〔非公開〕

TR-C-0148

熱画像と可視光画像の
ステレオ統合による
時系列距離画像獲得法の検討

大谷　淳　　　岸野　文郎
Jun OHYA　　　Fumio KISHINO

１９９６　３．１５

ＡＴＲ通信システム研究所

熱画像と可視光画像のステレオ統合による時系列距離画像獲得法の検討

(A Study of Time-sequential Range Image Acquisition from Stereo Pairs of Thermal and Intensity Images)

大谷 淳　　　岸野 文郎

Jun OHYA and Fumio KISHINO

ATR Communication Systems Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-02, Japan

Abstract

We propose a new method for acquiring time-sequential range images that provides dense range data in a viewpoint-independent way using only passive sensors. In our approach, stereo pairs of thermal and intensity images are synchronously acquired and are mutually registered. Since the points to be matched in the stereo thermal images are limited to those with the same temperature values, the thermal images are segmented into isotemperature regions whose contours are matched. Our method also exploits temporal correspondences between contours at successive time instants. Contour-based matching is done for the isotemperature regions independently at each time instant, and by temporal correspondence, possible matching pairs of contours are generated. By evaluating the similarities of the pairs, consistent and likely pairs are chosen. To supplement the sparse range data obtained from the contour matching process, information from the stereo intensity images is used. Intensity profiles along horizontal segments are matched by dynamic programming, where a horizontal segment is the intersection of an epipolar line with an isotemperature region. From corresponding pixel pairs obtained from the matching processes, the 3D coordinates of the points can be calculated. Experiments on real scenes, including a sequence showing moving human being, show promising results.

1 Introduction

In recent years, the problem of estimating the motion of objects in 3D scenes has received increasing attention in the computer vision community. In particular, non-rigid objects such as human beings are important targets for motion detection systems[1]. Just as a single range image can be used to analyze the 3D structure of a scene, time-sequential range images can greatly facilitate the detection of the 3D motion of non-rigid objects.

Existing methods for time-sequential range image acquisition involve fast laser range scanning[2] and stereo or multiple intensity image matching[3]. A laser range scanner that can acquire range images at video rate has been developed[2]. Range measurement by this method is generally accurate, and dense range data can be obtained. However, a shortage of laser reflection from dark colored objects results in large error or no data. Stereo intensity images can be used to get range data from dynamic scenes[3]. However, research using stereo images has mainly focused on motion parameter estimation, not on acquiring time-sequential range images. This is because correspondence detection in stereo images is still a difficult problem, and only sparse range data can be obtained from matching edge pixels. For static scenes, edge-based stereo algorithms[4], multiple-baseline stereo[5], and shape from motion[6] have all been studied, but these methods have not yet been applied to dynamic scenes.

To solve the problems of the existing methods, we proposed a passive sensor fusion approach that uses both stereo infrared video cameras and stereo video cameras for visible light[7]. In our approach, the thermal images facilitate stereo matching, and dense range data are acquired from the intensity images. Although thermal images have been previously used to discriminate objects in outdoor scenes[8], and thermal and intensity images have been combined to analyze outdoor scenes[9], our work is unique in that we acquire time-sequential range images. However, until recently, we were not able to achieve sufficiently dense range data acquisition[7].

In our method, stereo pairs of thermal and intensity images are synchronously acquired and are mutually registered. In the stereo thermal images, corresponding points must have the same temperature values. We segment these images into isotemperature regions. In the original method, we relied on contour-based matching of isotemperature regions at each time instant. In this paper, we describe how to match the contours of the isotemperature regions using temporal correspondence at successive time instants. To supplement the sparse range data obtained by contour matching, the intensity profiles, from the corresponding stereo intensity images, of pixels inside isotemperature regions are matched via dynamic programming.

Section 2 describes the principles of range estimation. Section 3 gives the details of our algorithm. Section 4 shows experimental results. Section 5 concludes the paper.

2   Range estimation by the integration of stereo pairs of thermal and intensity images

Objects emit electromagnetic radiation. The radiation at temperatures below $1,000$(K) peaks at wavelengths near 10 microns and is negligible at wavelengths of visible light(400 to 700nm). This means that the temperature of objects can be determined by measuring the amount of radiation at wavelengths of infrared light. The relationship between radiation W and temperature T(Kelvin) is given by

$$W = \epsilon \sigma T^4,$$ (1)

where ϵ is the emissivity and σ is the Stefan-Boltzman constant. For blackbodies $\epsilon = 1$; other objects have $\epsilon < 1$. Infrared video cameras[10] can measure radiation from the viewing direction of each pixel and, if the emissivity ϵ of an object is known, can convert the measured radiation to temperature according to Eq.(1).

Consider the case in which an object with a temperature distribution on its surface is observed by two infrared video cameras(Fig.1). Let the temperature of point A on the surface of the object be T_A, and suppose that point A is observed as pixel A_L in the left thermal image and pixel A_R in the right thermal image. The radiation W_A of A can be obtained by substituting the temperature T_A of A and the emissivity ϵ_A into Eq.(1). For ordinary objects, W_A is isotropic. Let W_{A_L} and W_{A_R} be the radiations observed at A_L and A_R. Then, the difference between W_{A_L} and W_{A_R} could be caused by differences in propagation characteristics between A and A_L and between A and A_R because of the differences in the type and length of the medium traversed. However, since gases emit much less radiation than solids, the propagation based differences are negligible (unless infrared cameras are observing distant places as in remote sensing systems). Therefore, the following relationship holds.

$$W_{A_L} = W_{A_R}$$ (2)

Now, if T_{A_L} and T_{A_R} are the observed temperatures at A_L and A_R, then

$$T_{A_L} = T_{A_R}$$ (3)

In stereo intensity images, on the other hand, intensity values in the left and right images generally differ from each other even if the pixel pair in the stereo image corresponds to the same point in a scene. Our method avoids this difficulty by utilizing the above mentioned property; that is, pixels to be matched in stereo thermal images are restricted to pixels with same temperature values.

To estimate range information, corresponding positions need to be found in the stereo thermal images. Thermal images can be segmented into isotemperature regions, which

3

consist of connected pixels having same temperature values. We have investigated region-based matching in stereo thermal images, but this does not always give satisfactory results because corresponding regions, due to occlusion, do not always have similar shapes in the left and right thermal images[11]. We therefore match the contours of isotemperature regions using both static and temporal correspondence, as we will describe later.

By matching contours, only sparse range data can be acquired. To get denser data, we also use stereo intensity images acquired by two visible light TV cameras. As shown in Fig.1, visible light and radiation from scenes including moving objects are separated by a mirror so that visible light is acquired by the TV cameras and radiation by the infrared video cameras. The infrared cameras and TV cameras are placed so that stereo pairs of thermal and intensity images are mutually registered and synchronously input to our algorithm. Pixels in the stereo intensity images that correspond to points inside isotemperature regions are matched via dynamic programming.

Our method achieves robust stereo matching, because stereo thermal images do not suffer from the differences due to viewpoint change that cause difficulty in matching stereo intensity images. Dense range data can be obtained by our method, because we combine stereo intensity images with stereo thermal images. In these respects, our method is appropriate for dynamic scenes including moving objects.

3 Algorithm

3.1 Outline

In our previous work[7], we obtained a range image at each time instant independent-ly. However, if temporal correspondence between time instants is used, more stable and robust stereo matching can be realized. In this paper, images from two successive time instants are utilized to improve stereo matching. In addition, for dense range data acquisition, we use information from stereo intensity images.

Figure 2 shows an overview of the algorithm proposed in this paper. As described in Section 2, isotemperature lines (contours of isotemperature regions) are matched first. Then, the results of the matching process are exploited for matching at the next time instant ((a) in Fig.2). For the initial time instant ($t = 0$ in Fig.2), which does not have a preceding time instant, isotemperature line matching is performed using the stereo pairs at $t = 0$ and subsequent time instant $t = 1$ (Mode 1). At other time instants ($t = i, i \geq 1$), we use the isotemperature matching results for $t = i - 1$ and the stereo pairs at $t = i$ (Mode 2).

At each time instant, isotemperature line matching, by either Mode 1 or Mode 2, can supply only sparse range data. To get denser range data, we use the intensity information from the stereo intensity images ((b) in Fig.2).

4

From the matching processes described above, corresponding pixel pairs are obtained, and the 3D coordinates of the pixel pairs are calculated by triangulation of the stereo cameras ((c) in Fig.2).

We now describe in more detail each of the modules shown in Fig.2, beginning with the modules of Mode 1 in Section 3.2 and followed by the modules of Mode 2 in Section 3.3. Most of the Mode 2 modules are similar to those in Mode 1. Dense range data acquisition using stereo intensity images is explained in Section 3.4.

3.2 Isotemperature line matching: Mode 1

To improve the robustness of the initial isotemperature line matching, the stereo thermal images from time instants $t = 0$ and $t = 1$ are used (rather than just a single pair). A block diagram of Mode 1 is shown in Fig.3. Each thermal image at $t = 0$ and $t = 1$ is segmented into isotemperature regions ((a) in Fig.3). Then, initial matching of contours of isotemperature regions is carried out independently at $t = 0$ and $t = 1$ ((b) in Fig.3). Next, correspondences between contour segments at $t = 0$ and $t = 1$ are found ((c) in Fig.3), and the matching of contour segments at $t = 0$ is finalized ((d) in Fig.3).

3.2.1 Segmentation into isotemperature regions

As shown in Fig.3, each thermal image at $t = 0$ and $t = 1$ is segmented into isotemperature regions. Let M be the number of temperature levels of the original stereo thermal images. The thermal images are quantized to $N(< M)$ temperature levels and segmented into isotemperature regions that consist of connected (8-adjacent) pixels having the same quantized temperature values. The subsequent processes (b), (c) and (d) in Fig.3 are carried out separately for each quantized temperature level.

3.2.2 Initial contour matching

Pixels on the contours of isotemperature regions are treated as feature points to be matched; there are no salient features inside isotemperature regions. In this paper, each contour pixel of an isotemperature region is categorized as one of three types depending on whether its left (type 1), right (type 2) or both (type 3) horizontal neighbors are outside the region, as illustrated in Fig.4. Pixels indicated by the white square in Fig.4 are equivalent to pixels inside regions and are not matched. Type 1 pixels are matched to type 1 or type 3; type 2 are matched to type 2 or type 3; type 3 can be matched to any of the three types.

Initial contour matching is performed independently for the stereo thermal image pairs at $t = 0$ and $t = 1$. Contour matching is performed in two steps: contour pixel matching and contour segment matching.

In contour pixel matching, for each contour pixel of a particular type in one thermal image, a matching pixel is sought in the other image, subject to the matching conditions listed above and to the epipolar constraint. If multiple candidates are found, the matching pixel is selected based on intensity image information. In the current implementation, we calculate the error in a potential match as follows. Let $I_L(i,j)$ and $I_R(i,j)$ denote the intensity values at (i,j) of the left and right intensity image, respectively, where $0 \le i \le xl - 1$ and $0 \le j \le yl - 1$, and xl and yl are the number of pixels for the x (horizontal) and y (vertical) direction. We choose as the most likely corresponding candidate pixel the pixel that minimizes

$$e = \sum_{\Delta x=-1}^{1} \sum_{\Delta y=-1}^{1} |I_L(i + \Delta x, j + \Delta y) - I_R(i + \Delta x, j + \Delta y)|. \qquad (4)$$

The error e in Eq.(4) is calculated for each corresponding candidate. If e is larger than a threshold, the candiate is considered not to correspond to the pixel and is rejected. The candidate pixel with minimum e is initially assigned to the pixel. If a pixel does not have any candidate, an identifier indicating this is assigned to the pixel.

However, the resulting matches may fragment contours. That is, contour pixels belonging to different isotemperature regions could be paired with neighboring contour pixels. To avoid this, contour segment matching is performed, where a contour segment is defined as a sequence of 8-adjacent contour pixels of the same type (Fig.4). The correspondence between contour segments is determined by the number of matched pixel pairs.

Examples of the contour segment matching are shown in Fig.5. In Fig.5(a), the isotemperature regions A_1, A_2 and A_3 are found in the left thermal image, and B_1 and B_2 are found in the right image. The type 1 contour pixels of the regions are numbered as shown in Fig.5(a), and the numbers written inside the contour pixels indicate the numbers of the matched pixels. For instance, pixel 1 of A_1 is matched to pixel 3 of B_1, and pixel 2 of A_1 is matched to pixel 14 of B_2. The matches of the two pixels are examples of the problem of contour pixel matching. In order to match contour segments $L(1)$ of region A_1, $L(2)$ of A_2, and $L(3)$ of A_3 in the left image with segments $R(1)$ of B_1 and $R(2)$ of B_2 in the right image, we observe, for example, that five contour pixels of A_1 are paired with pixels of B_1, while only two are paired with B_2. Therefore, as shown in Fig.5(b), $L(1)$ is judged to be matched to $R(1)$, and the matches of pixels 2 and 3 of $A_1(L(1))$ are changed to reflect this. Thus, the matched contour segment is determined by choosing the contour segment with the largest number of matched pixels. Finally, the contour segment matches $L(1)$-$R(1)$ and $L(2)$-$R(2)$ are determined, while $L(3)$-$R(1)$ is rejected because the number of matched pixels is smaller than that of $L(1)$-$R(1)$.

6

3.2.3 Temporal correspondence of contour segments

Since the initial matching of contour segments at $t = 0$ and $t = 1$ is performed independently, matching results could contain errors. However, using temporal correspondence between two stereo pairs, in addition to the spatial stereo matching within each pair, can make the contour segment matching more robust. Correspondences between contour segments at time instants $t = 0$ and $t = 1$ are computed in step (c) in Fig.3, as described below.

In the stereo thermal images at $t = 1$, we examine the contour segments located near the contour segments from $t = 0$. For example, referring to Fig.6, suppose that at $t = 0$ the contour segments $L\_1_{t=0}$ and $L\_2_{t=0}$ are found in the left thermal image, and $R\_1_{t=0}$, $R\_2_{t=0}$ and $R\_3_{t=0}$ are found in the right image. For each contour segment at $t = 0$, we search for a contour segment at $t = 1$ that satisfies the following conditions:

- satisfies the type constraint described in Section 3.2.2.

- the region to which the contour segment at $t = 0$ belongs overlaps the region to which the contour segment at $t = 1$ belongs.

In Fig.6, $L\_1_{t=0}$ is type 1 and therefore is matched to type 1 or type 3 contour segments at $t = 1$. Suppose $L\_1_{t=0}$ belongs to the isotemperature region $Rg_{t=0}$. According to the second condition, $L\_1_{t=0}$ is matched to a contour segment that belongs to a region that overlaps $Rg_{t=0}$. In Fig.6, the region $Rg_{t=1}$ overlaps $Rg_{t=0}$ in the hatched area. Therefore, either of the segments $L\_1_{t=1}$ or $L\_2_{t=1}$, both of which belong to $Rg_{t=1}$ and are type 1, could correspond to $L\_1_{t=0}$. As in this example, if there is more than one matching contour segment, the nearest one is chosen in the following way by dynamic programming (DP).

Let C_1 and C_2 be two contour segments whose component pixels are $C_1 = c_{1_0}, c_{1_1}, \cdots, c_{1_{N_1}}$ and $C_2 = c_{2_0}, c_{2_1}, \cdots, c_{2_{N_2}}$, where N_1 and N_2 are the numbers of the pixels included in C_1 and C_2 respectively. Then, in the DP matching between C_1 and C_2, if u is a mapping from C_1 to C_2, the similarity between C_1 and C_2 is represented by

$$D(C_1, C_2; u) = \sum_{i=0}^{N_1} d(c_{1_i}, c_{2_{u(i)}}). \tag{5}$$

In Eq.(5), as shown in Fig.7, $d(c_{1_i}, c_{2_j})$ is the distance between the pixels c_{1_i} and c_{2_j}, given by $d(c_{1_i}, c_{2_j}) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, where (x_i, y_i) and (x_j, y_j) are the x and y image coordiantes of c_{1_i} and c_{2_j}, respectively. DP matching is performed to find the u^* that minimizes $D(C_1, C_2; u)$ for $u \in U(N_1, N_2)$, which is the entire set of u. The contour segment that gives the minimum value for $D(C_1, C_2; u^*)$ is chosen as the nearest matching segment. In Fig.6, for $L\_1_{t=0}$, $L\_1_{t=1}$ is chosen by Eq.(5), because $L\_1_{t=1}$ is located closer to $L\_1_{t=0}$ than $L\_2_{t=1}$.

3.2.4 Matching contour segments

The temporal correspondence described in Section 3.2.3 is used to make final decisions for matching contour segments at $t = 0$ in (d) in Fig.3. However, if, for example, contour segments in the left stereo images at $t = 0$ and $t = 1$ correspond temporally, the matching segments in the right stereo images do not necessarily correspond temporally. This inconsistency could be caused by errors introduced in the initial contour matching described in Section 3.2.2 or in the temporal correspondence described in Section 3.2.3, but the inconsistent match can nevertheless be treated as a hypothesis for a possible match.

Let S_I be the set of contour segment pairs matched by the initial contour matching at $t = 0$. In the left image, if s_1, a contour segment at $t = 0$, is temporally matched to s_2, a contour segment at $t = 1$; if s_2 is matched with s_3, a contour segment at $t = 1$ in the right image by the initial contour based matching; and if s_3 is temporally matched to s_4, a contour segment at $t = 0$ in the right image; then s_1 is indirectly matched with s_4. Let S_L be the set of contour segment pairs matched indirectly, like s_1 and s_4. Similarly, let S_R be the set of contour segment pairs matched indirectly in the right image. Then, the set of possible matches (match hypotheses), S_T, is:

$$S_T = S_I \cup S_L \cup S_R. \tag{6}$$

Thus, compared with the set of initial matches (S_I), S_T includes more possible matches, which may result in more reliable matching results.

For example, in Fig.6, suppose that the contour segments $L\_2_{t=0}$ and $L\_3_{t=1}$ are temporally matched in the left image, and that in the right image, $R\_2_{t=0}$ and $R\_3_{t=0}$ are temporally matched to $R\_2_{t=1}$ and $R\_3_{t=1}$, respectively. Suppose that by the initial contour segment matching $L\_2_{t=0}$ is matched with $R\_2_{t=0}$, and $L\_3_{t=1}$ is matched with $R\_3_{t=1}$. Then, in Eq.(6), the elements of S_I are the pairs of $(L\_2_{t=0}, R\_2_{t=0})$ and $(L\_1_{t=0}, R\_1_{t=0})$. Elements of S_L include the pair $(L\_2_{t=0}, R\_3_{t=0})$, which is also an element of S_R. Thus, S_T includes three pairs, whereas S_I includes only two.

The hypotheses (S_T in Eq.(6)) are verified by the method described below so that consistent and likely matched contour segment pairs are determined. The basic idea for the hypothesis verification is to calculate the similarities between the pairs included in S_T, and to choose consistent pairs with high similarities, as in the previous section. The similarity function is different from the one in Eq.(5), however.

As illustrated in Fig.8, let C_1 and C_2 be two contour segments. Let the intensities of the pixels included in C_1 and C_2 be $I_{1_0}, I_{1_1}, \cdots, I_{1_{N_1}}$ and $I_{2_0}, I_{2_1}, \cdots, I_{2_{N_2}}$, where N_1 and N_2 are the numbers of pixels included in C_1 and C_2, respectively. Then, in the DP matching between C_1 and C_2, if v is a mapping from C_1 to C_2, the similarity between C_1

and C_2 is given by

$$Sim(C_1, C_2; v) = \sum_{i=0}^{N_1} I_{iv(i)}, \qquad (7)$$

where $I_{ij} = |I_{1_i} - I_{2_j}|$. DP matching is performed to find the v^* that minimizes $Sim(C_1, C_2; v)$ for $v \in V(N_1, N_2)$, which is the entire set of v.

The similarities of the contour segment pairs in S_T are calculated using Eq.(7), and the pair with the smallest similarity value is chosen as the matching contour segment pair. Among the remaining pairs, there could be pairs that overlap in the y direction with the contour segments already matched. For example, in Fig.8(b), if C_L is matched with C_{R_1}, C_{R_2} cannot be matched with C_L any more. Such pairs are eliminated, because a pixel in one image cannot be matched with more than one pixel in the other image. Again, among the remaining pairs, the pair with the smallest similarity is chosen, and the pairs inconsistent with the chosen pair are eliminated. This recursive process is repeated until all of the pairs are matched or eliminated. In the example of Fig.6, if the pair $(L\_2_{t=0}, R\_2_{t=0})$ is chosen by Eq.(7), the pair $(L\_2_{t=0}, R\_3_{t=0})$ is rejected.

3.3 Isotemperature line matching: Mode 2

At $t = i$ ($i \geq 1$), as shown in Fig.2, Mode 2 is used for contour segment matching. As illustrated in Fig.9, most of the modules of Mode 2 are similar to those of Mode 1.

At (a) in Fig.9, the left and right thermal images are segmented into isotemperature regions. The process is the same as that described in Section 3.2.1. At (b) in Fig.9, initial contour based matching is performed in the same way as that described in Section 3.2.2.

For the temporal correspondence ((c) in Fig.9), the results of contour segment matching at $t = i - 1$ are used. In the stereo thermal images at $t = i$, contour segments located near the contour segments at $t = i - 1$ are searched in the same way as in Section 3.2.3. Then, matching of contour segments at $t = i$ is determined based on the temporal correspondence at (d) in Fig.9. Here, contour segments at $t = i$ are indirectly matched using the temporal correspondence and matching results at $t = i - 1$, and by Eq.(7), consistent and likely contour segment pairs are obtained.

3.4 Dense range data acquisition

Only sparse range data can be obtained by the contour segment matching described in Sections 3.2 and 3.3. To acquire range data within isotemperature regions as well as on contours, stereo intensity images are used ((b) in Fig.2).

The basic approach is to match intensity profiles along horizontal segments, where a horizontal segment is defined as the horizontal row of pixels corresponding to the intersection of an epipolar line with an isotemperature region in a thermal image. However, there

9

are two cases to be considered, depending on whether there is correspondence between the pixels at both ends of a pair of horizontal segments. Figure 10 illustrates the two cases.

Figure 10 shows a scene, viewed by a stereo pair of cameras, in which two objects are situated in front of a wall. The objects and the wall have different temperatures T_1, T_2, T_3, and T_4, and object #1 itself has two regions with temperatures T_2 and T_3. These temperatures are observed as T_1', T_2', T_3', and T_4'. For the horizontal segments, along epipolar line l, in the T_3' regions of the stereo thermal images (i.e. the range L_1 in the left image and R_1 in the right image), the pixels at both ends of L_1 and R_1 correspond to each other; that is, E_L' corresponds to E_R', and F_L' corresponds to F_R', because the entire region for T_3 can be seen from both cameras, which results in small values of e in Eq.(4) for points E and F.

On the other hand, the ends of segments L_2 and R_2 do not correspond, because A_L' and B_L' in the left image are the projections of A_L and B_L, while A_R' and B_R' in the right image are the projections of A_R and B_R. That is, A_L', A_R' B_L', and B_R' correspond to four different points in the scene. Therefore, the entire range between A_R and B_L in the scene cannot be seen from both cameras.

As in the examples of Fig.10, if the pixels at both ends of a horizontal segment in one image of a stereo pair correspond to the pixels at both ends of a horizontal segment in the other image, the entire horizontal segment is judged to be observed by both cameras, and the intensity profiles on the horizontal segments are matched by DP.

Let A and B be the intensity profiles on L_i and R_j. Let the intensities of pixels on A and B be $A = a_0, a_1, \cdots, a_I$ and $B = b_0, b_1, \cdots, b_J$, where I and J are the numbers of pixels included in A and B respectively. Then, in the DP matching between A and B, if w is a mapping from A to B, the similarity between A and B is represented by

$$D(A, B; w) = \sum_{i=0}^{I} d(a_i, b_{w(i)}), \qquad (8)$$

where $d(a_i, b_j) = |a_i - b_j|$. DP matching is used to find the w^* that minimizes $D(A, B; w)$ for $w \in W(I, J)$, which is the entire set of w. Then, the corresponding pixel pairs on A and B are $(a_0, b_{w^*(0)}), (a_1, b_{w^*(1)}), \cdots, (a_I, b_{w^*(I)})$. In Eq.(8), $d(a_i, b_j)$ is defined even for a_i and b_j with no matching pixel so that multiple assignment to a pixel can be avoided. This is important also for dealing with occlusions. That is, in the example of Fig.10, no pixel in the left image could be assigned to the pixels corresponding to the range between A_R and A_L in the right image, because in the left image the range is occluded.

In our original algorithm[7], if one or both ends of a pair of segments did not correspond to each other, edge pixel matching was performed to reduce false matches due to occlusion, but only sparse range data could thus be obtained. In this paper, horizontal segments are matched based on DP. In Fig.11, the horizontal segments L_1, L_2, and L_3 in the left thermal

10

image and R_1, R_2, R_3, and R_4 in the right image are examples of horizontal segments both of whose ends do not correspond. The similarities $D(L_i, R_j; w^*)$ between the intensity profiles of L_i and R_j ($i = 1, 2, 3; j = 1, 2, 3, 4$) are calculated by Eq.(8). Then, horizontal segment matches are determined by finding the w'^* that minimizes $D'(L', R'; w')$ in the following equation:

$$D'(L', R'; w') = \sum_{i=1}^{I'} d'(L_i, R_{w'(i)}),\qquad(9)$$

where $d'(L_i, R_j) = D(L_i, R_j; w^*)$, $L' = L_1, \cdots, L_{I'}$, and $R' = R_1, \cdots, R_{J'}$ (in Fig.11 $I' = 3$ and $J' = 4$). Consequently, the corresponding horizontal segment pairs are $(L_1, R_{w'^*(1)}), \cdots, (L_{I'}, R_{w'^*(I')})$, and matched pixel pairs are obtained from Eq.(8).

Once corresponding pixel pairs are obtained by the methods in Sections 3.2, 3.3 and 3.4, calculation of 3D coordinates for the pairs is performed based on triangulation ((c) in Fig.2).

4 Experimental results and discussion

4.1 Experimental conditions

To test the effectiveness of our method, we implemented an experimental system as per Fig.1. The infrared video camera can acquire a thermal image within a 3,000 to 5,400 nm wavelength band with 256 temperature levels at a frequency of 30 frames per second. The separating mirror reflects about 75% of the incoming infrared light and transmits about 85% of the visible light.

The four images from the four cameras are input to a multiple image display capable of displaying four time-sequential images simultaneously. The displayed images, comprising the four subimages, are recorded by a digital VTR, and the recorded sequence is subsequently uploaded to a workstation for processing. Each of the four subimages 325×199 is pixels, and the thermal images, which include thermal noise at each pixel, are smoothed by a spatial smoothing operator and quantized to 32 levels.

4.2 Results

Our method was applied to a scene with a man raising his right hand in front of a wall. Stereo thermal and intensity images for the scene at three time instants are shown in Fig.12. The disparity map (for the left image) obtained by the old version of the algorithm[7] for the first time instant of Fig.12 is shown in Fig.13, where a pixel with no disparity indicates the absence of range data. As shown in Fig.13, only very sparse range data can be obtained in the area corresponding to the wall. Figure 14 shows the disparity maps for Fig.12 computed by the algorithm proposed in this paper. It is obvious that

11

much denser range data can be obtained by the new algorithm. Since we do not apply an upper limit for disparities, noise-like disparities due to false matches for very short contour segments can be seen in the area corresponding to the man, but globally contour segment matching and intensity profile matching are accurate. Good results were obtained also for other time instants.

5   Conclusions

This paper has presented a method for acquiring time-sequential range images by integrating stereo pairs of thermal and intensity images. In our algorithm, first, contour segments of isotemperature regions are matched using temporal correspondece at successive time instants. Then, to supplement the sparse range data obtained by contour matching, intensity profiles are matched via dynamic programming.

The experimental results for scenes of moving human beings show that the method proposed in this paper can obtain much denser range data than the old version. The experiments also show that the matching results are accurate.

Our method does not suffer from differences due to viewpoint change, because corresponding pixels are restricted to those with the same temperature values in the left and right stereo thermal images. Another property is that our method is a passive method that can nevertheless acquire dense range image sequences. Remaining problems are to make the computations faster and to improve matching accuracy.

References

[1] J.Ohya, Y.Kitamura, H.Takemura, F.Kishino and N.Terashima, "Real-time reproduction of 3D human images in virtual space teleconferencing," Proc. of IEEE Virtual Reality Annual International Symposium (VRAIS'93), pp.408-414, Sep. 1993.

[2] M.Yamamoto "Direct estimation of deformable motion parameters from range image sequence", Proc. of 3rd ICCV, pp460-464, Dec. 1990.

[3] A.M.Waxman and J.H.Duncan, "Binocular image flows: Steps toward stereo-motion fusion", IEEE Tr. PAMI, Vol.PAMI-8, No.6, pp715-729, Nov. 1986.

[4] Y.Ohta and T.Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," IEEE Tr. PAMI, Vol.PAMI-7, No.2, pp.139-154, Mar. 1985.

[5] M.Okutomi and T.Kanade, "A multiple-baseline stereo," IEEE Tr. PAMI, Vol.15, No.4, pp.353-363, Apr. 1993.

[6] R.C.Bolles, H.H.Baker and D.H.Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," Int. J. Computer Vision, Vol.1, No.1, 1987.

[7] J.Ohya and F.Kishino, "A new method for acquiring time-sequential range images by integrating stereo pairs of thermal and intensity images," Proc. of CVPR'93, pp.730-731, Jun. 1993.

[8] C.Caillas, "Autonomous robot using infrared thermal camera to discriminate objects in outdoor scene", SPIE Vol.1293, Applications of Artificial Intelligence VIII, pp801-812, 1990.

[9] N.Nandhakumar and J.K.Aggarwal, "Integrated analysis of thermal and visual images for scene interpretation", IEEE Tr.PAMI, Vol.10, No.4, pp469-481, Jul. 1988.

[10] J.Silverman, J.M.Mooney and F.D.Shepherd, "Infrared video camera", Scientific American, pp78-83, Mar. 1992.

[11] J.Ohya and F.Kishino, "Time-sequential range image acquisition method by stereo integration of thermal images and intensity images", Proc. of 45th National Convention of IPSJ, 7J-1, pp2-259 - 2-260, Oct. 1992 (in Japanese).

Fig.1 Principle

14
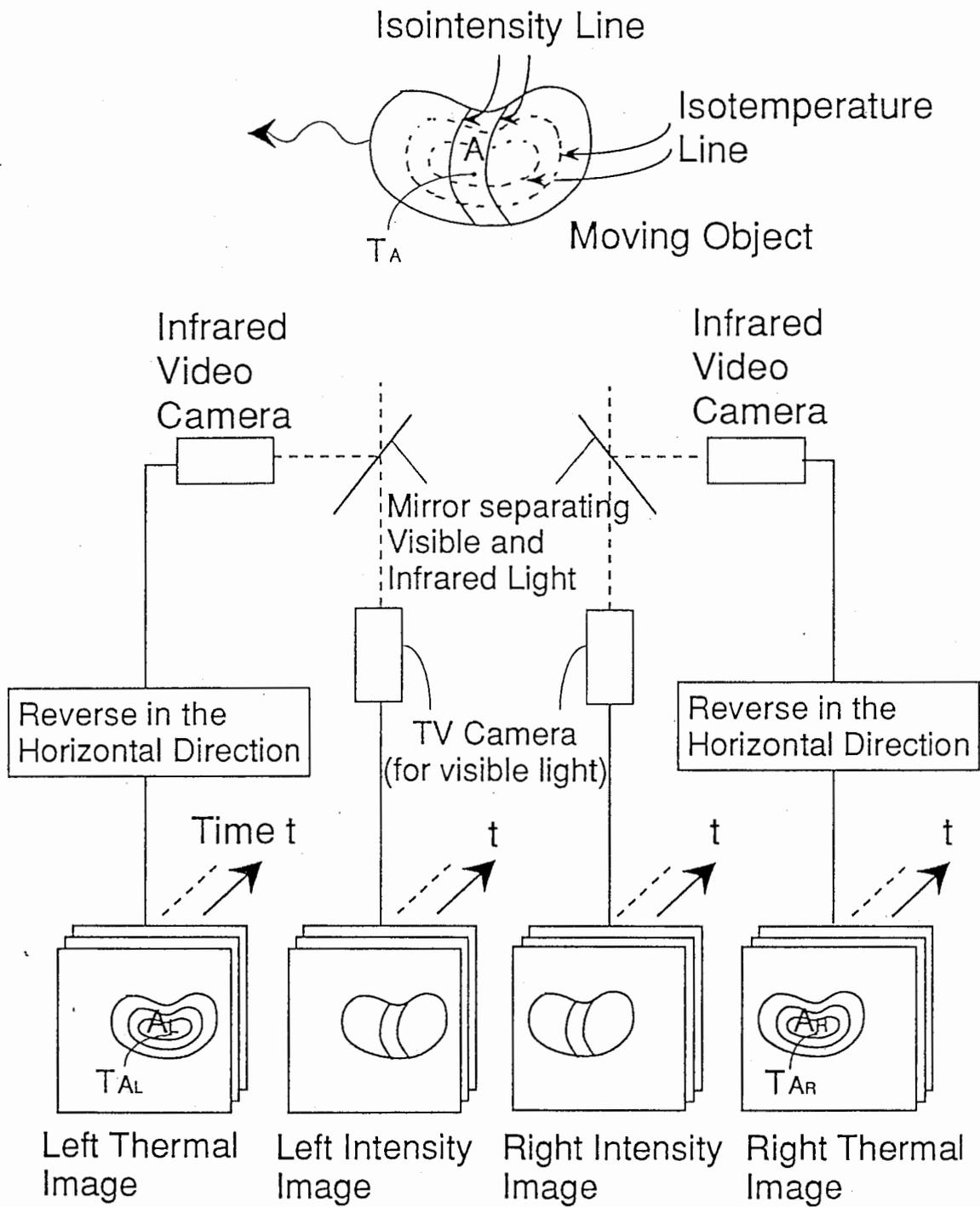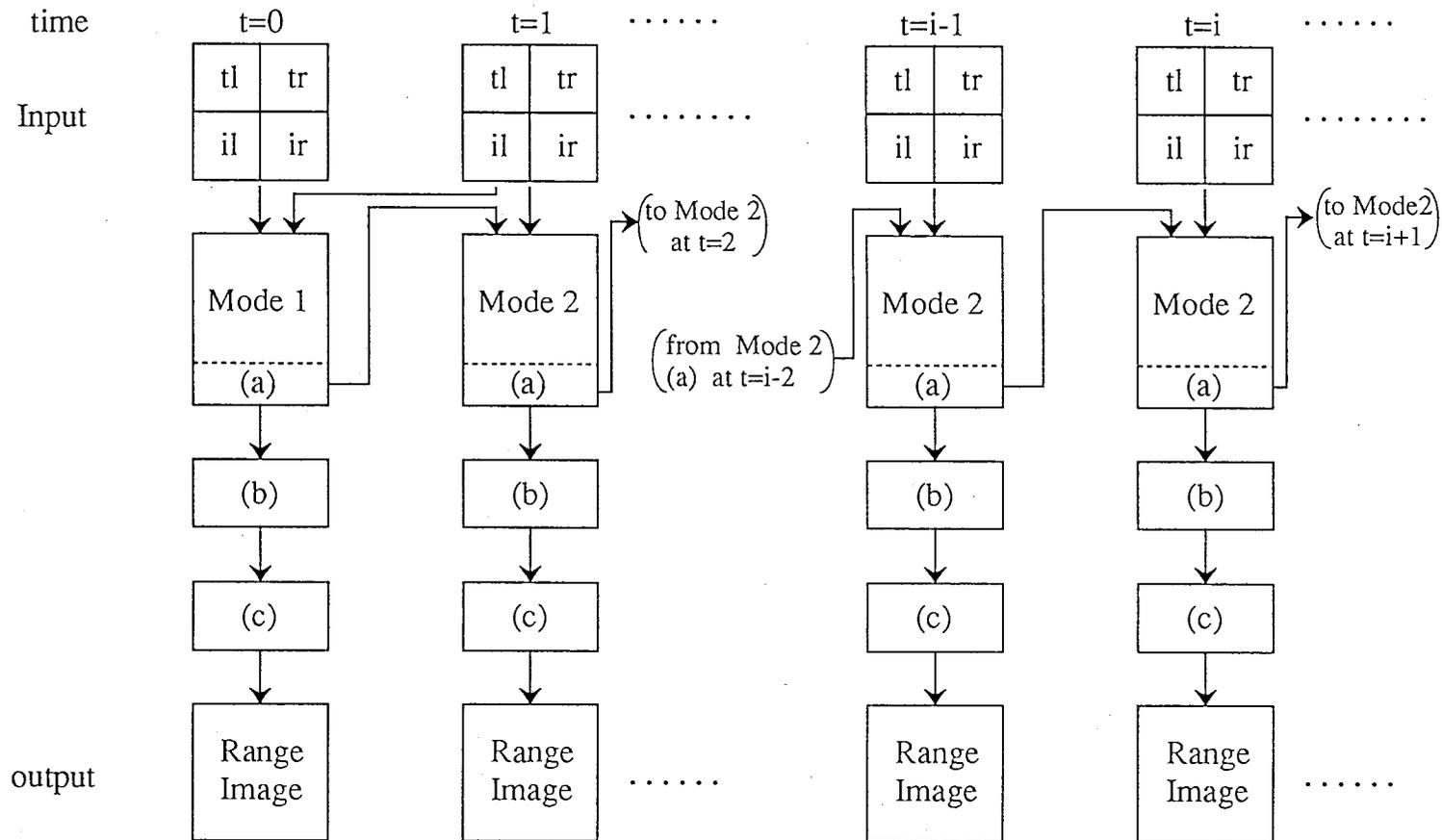
Fig.2   Overview of the Algorithm

tl , tr :   left and right thermal image, respectively
il , ir :   left and right intensity image, respectively
(a) :   Contour segment matching results
(b) :   Dense range data acquisition
(c) :   Range Calculation
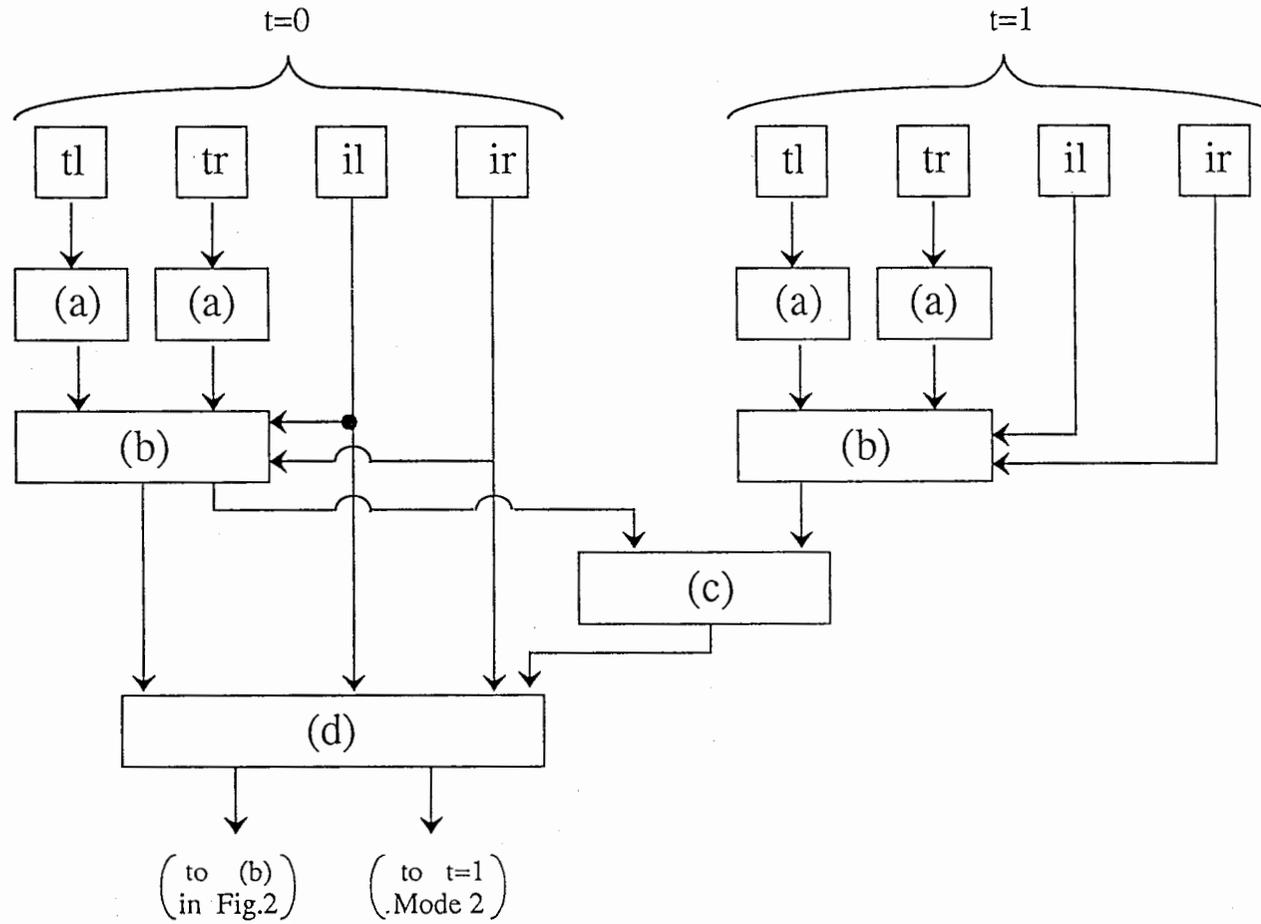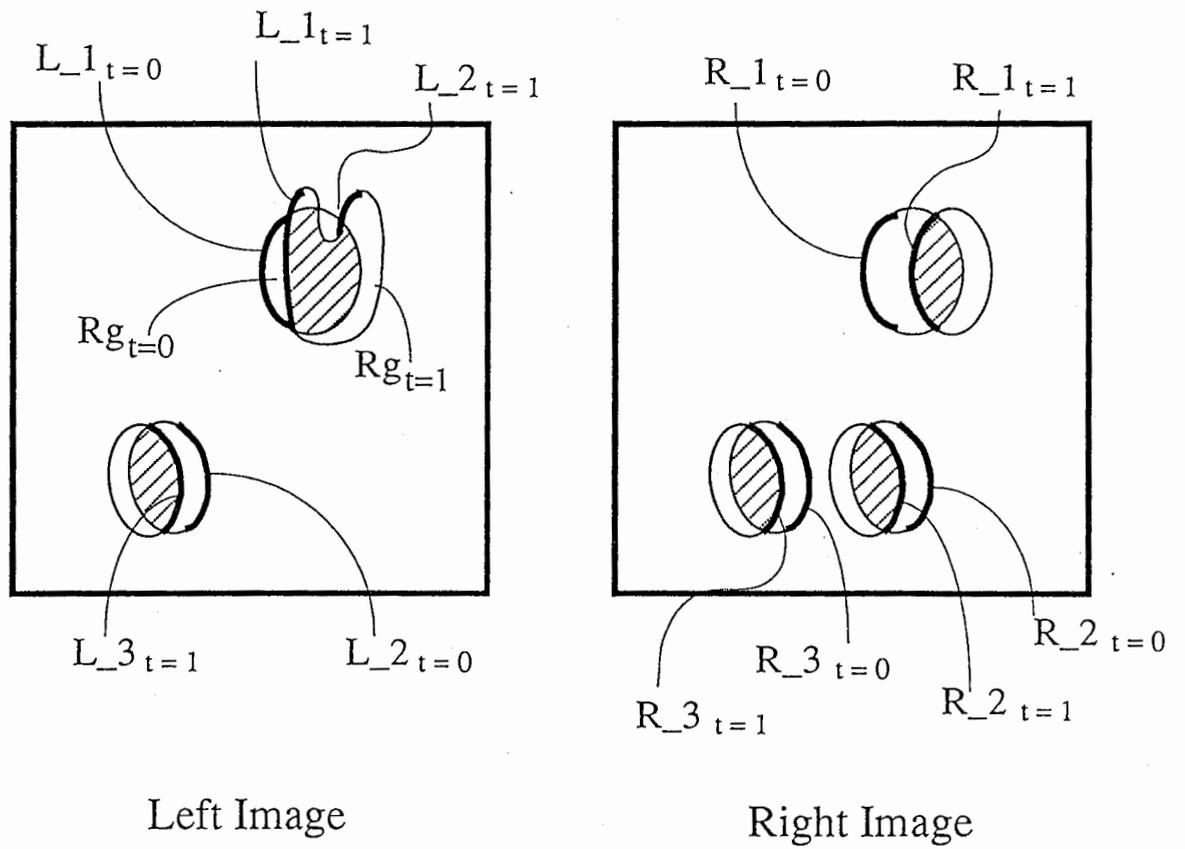
Fig. 3   Mode 1

(a)   Segmentation into isotemperature regions
(b)   Initial contour based matching
(c)   Temporal correspondence of contour segments
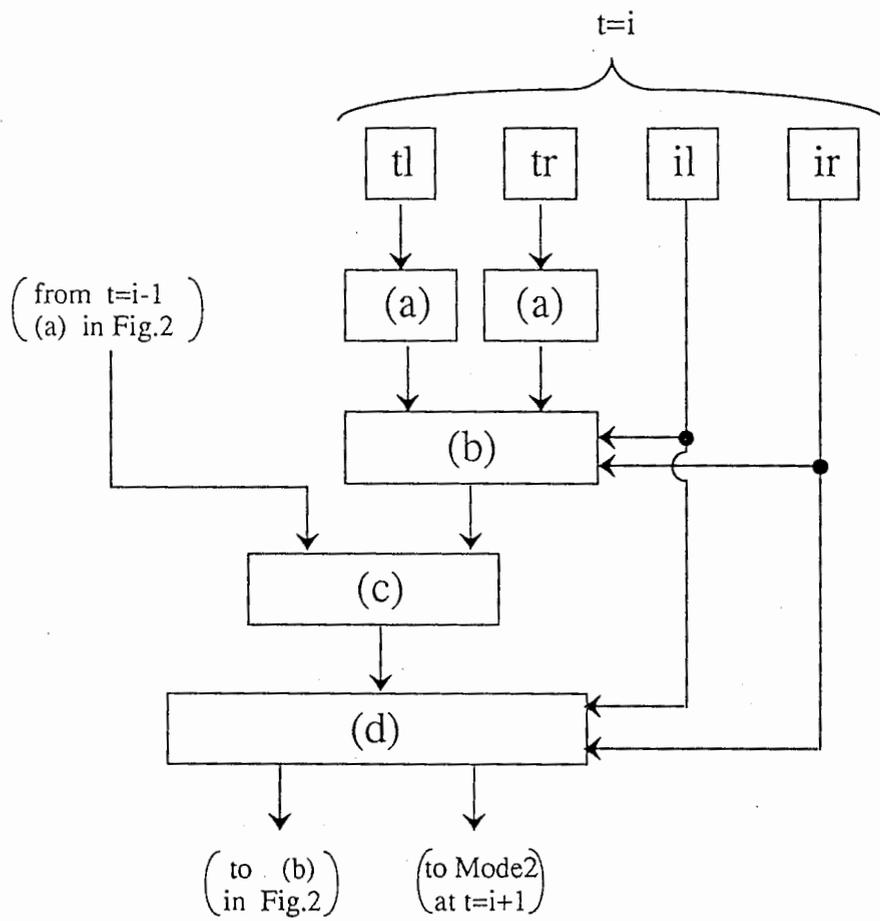(d)   Contour segment matching
tl, tr, il, ir : Refer to Fig.2
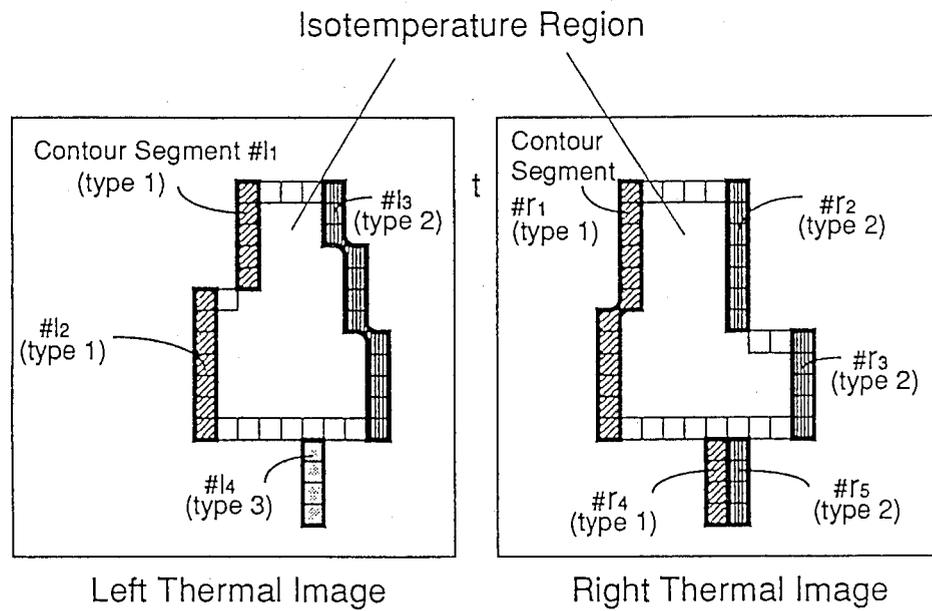
Fig.4 Contour segment

▨ : type1 pixel, whose neighboring pixels are Ⓝ▨Ⓨ , where Y and N indicate whether or not the pixels belong to the same isotemperature region as the contour pixel.

▥ : type2 pixel : Ⓨ▥Ⓝ
▦ : type3 pixel : Ⓝ▦Ⓝ
☐ : other contour pixel :

17
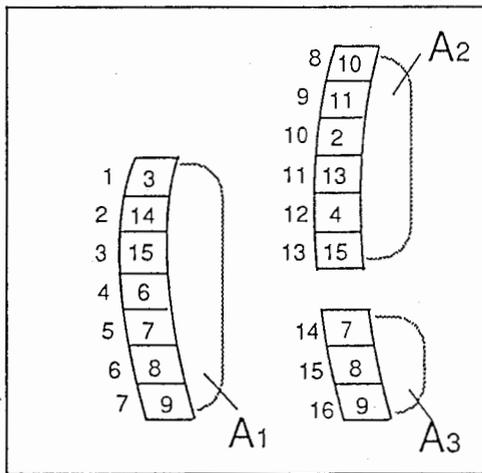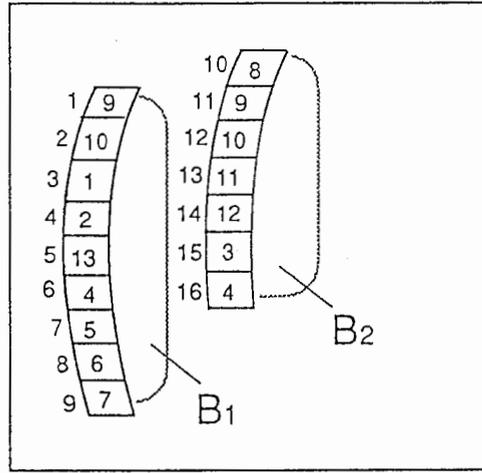
(a) Contour pixel matching

(b) Contour segment matching

Fig.5 Contour based matching

Left Image   Right Image

Fig. 6   Temporal Correspondence
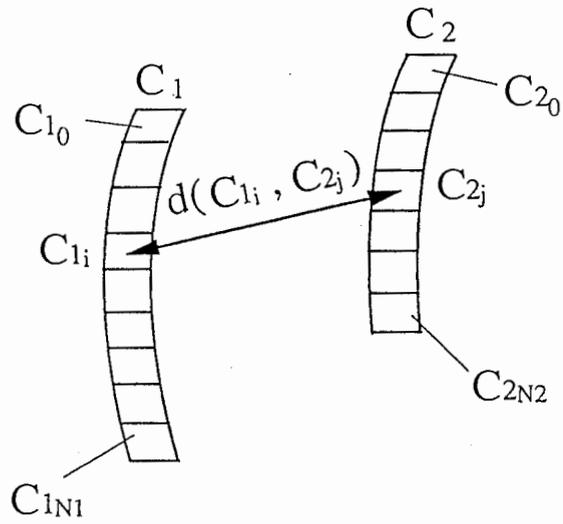
Fig.7   Distance between two contour segments
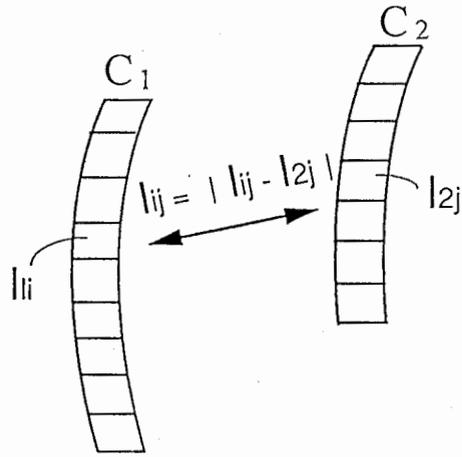
(a)
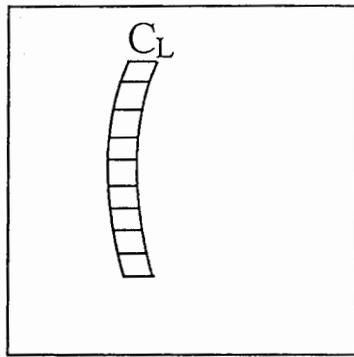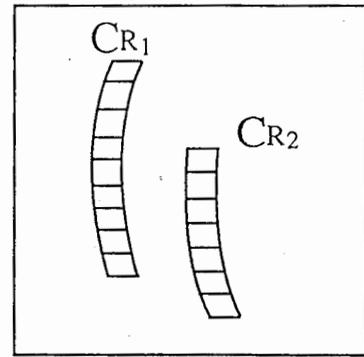
Left Image   Right Image

(b)

Fig.8   Contour Segment Matching

Fig. 9   Mode 2

   (a)   Segmentation into isotemperature regions
   (b)   Initial contour based matching
   (c)   Temporal correspondence of contour segments
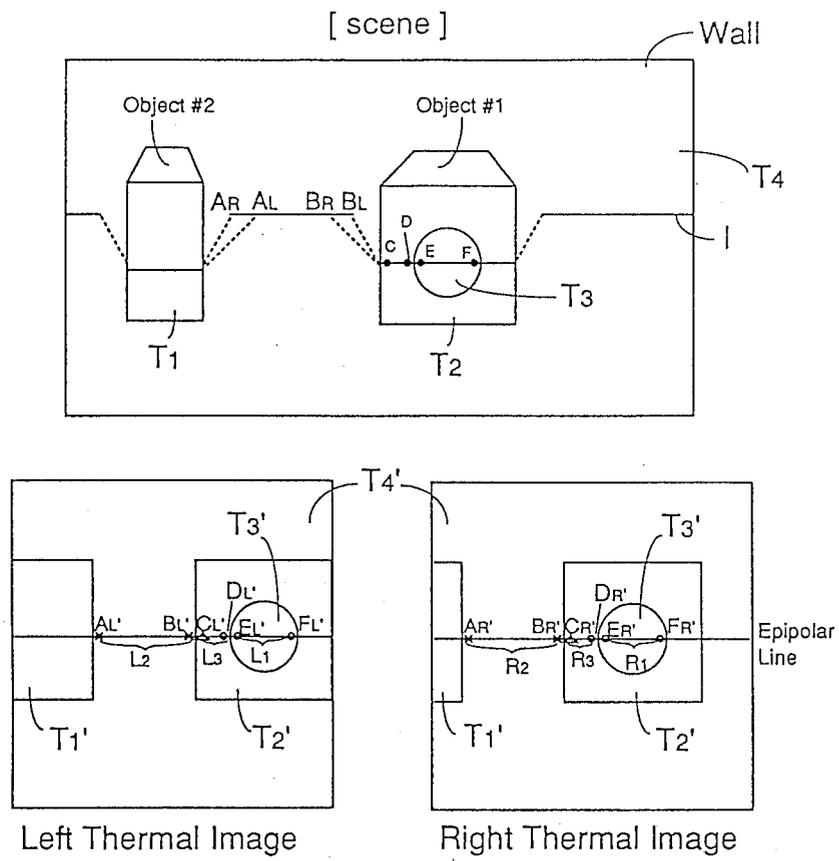   (d)   Contour segment matching
      tl, tr, il, ir : Refer to Fig.2

[scene]

Wall

Object #2   Object #1

A_R A_L B_R B_L

D

T4

C

E   F

I

T3

T1   T2

T4'

T3'   T3'

$D_{L'}$ $D_{R'}$

$A_{L'}$ $B_{L'}$ $C_{L'}$ $F_{L'}$ $F_{L'}$ $A_{R'}$ $B_{R'}$ $C_{R'}$ $F_{R'}$ $F_{R'}$ Epipolar
L_2 L_3 L_1 R_2 R_3 R_1 Line

T1'   T2'   T1'   T2'

Left Thermal Image   Right Thermal Image

Fig.10 Range estimation inside isotemperature regions

Same Temperature

L_1 L_2 L_3 R_1 R_2 R_3 R_4

Left Image   Right Image
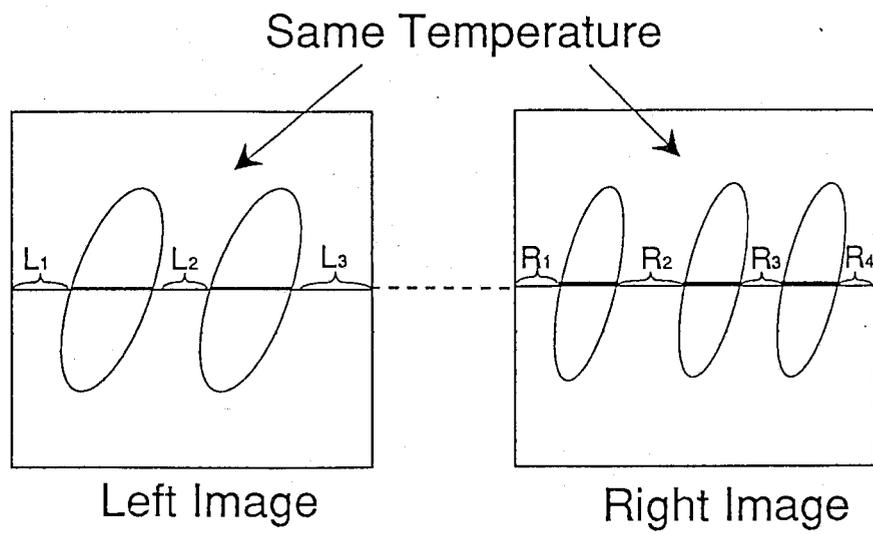
Fig.11 Horizontal segment matching

Left thermal image	Right thermal image	Left intensity image	Right intensity image

t=0

t=1

t=2
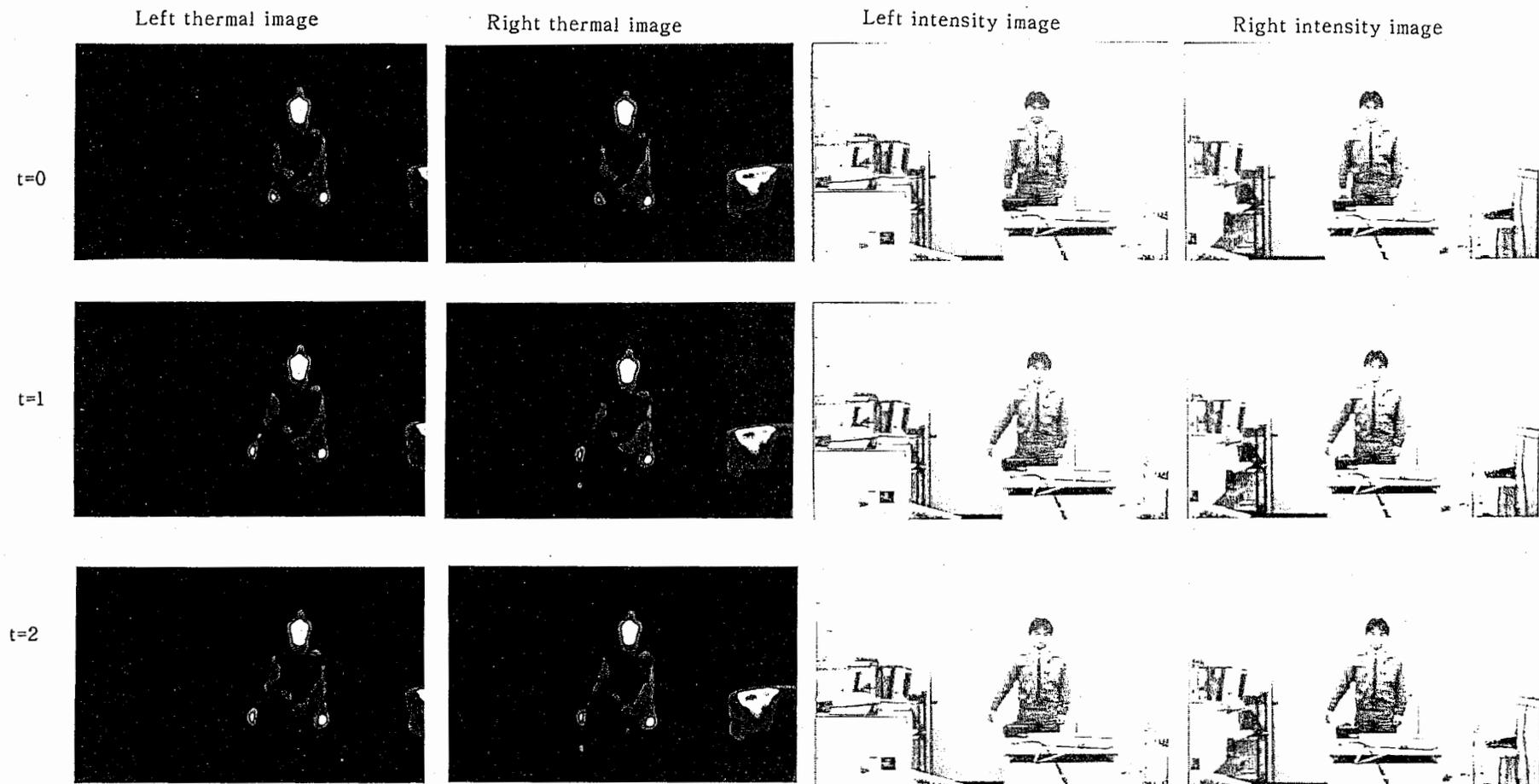
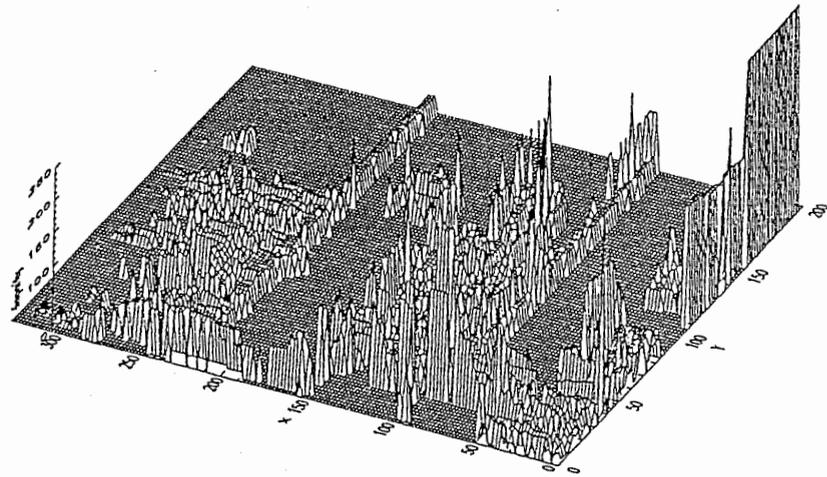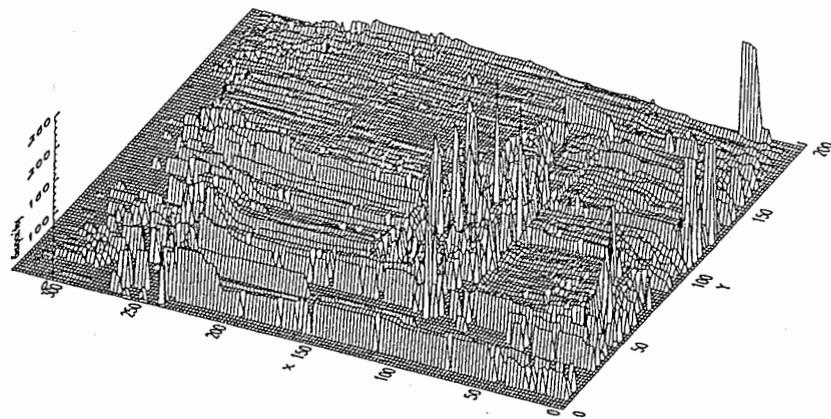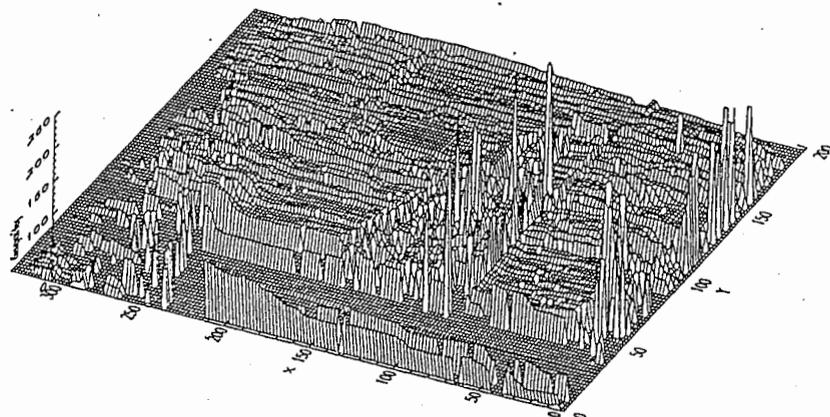Fig.12 Time-sequential images for scenes having a moving man

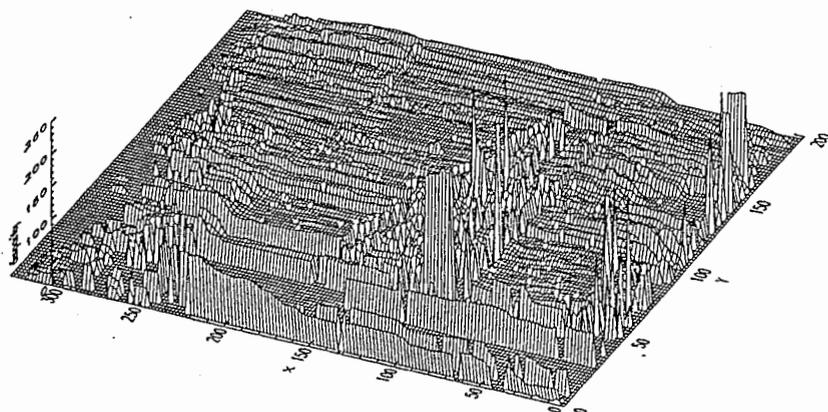Fig.13 Disparity map by the old version

$(t = 0$ in Fig.12$)$

$t = 0$

$t = 1$

$t = 2$

Fig.14 Disparity maps by the proposed method

26