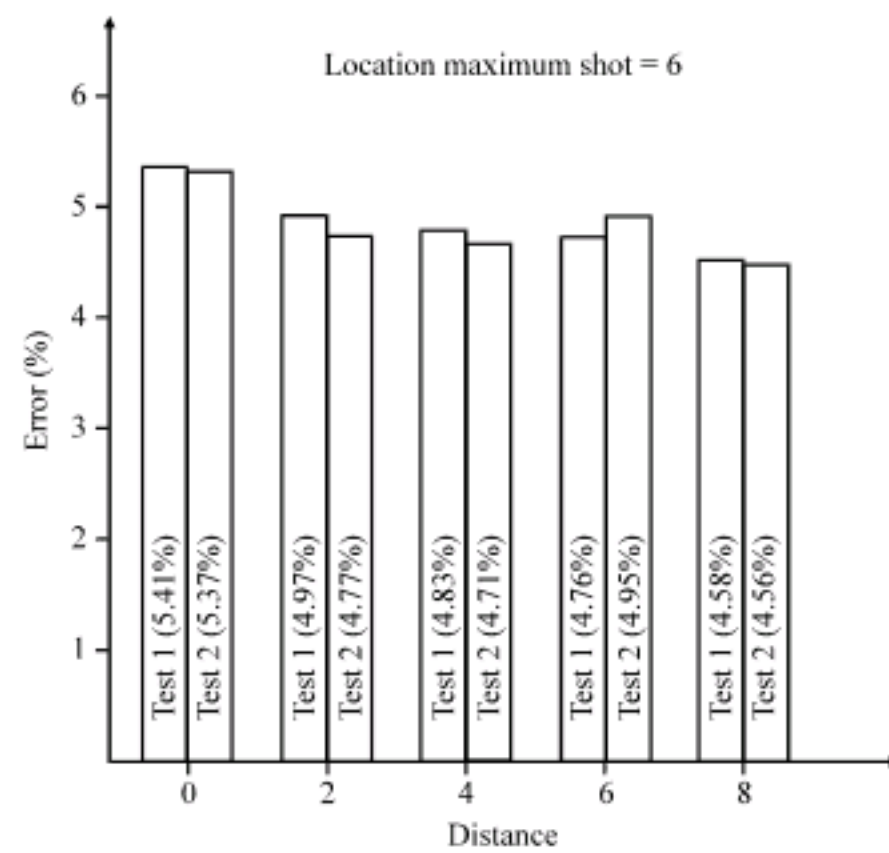


Fig. 21: Location maximum No. of aim is 6

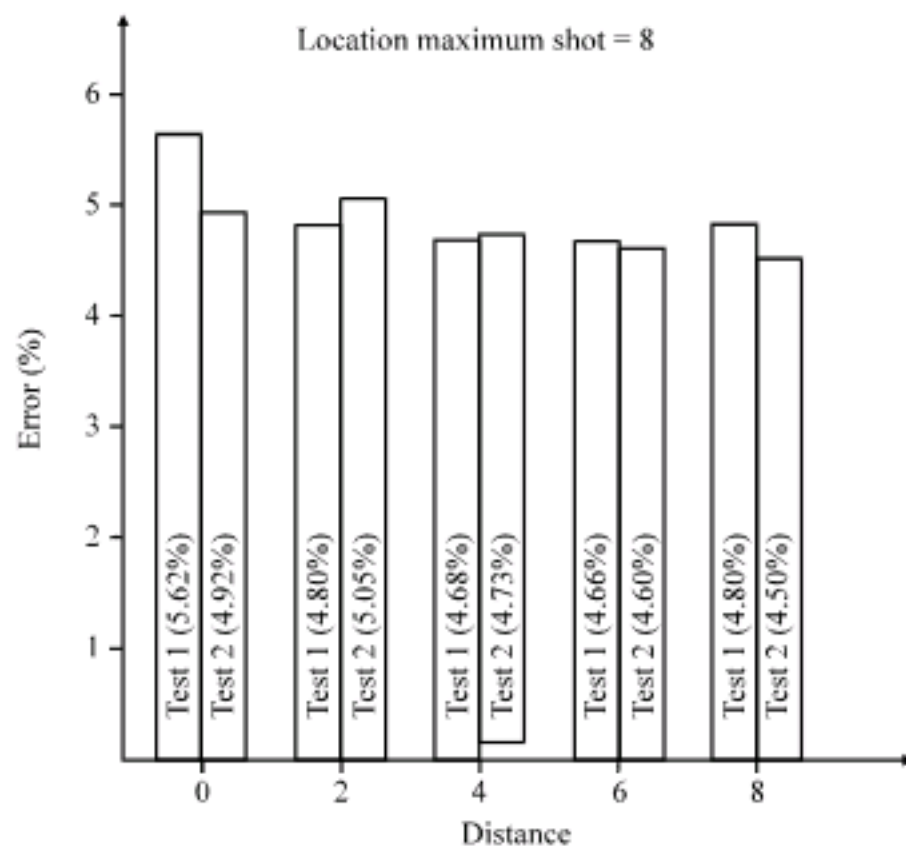


Fig. 20: Location maximum No. of aim is 4

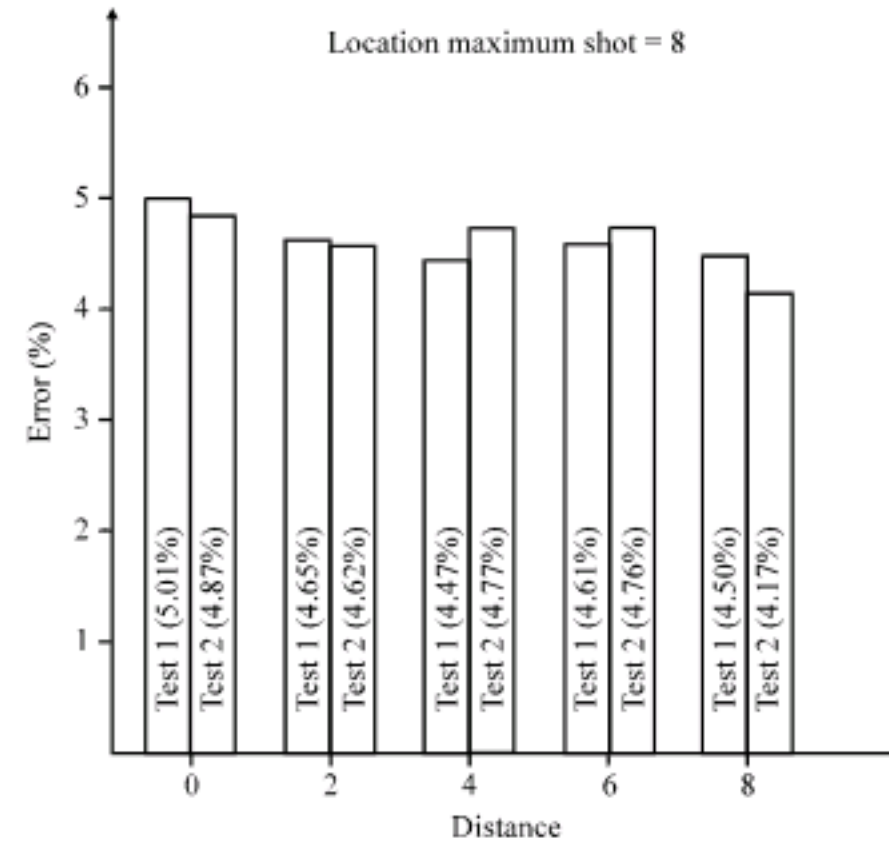


Fig. 22: Location maximum No. of aim is 8

respectively. In addition, the variables are the decisive least points and the sampling density.

The rules in choosing the sampling density when doing the goal's image analysis are different. The fundamental difference is that the digitization of the objects in an image into a collection of pixels introduces a form of spatial quantization noise that is not band-limited. It leads to the following results for the selection of the sampling density when one is interested in the measurement of area and (perimeter) length.

If one is interested in the image processing, one needs a sampling density based on classical signal theory -the Nyquist sampling theory. Similarly, choosing a sampling density based on the desired measurement accuracy (bias) and precision (CV) is also required. If the image analysis is requested. Moreover, in a case of

uncertainty, the higher of the two sampling densities (frequencies) will be necessary.

The final average data classified with the decisive least points are shown in Fig. 19-22. Besides the pixel of target, the sampling density, the decisive least points and the CPU speed play the essential role in the whole system. The final data reveals that both of the sampling density and the error are large. We can adjust the pixel of target by the bar of 0-8 in the software where the default target is 30x30.

THE INTEGRATION OF THE TRACKING SYSTEM AND THE VIRTUAL ENVIRONMENT

We designed an application of virtual reality to prove the characteristics of the image tracking system. The

operator can control the fire system on the warship and is in charge of shooting down the missiles. The benefit in the virtual reality environment is that the users will have the similar feeling without putting in the real environment.

Besides, the operator can control the device in time and will comprehensively realize the feasibility in military. In order to match up with the application of the tracking and the virtual reality, we redesign a virtual reality. Because there are plenty of 3D elements, we need a host with a better processing ability. Besides, moving parts of the application to another host to process independently is important. The hardware separates into two parts. First, as shown in Fig. 12, the PC1 has a tracking system. By inputting the image and analyzing the image data, the operator can use the mouse to circle the target and press the left key of the mouse to interact with virtual reality. The PC1 is used to imitate the LOCK and ATTACK of the cannonry shoot system.

On the other side, the specially used to work for the virtual environment, to show up the view and the contents through monitor and to show up the missile crash sound effect through the loudspeaker. When the conditions have been formed, the homologous results will be enforced.

In Fig. 23, there are two ways, including the left side computer (PC1) and the right side computer (PC2) that can connect with each other. First, the PC2 passes the missile position message used as image pathway to the PC1's image-input-equipment. Thereafter, the PC1 will

differentiate the missiles by the images. This is the single-way data communication. An alternative way is to make PC1 transform the attacking data to PC2 via the TCP/IP. The PC2 will transform the data. The status of the cannon and the missiles will be back to PC2. This is a two-way communication.

Virtual environment and programs: Virtual environment is built in a limited virtual place. The operator's view angle, who is in charge of shooting down the missiles, will be settled down on the top of the warship shown in Fig. 24. The missile will be produced randomly from the gray area of area A and the terminal point will be in the area B. The pathway of the missile is from the random star point to the random terminal point.

The simulation of the Virtual environment shown in Fig. 25 and 26, there's a visible alarm boundary, a limit to shoot down the missiles. The operator has to shoot down the missile before it cross the boundary. The application is designed for a vision tracking system, so that the operator can move the lock block in the PC1 in which the size of the lock block is 100x80. Basically, we can get the coordinates immediately if the object is within the lock block.

In order to increase the amusement and present the characteristic of the virtual reality, we set the coordinates of target. In order to make the games more fun, we also set that the target is truly in the attacking range when the

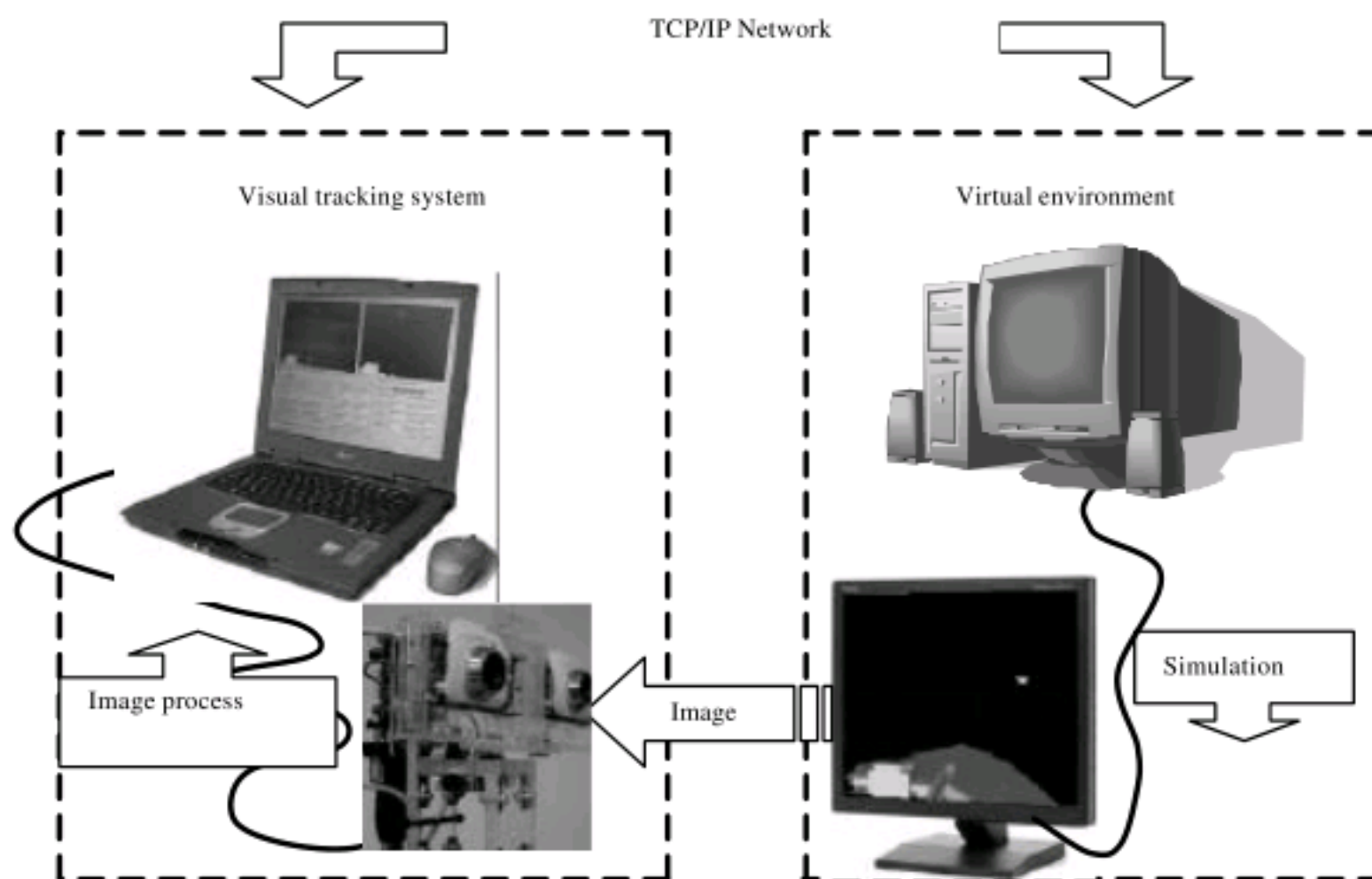


Fig. 23: The hardware of the tracking system and the virtual environment

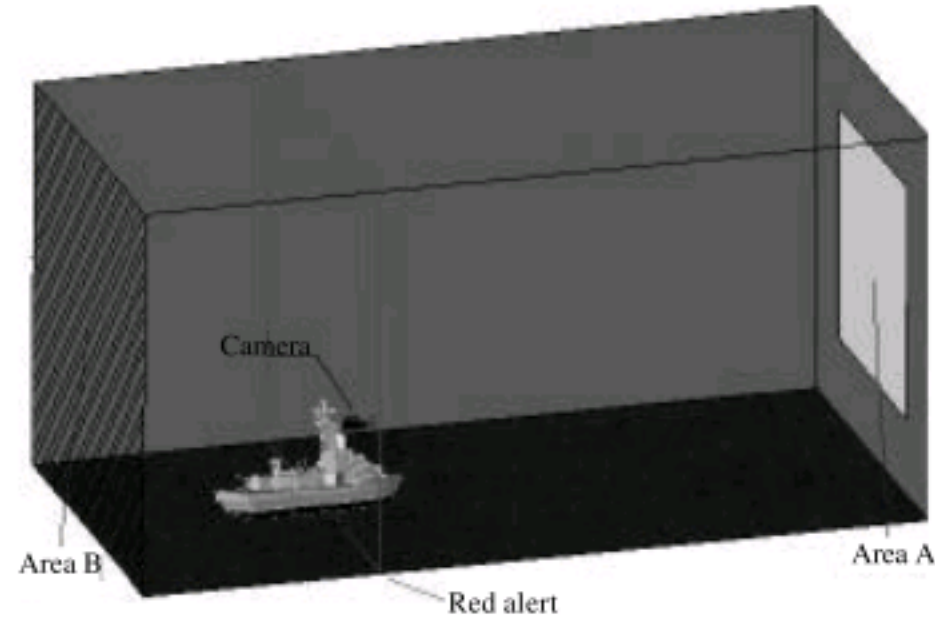


Fig. 24: The virtual environment



Fig. 25: The locked missile in the PC1



Fig. 26: The shot down missile in the PC1

system locks the target 10 times. Moreover, the aimed window will become red which means the final locking during attacking motion.

Before the target has been shot down, it will keep moving to the alarm area. The user will have no chance to shoot down the missile again. At the same time, the simulation virtual environment of the PC2 will make to be flash red alert.

CONCLUSION AND FUTURE OUTLOOK

The purpose in this study is to produce a high effective visual tracking system which is a complete and stable machine vision with low cost. The result reveals that the whole system is under cost free by using the webcams. The servo motor and the servo controller can be adopted. However, the system is unstable. The stability of the system will be improved expectedly if the light and the complex background can be eliminated.

REFERENCES

- Barnard, S.T. and W.B. Thompson, 1980. Disparity Analysis of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2: 330-340.
- Bertozzi, M. and A. Broggi, 1998. GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Trans. Image Process.*, 7: 62-81.
- Bjorkman, M. and D. Kragic, 2004. Combination of foveal and peripheral vision for object recognition and pose estimation. *Proceeding of the IEEE Int. Conf. Robotics Automation*, Apr. 26 -May 1, Centre for Autonomous Syst. and Comput. Vision and Active Perception Lab, R. Inst. of Technol., Stockholm, Sweden, pp: 5135-5140.
- Black, J. and T. Ellis, 2006. Multi camera image tracking. *Image Vision Comput.*, 24: 1256-1267.
- Henstock, P.V. and M.C. David, 1966. Automatic gradient threshold determination for edge detection. *IEEE Trans. Image Processing*, 5: 784-787.
- John, M., G. Ovidiu and F. Paul Whelan, 2003. Robust 3-D landmark tracking using trinocular vision. *Proc. SPIE*, 4877: 221-229.
- Kidono, K., J. Miura and Y. Shirai, 2002. Autonomous visual navigation of a mobile robot using a human-guided experience. *Robot. Autonomous Syst.*, 40: 121-130.
- Kim, S., I.S. Kweon and I. Kim, 2003. Robust model-based 3d object recognition by combining feature matching with tracking. *Proceedings of the IEEE International Conference on Robotics and Automation*, Sept. 14-19, IEEE Xplore London, pp: 2123-2128.
- Kuo, H.C. and L.J. Wu, 2002. An image tracking system for welded seams using fuzzy logic. *J. Mater. Process. Technol.*, 120: 169-185.
- Lobo, J., C. Queiroz and J. Dias, 2003. World feature detection and mapping using stereovision and inertial sensors. *Robot. Autonomous Syst.*, 44: 69-81.
- Mohan, R. and R. Nevatia, 1989. Using perceptual organization to extract 3D structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11: 1121-1139.
- Murray, D. and C. Jennings, 1997. Stereo vision based mapping and navigation for mobile robots. *Proceedings of the IEEE International Conference Robotics and Automation*, Apr. 20-25, IEEE, pp: 1694-1699.
- Ohya, A., A. Kosaka and A. Kak, 1998. Vision-based navigation by a mobile robot with obstacle avoidance using single-camera vision and ultrasonic sensing. *IEEE Trans. Robot. Automation*, 14: 969-978.
- Olson, C.F. and D.P. Huttenlocher, 1997. Automatic target recognition by matching oriented edge pixels. *IEEE Trans. Image Process.*, 6: 103-113.
- Seara, J.F. and G. Schmidt, 2004. Intelligent gaze control for vision-guided humanoid walking: Methodological aspects. *Robot. Autonomous Syst.*, 48: 231-248.
- Starck, J.L., F. Murtagh, E.J. Candes and D.L. Donoho, 2003. Gray and color image contrast enhancement by the curve let transform. *IEEE Trans. Image Process.*, 12: 706-717.
- Winters, N. and J. Santos-Victor, 2002. Information sampling for vision-based robot navigation. *Robot. Autonomous Syst.*, 41: 145-159.
- Yau, W.Y. and H. Wang, 1999. Fast relative depth computation for an active stereo vision system. *Real Time Imag.*, 5: 189-202.
- Yue Shigang, F., R. Claire, S. Matthais Keil, C. Jorge and S. Richard, 2006. A bio-inspired visual collision detection mechanism for cars: Optimisation of a model of a locust neuron to a novel environment. *Neurocomputing*, 69: 1591-1598.