

SHAPE INTEGRATION FOR 3D RECONSTRUCTION OF ARCHAEOLOGICAL SITES

Maki NAGANO Kosuke SATO Kunihiro CHIHARA
Grad. Student Associate Professor Professor
Grad. School of Information Science, Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara, 630-0101
E-mail: {maki-n, sato, chihara}@is.aist-nara.ac.jp
JAPAN

Commission V, Working Group 5

KEY WORDS: shape integration, 3D-reconstruction, stereo camera, archaeology, sensor fusion, camera calibration

ABSTRACT

The authors proposed a 3D measuring method, which is suitable for archaeological sites. Archaeological sites are especially difficult to measure because of the characteristics of their shapes and colors. Thus, the authors have selected a measuring method, which uses a stereo camera and an instrument for carrying out point surveys.

The method measures the entire shape of an object by integrating two kinds of sensed data: (1) dense local points obtained by pixels in stereo images ("a partial surface" is made from a stereo pair.) and (2) scattered points on the object, which are surveyed precisely and manually. The authors focused on fitting the partial surfaces to the surveyed points, and adopted an adjusting process, which contains two feedback loops. The method is evaluated by applying to the measuring of an archaeological site of a tile kiln. As a result, a precise 3D image of the site is made, and the authors succeeded in visualization of its 3D image.

1 INTRODUCTION

Recently, many multimedia databases of cultural heritages have been developed using digital archiving techniques (Hong, 1992, IEEE, 1993, Ueshima, 1988). In those databases, 3D shape and surface color of the heritages have useful information for their analysis, classification and visualization. Recent researches on computer vision enabled us to measure the precise 3D shape of objects automatically or semi-automatically (Sakaguchi, 1991, Bergevin, 1995). However, every measuring algorithm puts its own restrictions on the measuring objects. Still now, complex-shaped outdoor objects, such as archaeological sites, are especially difficult to measure for the following reasons.

- The active ways of casting laser light is not appropriate, because
 - the surface of archaeological sites often have dark color.
 - the power of sunlight changes greatly in a short time.
- Marking a great deal and mounting heavy machines on the site should be avoided, because the surface is very fragile.

In this paper, the authors proposed a 3D measuring method, which is suitable for archaeological sites.

The method uses (1) stereo images from multiple view-points and (2) scattered points, which are surveyed precisely and manually. Each stereo image is used for reconstructing a local shape ("a **partial surface**") in detail. The "**surveyed points**" are used for integrating the local shapes. As the integration result, data is expressed as a set of colored points on 3D space. In other words, we obtain "a **3D image**" by the proposed method.

The result of stereo matching (disparity image) always contains inaccurate values. The authors focus on a fitting process, which contains two feedback loops. Through this process, the inaccurate partial surfaces are gradually fitted to the surveyed points.

At first, the authors mention an acquisition of two kinds of data in section 2. Secondly, section 3 describes the integration processes of the two kinds of data. An experiment of measurement is shown in section 4, and the experimental result is discussed in section 5.

2 DATA ACQUISITION ON THE FIELD

The proposed 3D-reconstruction method uses two kinds of data acquired on the field (Figure 1). (1) Multiple

stereo images are taken by the stereo camera, and (2) the surveying instrument measures a set of 3D coordinates scattered on the surface.

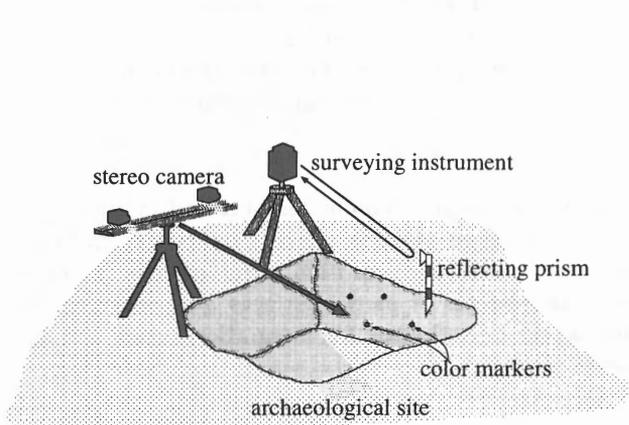


Figure 1 Two kinds of data are collected with dedicated instruments.

(1) A stereo camera (two cameras mounted on a pole in parallel), (2) an instrument for carrying out point surveys.

Voluntary viewpoints of the stereo images are allowed for convenience' sake. Therefore, camera parameters such as the positions and directions are unknown.

Archaeological investigation sometimes involves point survey using an instrument, which consists of range finding part and angle meters. 3D coordinates of distinctive features often have important meanings for archaeological analysis. In this paper, precisely surveyed points, which are randomly scattered on the surface, are used for adjusting integration parameters including the stereo camera parameters.

These integration parameters are very difficult to decide their optimal solutions. (Obtaining these solutions is one of the camera calibration(Hartlay, 1992).) Therefore, small numbers of color markers are used for rough estimation of the parameters (Figure 1). These reference points are surveyed precisely, and their colors are recorded with the 3D coordinates. It is not so difficult to extract the reference points from images automatically, because vivid colored markers are used, and archaeological sites generally have dark colors such as brown and black to the contrary. The color of a reference point in stereo images indicates the corresponding point in the set of surveyed points.

(The overlapped part of each stereo image has to contain more than three marker points.) Note that only a small number of markers are allowed because of the fragility of archaeological sites.

3 DATA INTEGRATION

This section describes the integration of the two kinds of data written in section 2. At first, the integration parameters are written in detail in subsection3.1. Secondly, the optimization process of the parameters is written precisely in subsection3.2.

3.1 Parameters for Shape Integration

3.1.1 Internal Parameters

The geometry of taking stereo images is arranged to simplify the stereo matching process(Figure 2). Two cameras are mounted on a pole in parallel, and each of those sight axes is perpendicular to the pole.

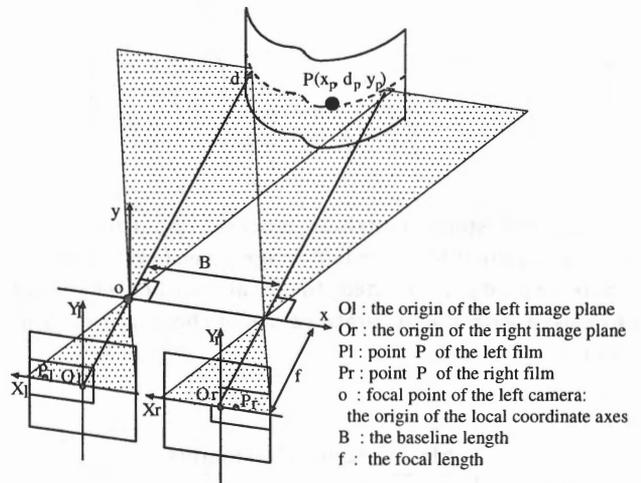


Figure 2 Geometry of stereo images.

Let $(X_l, Y_l), (X_r, Y_r)[m]$ be the left and the right coordinates on the focal plane, when O_l, O_r are their origins. The local coordinate $(x_p, d_p, y_p)[m]$ of a point P is calculated as follows:

$$\begin{pmatrix} x_p \\ d_p \\ y_p \end{pmatrix} = \frac{B}{X_l - X_r} \begin{pmatrix} X_l \\ f \\ Y_l \end{pmatrix} \quad (1)$$

Where B means the baseline length, and f means the focal length. The local coordinate is fixed on the stereo cameras, and the origin (o) is the focal point of left camera. The three axes(x, d, y) of the local coordinate fixed on the right image plane. Both of x, y axes are parallel to the image plane and d axis points to the direction of depth.

The authors restore depth information from stereo images using stereo matching (binocular stereo). A disparity image is made through the process of stereo matching. Slight rotation and slip occur when taking images, which would result in failures of searching in stereo matching, the authors revise the rotation or slipping using the correspondence of two markers in the both images.

For reducing the image processing time, left and right images are clipped out only overlapped part. This process brings four additional parameters, which mean the offset of clipping out two images from the original images. Where $(a_l, b_l), (a_r, b_r)$ [pixel] are the coordinates of the left-bottom of the cutout images on the image planes. New origin on the each of two images are the left-bottom point, and let $(u_l, v_l), (u_r, v_r)$ [pixel] be new coordinates of point P on the both images (left: P_l , right: P_r), let r [m/pixel] be the pixel sampling pitch, the local coordinates are formulated as follows:

$$\begin{pmatrix} x_p \\ d_p \\ y_p \end{pmatrix} = \frac{B}{(u_l - u_r) + (a_l - a_r)} \begin{pmatrix} u_l + a_l \\ f/r \\ v_l + b_l \end{pmatrix} \quad (2)$$

Through the stereo matching process, disparities $(u_l - u_r)$ are obtained for each left image pixels(P_l). If an accurate disparity is obtained, the local coordinates should be derived with the accurate values of the parameters in Table 1.

Table 1 Internal Parameters

parameters	meanings
a_l, b_l	offset of cutting out the images[pixel]
$a_l - a_r$	offset of the disparity[pixel]
f/r	the focal length[m] per pixel sampling pitch[m/pixel]
B	baseline length[m]

However, the stereo matching process is not so reliable, and the obtained disparity image contains a lot of wrong values.

Then we apply the threshold of the correlation coefficient, and label the pixel when none of the rectangle's coefficient is over the threshold, although there still exists much noise in the disparity image.

Note that the parameters may not be accurate at this point of time, the integration process(see section 3.2), which is the essential issue of the paper, effectively corrects them gradually using the noisy disparity image.

3.1.2 External Parameters

Table 2 External Parameters

parameters	meanings
θ, ϕ, ψ	rotation angles[radian]
δ, ζ, η	translation distances[m]

The partial surface coordinates obtained in subsection 3.1.1 is fixed on cameras, so it is difficult to integrate the other partial surfaces from many viewpoints. These local coordinates of partial surfaces are converted into the **world coordinates** for integration, using the surveyed points.

Let θ, ϕ, ψ be each of the rotation angles around the axes of x, d, y , and δ, ζ, η be the translation distance in the directions of each axis (see Equation (1)), the linear coordinate transformation is expressed as following.

Put

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{bmatrix} \begin{bmatrix} \cos \psi & -\sin \psi & 0 \\ \sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

then

$$\begin{pmatrix} x' \\ d' \\ y' \end{pmatrix} = A \begin{pmatrix} x \\ d \\ y \end{pmatrix} + \begin{pmatrix} \delta \\ \zeta \\ \eta \end{pmatrix} \quad (4)$$

where (x', d', y') means the world coordinates. The initial values of these six parameters are roughly estimated using the correspondence of three marker points, although strains or distortions on data acquisition may affect on the accuracy of the result. The other surveyed points help to optimize the values of 11 parameters including these six parameters. Table 2 shows the six external parameters, which are not calibrated accurately at this point, and they should be optimized later.

3.2 Integration Process

3.2.1 Process Outline

The two kinds of data are integrated through a double loop procedure. Figure 3 is the conceptual figure of the integration process.

One of the disparity images converges onto the surveyed points as follows:

1. Convert a disparity image into the local coordinates using roughly estimated internal parameters (subsection 3.1.1).
2. Calculate the rough values of external parameters using the correspondences between the world coordinates of surveyed points and local coordinates of a partial surface (subsection 3.1.2).
3. Convert the local coordinates into the world coordinates. Then examine a certain threshold value, and quit the process if the coordinates are fitted onto the surveyed points.
4. Find a new set of one-to-one correspondences between a pixel in the partial surface and a surveyed point (subsection 3.2.2): select the nearest points when converting with parameters at the point of time
5. Correct (subsection 3.2.3) 11 parameters using the correspondence as stated above, and back to the process 3.

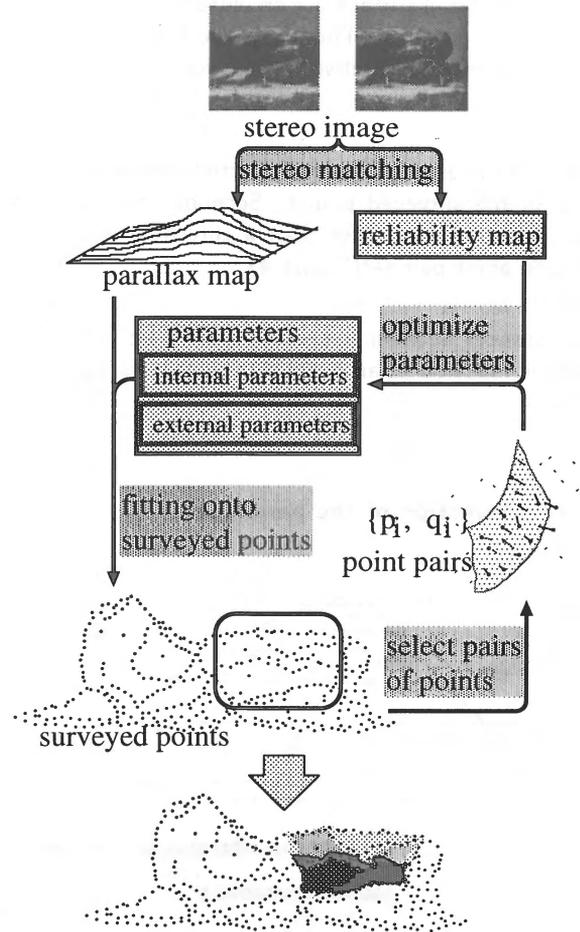


Figure 3 Flow of the surface integration.

3.2.2 Selection of One to One Correspondences of Two Data

The point selection process (process 4) is described in subsection 3.2.2, and the parameter optimization process (process 5) is described in subsection 3.2.3.

After the above procedure, all of the local coordinates from different viewpoints are integrated into a 3D image. A pixel of a 3D image has six parameters (3D coordinates and RGB values).

According to the instability of stereo correspondence search, the disparity image contains a lot of noises, which still remain in the 3D image as floating points in the air. Those floating points are removed by 3D region growing.

3D region growing:

A set of proper points are defined. At first, one pixel in the 3D image is the element of the set. The points within a threshold distance are taken into the proper set one by one until no more point will join the set.

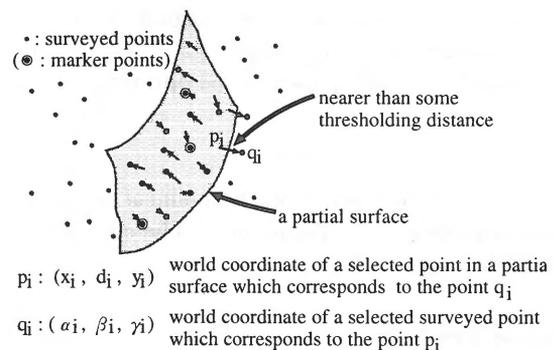


Figure 4 The selection of correspondences. In this process, the nearest point in a partial surface is selected for each surveyed point.

To derive more precise parameters, or to optimize the parameters, we should select which one of the point in the partial surface is to agree with which one of the surveyed points (Besl, 1992). The surveyed points are sparse, but

scattered on the entire surface. On the contrary, the points in a partial surface are crowded, but exist only on a part of the surface. Therefore, the first process is to find correspondences between them for the optimization process.

At first, the nearest point in the partial surface is selected for all of the surveyed points. Secondly, the point pair whose distance is farther than a threshold is excluded from the point pair set(Figure 4).

The distances between these point pairs are used as the objective function of an optimization process(subsection 3.2.3).

3.2.3 Correction of the parameters

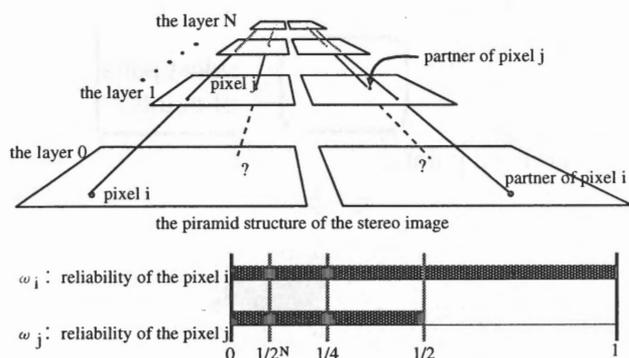


Figure 5 The pyramid structure of image resolution determines the reliability of the disparity. This is the resolution pyramid of stereo images: step 0 to N, where step 0 means the original image, which is most precise, and step n+1 is the half size of step n. If a correlation coefficient exceeds the threshold in some region of step n and doesn't exceed in every region of step n-1, the reliability of the pixel is $1/2^n$. The obtained reliability image is used in the optimization process.

After the point pair selection, a nonlinear optimization process is applied to the parameters. The objective function is following.

objective function:

$$\frac{1}{N} \sum_i \frac{1}{\omega_i} |(\alpha_i - x'_i, \beta_i - d'_i, \gamma_i - y'_i)| \rightarrow \text{minimum} \quad (5)$$

Where, N is the number of selected pairs of points(see subsection 3.2.2), $(\alpha_i, \beta_i, \gamma_i)$ means the coordinates of a point q_i in the surveyed points, (x'_i, d'_i, y'_i) means the coordinates of a point p_i in the partial surface,

and the parameter ω_i means the "reliability" of the disparity $((u_l - u_r)_i)$ of the partial surface. Its necessity and meaning is as follows.

Necessity of reliability ω_i :

The stereo matching process is based on the correlation coefficients of pixel intensity in small regions for finding the match. Due to occlusion problem, the proper target may not exist in the other image. Therefore, if the coefficient peak is smaller than a threshold, the point in the disparity image is excluded from the integration process(see subsection 3.1.1). However, the coefficient values could be lower because of the object's surface is rough microscopically while it looks smooth surface macroscopically. To cope with this phenomenon, we applied image pyramid method on stereo matching, where if we cannot find a match in the lower step(fine-resolution), we adopt the result of the higher step(coarse-resolution), although it may not be so accurate.

Figure 5 shows the meaning of reliability, which is associated with the image pyramid method used in the stereo matching.

The initial values of the parameters are obtained using the two data correspondence of marked points. Therefore, it can be assumed that the optimal solutions are near enough to the initial values, and that a local optimal solution is equal to the global optimal solution.

If the integration result is accurate enough, the average distance of the nearest M (a constant) point pairs out of all point pairs (N pairs) are shorter than a constant threshold. If the value of the objective function got worse than the former iteration of selection (subsection 3.2.2), it judges that the optimization of the parameters is finished. If not being judged as finished, it restarts from the process of the selection using new parameters.

4 EXPERIMENTAL RESULTS

Table 3 Initial values of 11 parameters

parameters	initial value
$a_l/r, b_l/r$	-460, -600 [pixel]
$(a_l - a_r)/r$	1000 [pixel]
f/r	3100 [pixel]
B	0.8 [m]
θ, ϕ, ψ	-0.75779, -0.47925, 0.45981 [radian]
δ, ζ, η	4.86775, 8.07292, -0.51093 [m]

In order to evaluate the proposed method, the authors measured an actual archaeological site in Oyama, Tochigi, Japan.

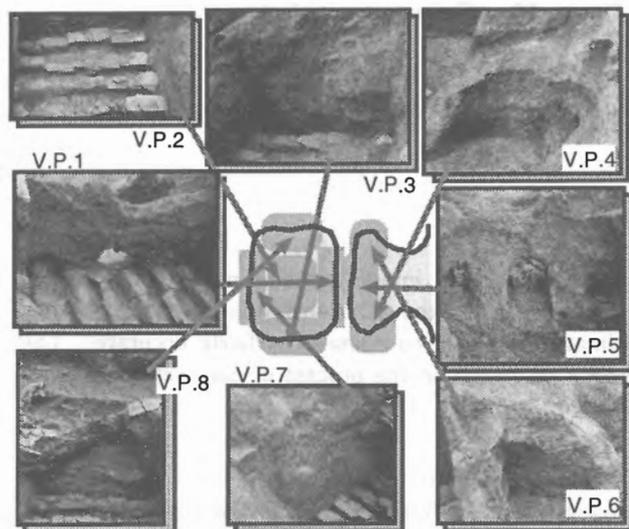
Table 4 The last values of parameters

parameters	result of the final optimization
$a_l/r, b_l/r$	-462.08, -611.33 [pixel]
$(a_l - a_r)/r$	1082.4 [pixel]
f/r	3118.7 [pixel]
B	0.81452 [m]
θ, ϕ, ψ	-0.78924, -0.47390, 0.51704 [radian]
δ, ζ, η	4.91814, 7.96012, -0.25621 [m]

Figure 6(a) shows the whole view of the site, and (b) shows the left images of a stereo. The site size of the inside (the left square part) is about 2 meters square, and the depth is about 1.3 meters.



(a) The whole site



(b) Left stereo images from 8 viewpoints

Figure 6 Measured object: the site of a tile kiln. "Oyama-fudoubara kawara kamaato": an oven of ceramics. Oyama, Tochigi, Japan, the 7th century.

All of the images were taken from upwards obliquely. The left square of (a) is the inside of the kiln, and from the holes of right part, people blew on the fire.

On the field, a 35mm still camera with a less distorted

28mm(F2.8) lens took stereo photographs. The conditions of the experiment are as follows.

- Overlapping percentage of a scene in the films was more than 50%
- The resolution of the film reader was about 2820 dot per inch .
- Baseline length was about 80cm.
- Distance from camera to objects was within the range of 1.24~2.24m.

In the process of stereo matching, the window size of calculating correlation coefficients was 15×15 [pixel].

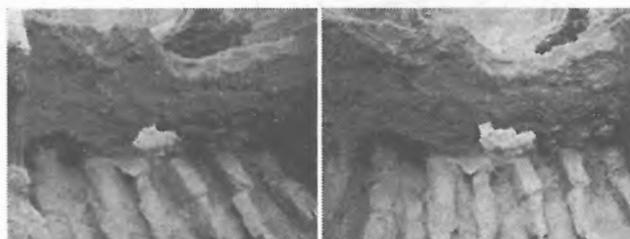
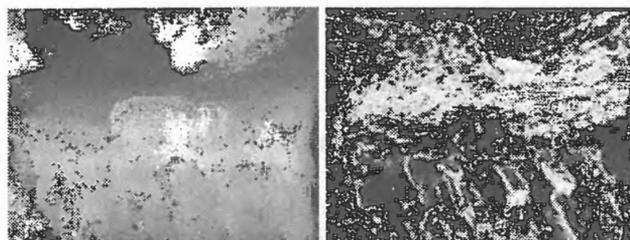


Figure 7 A stereo image

Figure 7 shows one of the stereo pairs, which were taken from viewpoint 1. Note that the overlapping parts were clipped out from the original images. Both of the image size were 2300×1750 [pixel]. We used three markers out of nine in overlapped part of the images for the calculation of initial values of external parameters. Table 3 shows the obtained initial values.



(a) A disparity image (b) A reliability image

Figure 8 The result of stereo matching. The disparity image shows depth information of the left image of Figure 7. Brighter grays indicate farther points from cameras. In the reliability image, lighter grays mean higher resolution of disparity.

Figure 8(a) is the disparity image obtained from the stereo image (Figure 7), where the black areas indicate the pixels for which the stereo matching was failed, elsewhere, brighter grays indicate farther part from cameras. The image size is 460×350 [pixel], which is smaller than the original stereo images. Figure 8(b) is the reliability image, which indicates the resolution of the disparity. Brighter areas mean that reliability is high.

1386 points were surveyed precisely using an optimal point survey instrument. These points scatter on the surface of the site randomly, and they are shown from three viewpoints in Figure 9. The point set includes 130 marker points in nine colors.

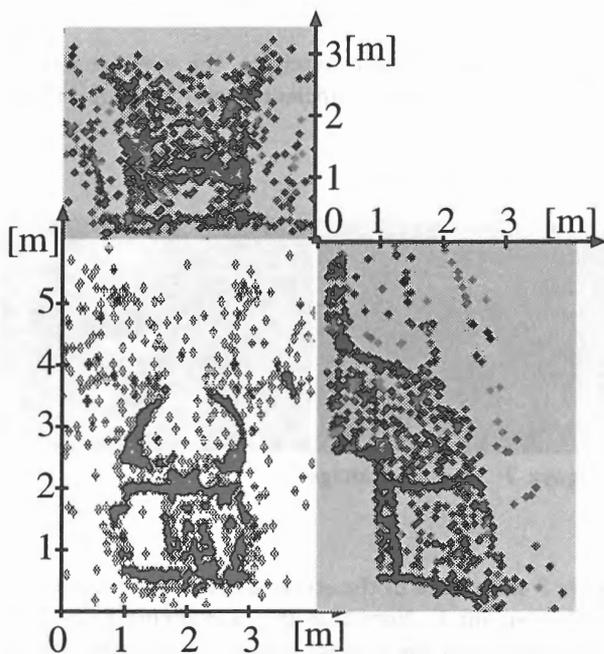


Figure 9 Surveyed points from three points of view. White region of left-bottom is the points we look down from upper part, where the bottom square is inside of the kiln. The other two figures are the horizontal appearances of the surveyed points, where we can confirm that the level of outside is lower than the inside of the kiln.

We took 0.05 [m] for the thresholding value of distance in the process of selection (see subsection 3.2.2). The iterative selection process was done four times, optimizing process was done in each of iteration, then the final optimal values of parameters were obtained as shown in Table 4. The local coordinates are translated to the world coordinates using these parameters.

Similar processes were applied to the other partial surfaces, and whole shape of the site was reconstructed as a 3D image. There existed a lot of and isolating points in the air at the point. Therefore, the 3D region growing

(see subsection 3.2.1) was applied to the 3D image using a threshold distance of 0.05 [m]. Within the $653,923$ points of 3D image, $569,226$ pixels remained after the region growing process.

Figure 10 shows the reconstructed 3D image from six viewpoints. The 3D image is colored using the pixel values of the original left images.

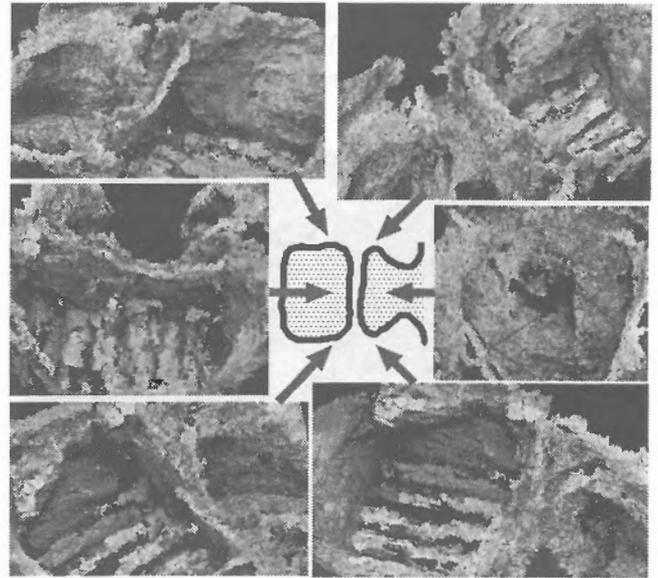


Figure 10 3D-reconstructed site

5 Discussion on optimization process

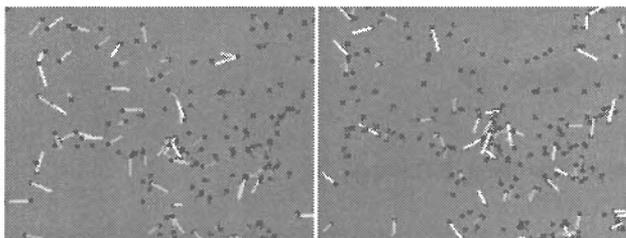
Although disparity image contains many wrong values from false target, the 3D transformation from local coordinates to world coordinates is fairly accurate. This section discusses on the process of parameter optimization.

At first, (1) - (3) of Figure 11 visualize the one to one correspondences after each iteration of selection process. Black dots mean the selected pixels within the scene of Figure 7. The line from each dot is a 3D vector of the same length being projected onto the image plane, and the line color means the distance between the two corresponding points: brighter gray vector indicates farther point (white is exactly 0.05 m.).

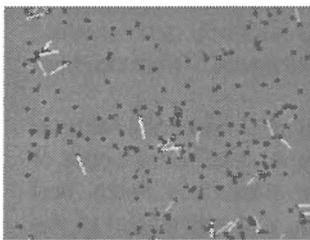
Figure 11(1) agrees that the parameters are not appropriate, because the correspondence vectors have a biased distribution, and there are many brighter gray vectors. After optimization process, some improvements in the distribution are confirmed. Figure 11(3) seems that the

errors are distributed rather equally into every vector. The vector colors are darker on the average compared with (1): distances of two points are shortened as a whole.

Figure 12 shows the trend of the optimization process. The horizontal axes of both graphs are the iteration of the selection. The vertical axis of graph (a) is the number of selected pairs of points. The vertical axis of graph(b) is the average distance of nearest 100 pairs of the points.



(1) The first selection (2) The second selection



(3) The last (fourth) selection

Figure 11 The selected points in a disparity image: Each one has a correspondence in the surveyed points. Black points are the selected points, and lines from those points indicates the unit 3D vector projected. The brighter gray line means longer distance between the two points.

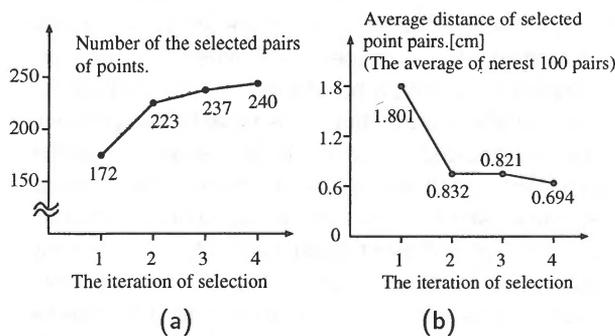


Figure 12 Progress in the optimization process. The parameters are exact if there are many selected points, and if the average distance of every pairs of world coordinates is short.

6 CONCLUSIONS

The authors have proposed a 3D-reconstruction method, which can be applied to outdoor object of complex shape, combining detailed stereo matching and optical point survey.

The integration parameters need not to be precisely decided manually, but they need to be roughly estimated. The marker points help with making sure of the epipolar restriction and deciding the initial values of the integration parameters semi-automatically. The errors in a rough estimation are corrected gradually according to the optimization process using a cloud of surveyed points, which are scattered randomly on the surface of the object.

This kind of easy photo-surveying using computer vision technology might assist archaeological researches.

Acknowledgment

The archaeological site was provided by Oyama-city Board of Education. The experiments are sponsored by Historic-Sites Repairing Lab., Tsukahara Ryokuchi Institute. The authors wish to thank Prof. N. Katata of Tezukayama University, and Mr. M. Okayasu for helpful discussion in the field of archaeology.

References

- IEEE Computer Society, 1993. Special Issue on Scientific Databases Bulletin of the Technical Committee on Data Engineering, 16(1).
- J. K. Hong, 1992. A motion picture archiving technique, and its application in an ethnology museum. DEXA92
- P. J. Besl and N. D. McKay, 1992. A method for registration of 3-d shapes. IEEE Transaction of Pattern Analysis and Machine Intelligence, 14(2), pp.239-256.
- R. Bergevin, D. Laurendeau, and D. Pousart, 1995. Registering range views of multipart objects. Computer Vision and Image Understanding, 61(1), pp.1-16.
- R. I. Hartley, 1992. Estimation of relative camera positions for uncalibrated cameras. Proc. ICCV, pp 579-587.
- S. Ueshima, K. Ohtsuki, J. Morishita, Q. Qian, H. Oiso, and K. Tanaka, 1995. Incremental data organization for ancient document databases. Proc. of the DASFAA95, pp.457-466.
- Y. Sakaguchi, H. Kato, K. Sato, S. Inokuchi, 1991. Acquisition of entire surface data based on fusion of range data. The Trans. of the IEICE, E74(10), pp.3417-3422.