Real-Time Target Localization and Tracking by N-Ocular Stereo

Takushi Sogo
Department of Social Informatics
Kyoto University
Sakyo-ku, Kyoto 606–8501, Japan

Hiroshi Ishiguro
Department of Computer and
Communication Sciences
Wakayama University
930 Sakaedani, Wakayama 640–8510, Japan

Mohan M. Trivedi
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093–0407, U.S.A.

Abstract

In recent years, various practical systems using multiple vision sensors have been proposed. In this paper, as an application of such vision systems, we propose a real-time human tracking system consisting of multiple omnidirectional vision sensors (ODVSs). The system measures people's locations by N-ocular stereo, *which is an extension of trinocular stereo, from omnidirectional images taken with the ODVSs. In addition, the system employs several compensation methods for observation errors in order to achieve robust measurement. We have evaluated the proposed methods in the experimentation using four compact ODVSs we have originally developed.*

1. Introduction

Recent progress of multimedia and computer graphics is developing practical applications based on simple computer vision techniques. Especially, the practical approach recently focused on is to use multiple vision sensors with simple visual processing. For example, several systems track people or automobiles in the real environment with multiple vision sensors [1, 2, 10, 13, 14] and other systems analyze their behaviors and so on [3]. Compared with systems using a single vision sensor [6, 9, 12, 21], these systems enable to observe a moving target in a large space for a long time. However, they need to use many vision sensors to seamlessly cover the environment since a single standard vision sensor itself has a narrow range of view. On the other hand, an omnidirectional vision sensor (ODVS) provides a wide range of view. In addition, use of multiple ODVSs provide rich and redundant visual information, which enables robust recognition of the targets. Thus, multiple ODVSs opens up a new application area of computer vision with their wide range of view.

Figure 1. Compact omnidirectional vision sensor

As an application of such vision systems, we propose a real-time human tracking system using multiple ODVSs [16]. We have originally developed low-cost and compact ODVSs as shown in Figure 1 [7] and used for this research. The system detects people, measures azimuth angles with the ODVSs, and determines their locations by triangulation as shown in Figure 2 (a). In this system, the following problems in the stereo using ODVSs (called omnidirectional stereo) should be considered:

- Correspondence problem among multiple targets
- Measurement precision of target locations

The former problem also occurs in conventional stereo using two or more vision sensors [11, 14, 17]. However, in our system it is more difficult to verify the correspondence of targets with visual feature, since the baseline of ODVSs

0-7695-0704-2/00 $10.00

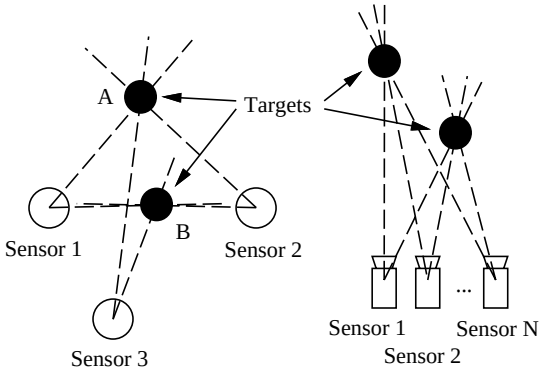


Figure 2. Omnidirectional stereo and conventional stereo

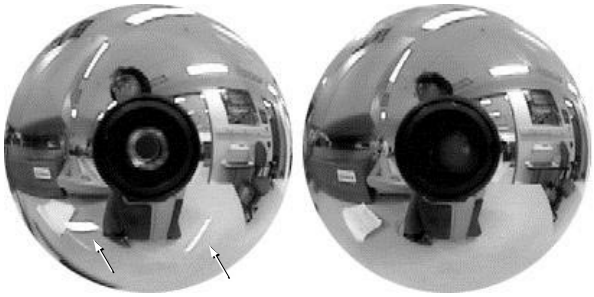
is much longer than that of conventional stereo. The latter problem is that the measurement precision of a target location becomes very low when the target locates along the baseline of two sensors [8]. For example, the target B in Figure 2 (a) locates on the baseline of sensors 1 and 2, so that the target location measured by them is unstable. Generally, this problem does not occur in conventional stereo, but in omnidirectional stereo, since omnidirectional stereo assumes arbitrary locations of sensors and targets. Furthermore, in the real-time human tracking system, deformative human bodies should be properly handled.

In order to solve the above problems, we have extended trinocular stereo [4, 23]. The extended method, called *N-ocular stereo*, verifies correspondence of multiple targets without visual features. In addition, we have developed several compensation methods of observation errors for measuring target locations more robustly and accurately.

In the following, we first introduce the development of a *low-cost and compact ODVS*, which makes systems practical that consist of many ODVSs. Then, N-ocular stereo is explained, which measures target locations using multiple ODVSs. Furthermore, the simplified N-ocular stereo and error compensation methods for real-time processing are described in detail. Finally, we show experimental results of tracking people by the real-time human tracking system.

2. Development of compact ODVSs

The ODVS has been first proposed by Rees [19] in the patent submitted to US government in 1970. Then, Yagi [24], Hong [5] and Yamazawa [26] developed again in 1990, 1991 and 1993, respectively. Recently, Nayar [15] has geometrically analyzed the complete class of single-lens single-mirror catadioptric imaging systems and developed an ideal ODVS using a parabola mirror.



With a black needle

Figure 3. Internal reflections

In these previous works, researchers developed ODVSs as prototypes and investigated properties of *Omnidirectional Images* (ODIs) taken by ODVSs. Therefore, the developed ODVSs were not so compact and their costs were high. In order to develop practical vision systems, we have proposed original ideas and developed *low-cost and compact ODVSs* [7]. Figure 1 shows the developed compact ODVS, the height of which is about 6cm including a CCD camera unit.

One of the important components of the ODVS is an apparatus supporting the mirror. For the supporting apparatus, there are two requirements:

- Eliminating internal reflections by the supporting apparatus.
- Precise surface for acquiring non-distorted ODIs.

Many types of ready-made clear hollow cylinders are available and the surface precision is sufficiently high. However, the clear hollow cylinder has a serious problem of internal reflections (see the left image in Figure 3). Our idea to eliminate the internal reflections is to equip a black needle along the main axis of the cylinder. The light reflecting on the internal surface of the hollow cylinder and projected onto the image plane passes through the main axis of the cylinder. Therefore, the black needle set along the main axis completely eliminates the internal reflections (see the right image in Figure 3).

3. Multiple camera stereo

3.1. The correspondence problems and trinocular stereo

In order to detect targets and measure their locations, our system uses multiple ODVSs. Since each ODVS is fixed in the environment, the system can detect targets in omnidirectional images by background subtraction. Then, it measures the azimuth angles of the targets. If the locations and the orientations of the ODVSs are known, the locations of the targets can be measured from the azimuth angles by triangulation (see Figure 4).

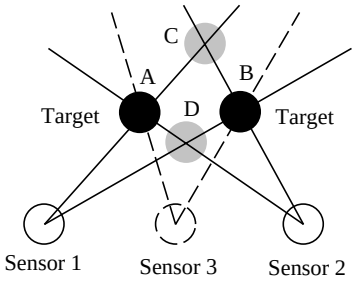


Figure 4. The correspondence problem and trinocular stereo

In triangulation, multiple targets in the environment cause the correspondence problem. In Figure 4, for example, there are two targets (the black circles indicate actual target locations), however, from azimuth angles observed by the sensors 1 and 2, it is estimated by triangulation that the targets may exist at A through D in Figure 4. In general, this correspondence problem can be solved by using visual features of the targets. In our system, however, it is difficult to verify the correspondence of targets with visual features, since ODVSs observe targets from various points of view and their visual features may differ. Alternatively, the correspondence problem can also be solved by using three or more sensors. In Figure 4, the locations C and D are verified with the sensor 3, then they are eliminated since the sensor 3 does not observe the targets in these directions. This technique is known as *trinocular stereo* [4, 23], and it can be applied to our system for verifying the target correspondence.

3.2. Problems of previous methods

3.2.1 Observation errors

When applying trinocular stereo to actual systems, observation errors should be considered. In Figure 4, for example, the lines indicating azimuth angles of the target A and B exactly intersect at one point, however, in practice they do not intersect in this manner because of observation errors. In this case, clusters of intersections are considered as target locations [18, 20, 22], in general.

However, information of vision systems is much noisy compared with that of radar systems. Furthermore, our system cannot precisely detect the azimuth angles of targets because of the following reasons:

- If targets locate near sensors, they are widely projected on the ODVSs.
- Targets are humans which have deformative bodies. In addition, each vision sensor observes the target from various points of view.

The previous methods for localizing targets do not consider these problems. In addition, binocular stereo using ODVSs has a low-precision problem with respect to targets locating along the baseline of the sensors [8]. These problems should be carefully considered in our approach.

3.2.2 Computational costs

In order to solve the correspondence problems and to measure target locations properly, each azimuth angle of the targets detected by the sensors should have an association with at least one of measured locations. Assignment of azimuth angles is an optimization problem of NP-hard [18]. Several methods for solving it have been proposed so far [18, 22], however, these methods need iterative computation (more than 10 or 300 times). Therefore, a more efficient method is needed for real-time processing.

The aim of N-ocular stereo proposed in this paper is to solve the correspondence problem with low computational cost within acceptable quality of solutions, rather than to compute optimal ones. Basically, N-ocular stereo is also NP-hard. However, it does not require iterative computation. In addition, the computational costs can be reduced by checking the apparent sizes of targets and ignoring small (distant) ones.

In the following, we mainly discuss the handling of the observation errors.

4. Localization of targets by N-ocular stereo

4.1. Basic algorithm

In trinocular stereo, three vision sensors are used to measure the target location and to verify the correspondence. On the other hand, in N-ocular stereo, more than three vision sensors are used. This is based on the idea that observation errors are reduced by using much visual information.

The basic process of N-ocular stereo is as follows:

1. Measure the location of a target from azimuth angles detected by a pair of arbitrary vision sensors as shown in Figure 4 (binocular stereo).
2. Check if another sensor observes the target at the location measured with $(N-1)$-ocular stereo. If so, the location is considered as a result of N-ocular stereo (see A and B in Figure 4). Iterate this step from $N=3$ to $N=$(the number of sensors).
3. Finally, the locations measured with only two sensors (C and D in Figure 4) are considered as wrong matchings, and erased from the list of candidates.

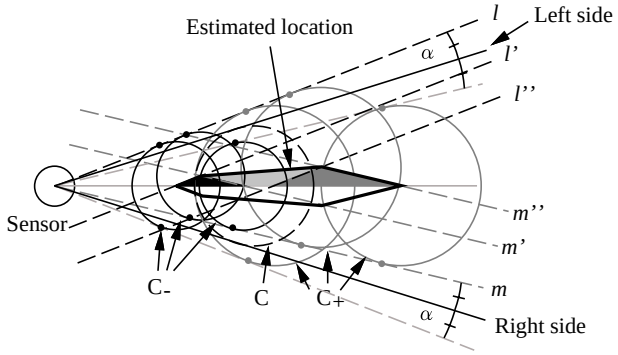


Figure 5. Localization of a target considering observation errors

4.2. Localization of targets and error handling

As described in Section 3.2, observation errors of azimuth angles should be considered when measuring people's locations, since the human body deforms every moment and is widely projected on the ODVSs. Here, we suppose the human body is represented with a circle of a constant radius, and the location of a person is represented as the center of the circle. The errors in the model matching can be handled with the following two parameters:

- α: Detection errors of the right and left side of a target
- β: An error of the human model, i.e., the error of the circle's radius

With the parameters α and β, the center of the circle is localized within the hexagon as shown in Figure 5. It is computed as follows: suppose that a target C with a radius r is observed from the sensor, and the detection error of the right and left side of the target is α, as shown in Figure 5. First, a circle C_- with a radius $(r - \beta)$ is considered. The black region in Figure 5 indicates a possible region for the center location of the circle C_-, on condition that the right and left side of the circle C_- are projected within $\pm\alpha$ from those of the target C, respectively. Here, the straight lines l and m are parallel to l' and m', respectively, and the black region indicates only the upper half of the possible region for the circle C_-. In the same way, the dark gray region indicates a possible region for the center location of a circle C_+ with a radius $(r + \beta)$. Here, the straight lines l'' and m'' are parallel to l and m, respectively. Hence, the center of the circle whose radius is from $(r - \beta)$ to $(r + \beta)$ exists in the merged region of the black, the dark gray and the light gray regions (Figure 5 shows only the upper half of the region). This region indicates the location of the target C.

In the above method, target matchings can be verified by checking if the hexagons overlap each other. Then, in the first step of N-ocular stereo, the target is localized at the overlapped region of two hexagons as shown in Figure 6. In the same way, in the second step, the target is localized at the overlapped region of N hexagons. If let α and β smaller, the overlapped region also becomes smaller; and when it finally becomes a point, it can be considered as the location of the target.

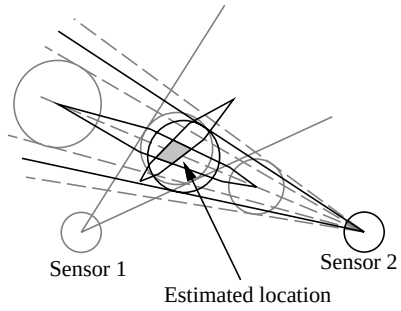


Figure 6. Localization of a target by binocular stereo

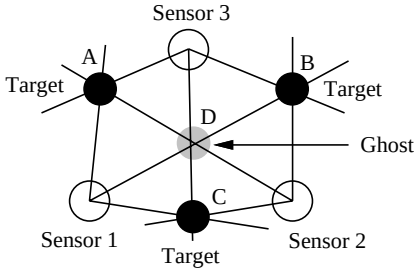


Figure 7. False matchings in N-ocular stereo

4.3. False matchings in N-ocular stereo

N-ocular stereo can solve the correspondence problems of multiple targets in most cases, however, it cannot solve the target correspondence with a particular arrangement of targets. In Figure 7, for example, it is estimated by N-ocular stereo that targets exist at up to four locations of A through D, including a false one (called a ghost target). In general, there is no way to eliminate the ghost target except to observe the motion of the intersections for a while [22].

The false matching in N-ocular stereo occurs when an azimuth angle of a target is associated with multiple locations (in Figure 7, an azimuth angle observed by the sensor 1 is used for the locations B and D). Therefore, if all of azimuth angles which are used for measuring a target location are also used for other locations, it is estimated that the location may be a false matching (the location D in the case of Figure 7).

In the implemented system, the false matchings are considered in the process of target tracking. In the process, each of the measured locations is related to the nearest one of previously measured locations, and the locations of false

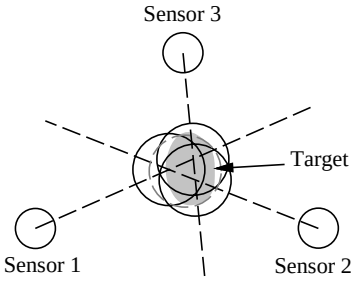


Figure 8. Simplified N-ocular stereo

matchings are checked after those of correct matchings.

5. Implementing N-ocular stereo

5.1. Simplified N-ocular stereo

In N-ocular stereo described in the previous section, the verification costs of overlapped regions of hexagons and that of convergent operations are very high, and it is difficult to perform real-time computation. Therefore, we have simplified N-ocular stereo as follows:

1. In the first step (binocular stereo), place a circle at the intersection of the azimuth angles detected by arbitrary two sensors, and consider the circle as the target location (see three black circles shown in Figure 8). Here, the radius of the circle is assumed to be 30cm since the targets are people.

2. In the second step (N-ocular stereo), check if the circles overlap each other to verify if the Nth sensor observes the target. If the circles overlap each other, place a new circle with a radius of 30cm at the center of gravity of the circles. It is considered as the target location measured with N sensors.

5.2. Error handling in the simplified N-ocular stereo

In the simplified N-ocular stereo, errors α and β described in Section 4.2 are handled as follows.

5.2.1 α: Detection errors of the right and left side of a target

As described in Section 3.2, binocular stereo using ODVSs has a low-precision problem with respect to targets locating along the baseline of the sensors [8]. In the simplified N-ocular stereo, this problem causes the following problems: Figure 9 (a), (b) and (c) show examples in the step of binocular stereo, where there is a target whereas no circle is placed since there is no intersection on account of observation errors of azimuth angles. Figure 9 (d) shows another example in the step of N-ocular stereo, where the

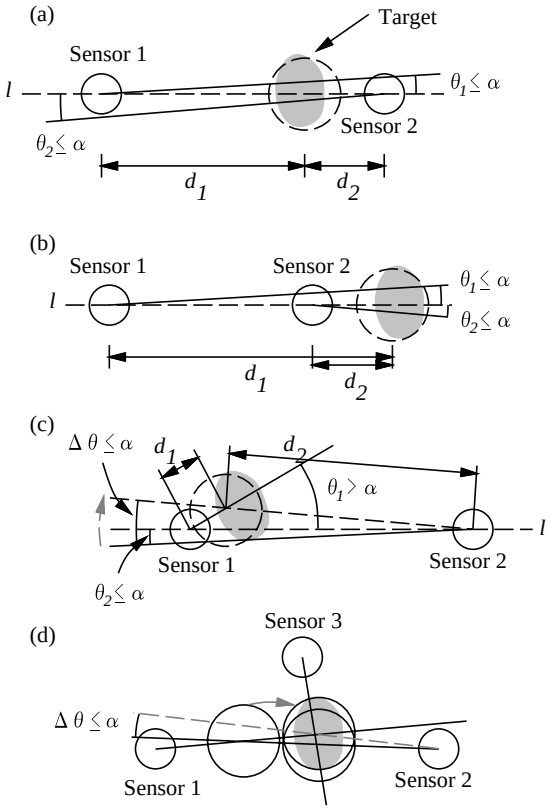


Figure 9. Error compensation

target cannot be localized since the circles which are placed in the step of $(N-1)$-ocular stereo do not overlap each other on account of observation errors.

Here, we introduce the following techniques to cope with the above problems.

When there is no intersections with respect to two lines: If the angle between the baseline l of the two sensors and each of azimuth angles detected by the sensors (let these be θ_1 and θ_2) are equal to or less than α (see Figure 9 (a) and (b)), consider that a target exists on the baseline l. Then, locate the target in such a way that the ratio of the distances between the target and each sensor (let this be $d_1 : d_2$) matches that of the apparent sizes of the target observed by the sensors. If one of the azimuth angles (let this be θ_2) is equal to or less than α, consider that a target exists on the line representing the other azimuth angle (θ_1). Then, correct the azimuth angle (θ_2) with $\Delta\theta$ ($\Delta\theta \leq \alpha$), and locate the target in such a way that the ratio of the distances $d_1 : d_2$ is close to that of the apparent sizes of the target.

When two circles do not overlap each other: If the circles overlap each other by correcting one of the azimuth angles with $\Delta\theta$ ($\Delta\theta \leq \alpha$), consider that they overlap

Figure 10. Overview of the real-time human tracking system

each other (see Figure 9 (d)).

5.2.2 β: An error of the human model

After the target is localized, the apparent size of the target reflected on each sensor can be computed from the distance between the sensor and the measured target location. If it differs by more than β from the actual size observed by the sensor, consider that the measured location is a false matching.

6. Experimentation

6.1. Hardware configuration

We have developed a real-time human tracking system (Figure 10), which measures people's locations based on N-ocular stereo and tracks them in real time. The system consists of four ODVSs, and omnidirectional images taken with the sensors are merged into one image with a quadrant image unit, then sent to a standard image capture card (Matrox Meteor, 640×480 pixels) on a PC (Pentium II 400MHz with 128MB memory). The four ODVSs are arranged in the center of a room (9m×7m) at a height of approximately 1m. In this system, the locations and the orientations of the sensors are measured before tracking.

The system detects targets in the omnidirectional images by background subtraction, since the sensors are fixed in the environment. The top of Figure 11 shows an unwrapped image, and the bottom graph shows the vertical sum of the difference at each pixel. The targets A, B and C are detected with a threshold shown with the broken line in Figure 11, which is determined by taking 10 frames. The centroid is regarded as the azimuth of the target.

6.2. Precision of N-ocular stereo

In this experimentation, the resolution of the omnidirectional image as shown in Figure 3 is approximately 400 pixels along the circumference of the image, and that of

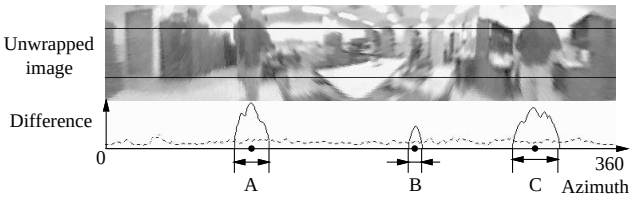


Figure 11. Detecting targets by background subtraction

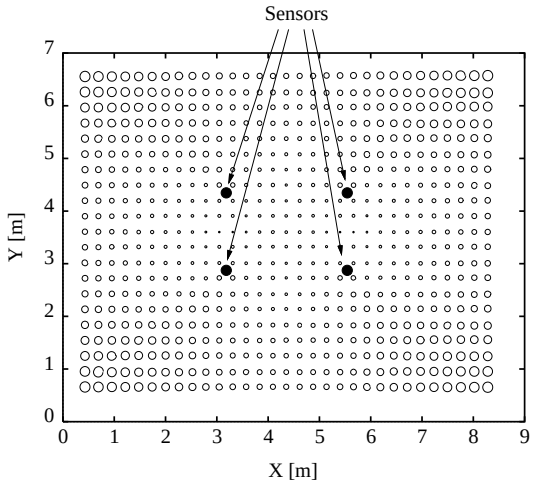


Figure 12. Uncertainty of binocular stereo

the unwrapped image as shown in Figure 11 is 640 pixels. Hence, the ODVS has approximately a resolution of 1 degree. Figure 12 shows uncertainty of binocular stereo with a resolution of 1 degree. The arrangement of sensors is same through the experimentation in this section. Each circle indicates the error range of binocular stereo at that location using two sensors which give the best precision (the diameter of the circles in Figure 12 is twice of the actual error range). In Figure 12, the minimum error range is approximately 0.7cm and the maximum is approximately 5cm. We can find that multiple ODVSs provide fine precision in a wide area.

Figure 13 shows the error range of target locations measured by the system. Here, we have used a white cylinder with a diameter of 30cm as a target, and placed it at precisely measured marks on the floor. The circles A through N in Figure 13 indicate the locations of the marks and '+' indicates cylinder locations measured by the system over 100 frames. Thus, the distribution of the measured locations (Figure 13) is analogous to that of the uncertainty of binocular stereo (Figure 12). Table 1 shows averages and errors (distances between measured and actual target locations) of the measured locations. The maximum error is 0.17m at the location A.

In Figure 13, we can find that the target locations are

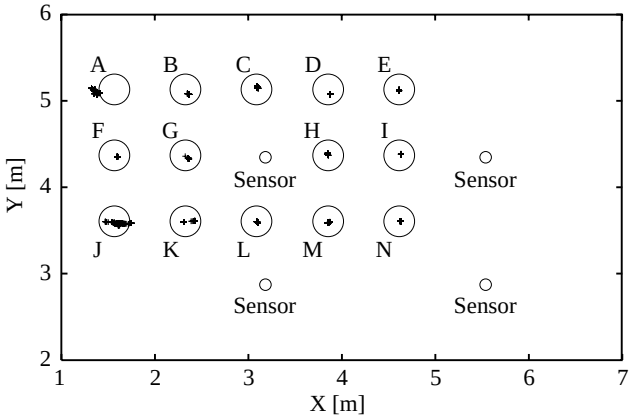


Figure 13. Precision of localization

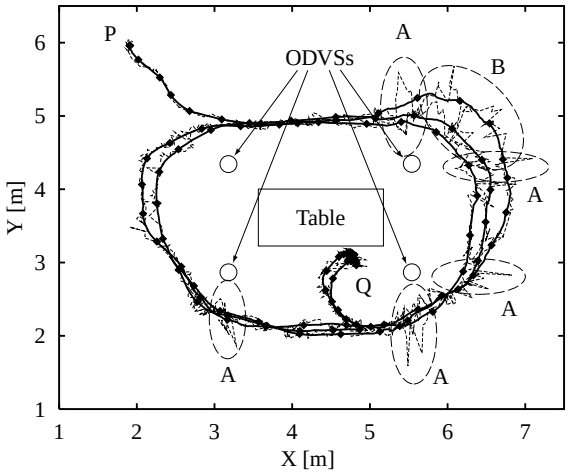


Figure 14. Trajectories of a walking person

Table 1. Averages and errors of the measured locations

Loc	Average [m]	Err [m]	Loc	Average [m]	Err [m]
A	(1.35, 5.11)	0.170	H	(3.85, 4.38)	0.043
B	(2.37, 5.07)	0.098	I	(4.63, 4.38)	0.061
C	(3.10, 5.15)	0.056	J	(1.61, 3.58)	0.092
D	(3.88, 5.08)	0.086	K	(2.40, 3.61)	0.114
E	(4.61, 5.12)	0.041	L	(3.10, 3.59)	0.053
F	(1.60, 4.35)	0.077	M	(3.85, 3.59)	0.047
G	(2.36, 4.34)	0.077	N	(4.63, 3.61)	0.056

measured within 5cm error if the target locates within 3m from three sensors (in N-ocular stereo, at least three sensors need to simultaneously observe the same target for measuring its location). However, the precision depends on the number of sensors, the arrangement of the sensors, the precision of background subtraction, and so on.

6.3. Tracking people

Figure 14 shows trajectories of a walking person for one minute, with the same arrangement of sensors as the experimentation in Section 6.2. The solid lines show the trajectories, and dots on the lines show the person's locations at intervals of 1/2 second. As shown in Figure 14, the system could track the person without loosing sight. In this experimentation, the person's location measured by N-ocular stereo is smoothed during 1/2 second, so that there is a delay of about 1/4 second.

The broken lines in Figure 14 indicate the person's locations at every frame before smoothing. There are large errors around A and B. This is because (1) binocular stereo using ODVSs has a low-precision problem with respect to targets locating along the baseline of the sensors (this corresponds A in Figure 14), and (2) the result of background subtraction becomes noisy if the color of person's clothes is similar to that of the background (this corresponds B in Figure 14). In the latter case, general noise filtering techniques such as the Kalman filter may not be able to successfully eliminate the noise, since the noise is different from white noise. It is effective to add additional sensors to cope with this kind of noise.

In this implementation, the system could simultaneously track three people at video rate (30 fps). The experimental results show that the system completely tracked one person, and correctly tracked two persons for 99% of the time, and three persons for 89% of the time. Tracking errors occured in the following cases:

- When two or more people moved closer with each other, the system recognized them as one person in background subtraction, or could not correctly identify them in the tracking phase.
- When a person moved behind another person, the system could not measure its location.

In order to solve the former error, a more sophisticated method is needed for detecting people. The latter error will be solved by adding ODVSs into the environment.

7. Conclusion

In this paper, we have proposed N-ocular stereo for verifying the correspondence among multiple targets and measuring their locations, using multiple omnidirectional vision sensors (ODVSs). In addition, several methods have been developed for compensating observation errors, in order to cope with the precision problem in omnidirectional stereo. We have developed a real-time human tracking system with four compact ODVSs we have originally designed, and shown that the system can robustly track people in real time only with visual information.

The developed system can show live images of tracked people as shown in Figure 15 as well as record the trajectories of them. The images are taken with the ODVSs,

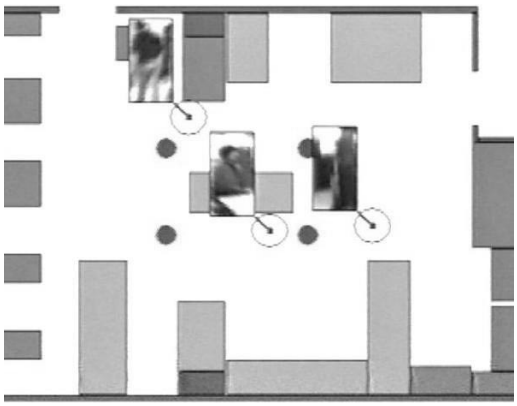


Figure 15. Showing people's locations and their images

and zoomed in and out according to the distance between the people and the sensor. The side views of the people also enable the system to identify the people with the colors of their clothes, to recognize their behaviors by observing motions of their head and arms, and so on. In addition, more robust recognition will be achieved by using redundant visual information taken from various points of view with multiple ODVSs.

There are various application systems which take the above advantages. For example, the system can record people's trajectories and their behaviors, so that it is useful for a monitoring system. It is also useful for analyzing people's behaviors, since it can automatically collect an enormous amount of people's trajectories. Furthermore, in gesture recognition [25], information of people's locations and trajectories will be useful for correct gesture recognition. Thus, we believe this practical system opens up new application areas of computer vision.

References

[1] J. Boyd, E. Hunter, P. Kelly, L.-C. Tai, C. Phillips, and R. Jain. MPI-Video infrastructure for dynamic environments. In *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, pages 249–254, 1998.

[2] R. Collins, Y. Tsin, J. Miller, and A. Lipton. Using a DEM to determine geospatial object trajectories. In *DARPA Image Understanding Workshop*, 1998.

[3] W. Grimson, C. Stauffer, R. Romano, and L. Lee. Using adaptive tracking to classify and monitor activities in a site. In *Proc. CVPR*, pages 22–29, 1998.

[4] E. Gurewitz, I. Dinstein, and B. Sarusi. More on the benefit of a third eye. In *Proc. ICPR*, pages 966–968, 1986.

[5] J. Hong, X. Tan, B. Pinette, R. Weiss, and E. Riseman. Image-based homing. In *Proc. IEEE Int. Conf. Robotics and Automation*, pages 620–625, 1991.

[6] S. Intille, W. Davis, and A. Bobick. Real-time closed-world tracking. In *Proc. CVPR*, pages 697–703, 1997.

[7] H. Ishiguro. Development of low-cost compact omnidirectional vision sensors and their applications. In *Proc. Int. Conf. Information Systems, Analysis and Synthesis*, pages 433–439, 1998.

[8] H. Ishiguro, M. Yamamoto, and S. Tsuji. Omni-directional stereo. *IEEE Trans. PAMI*, 14(2):257–262, 1992.

[9] N. Johnson and D. Hogg. Learning the distribution of object trajectories for event recognition. *Image and Vision Computing*, 14(8):609–615, 1996.

[10] T. Kanade, R. T. Collins, A. J. Lipton, P. Anandan, P. Burt, and L. Wixson. Cooperative multi-sensor video surveillance. In *DARPA Image Understanding Workshop*, volume 1, pages 3–10, 1997.

[11] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka. A stereo machine for video-rate dense depth mapping and its new applications. In *Proc. CVPR*, pages 196–202, 1996.

[12] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russell. Towards robust automatic traffic scene analysis in real-time. In *Proc. ICPR*, volume 1, pages 126–131, 1994.

[13] A. Lipton, H. Fujiyoshi, and R. Patil. Moving target classification and tracking from real-time video. In *DARPA Image Understanding Workshop*, pages 115–122, 1998.

[14] T. Mori, Y. Kamisuwa, H. Mizoguchi, and T. Sato. Action recognition system based on human finder and human tracker. In *Proc. IROS*, pages 1334–1341, 1997.

[15] S. Nayar and S. Baker. Catadioptiric image formation. In *DARPA Image Understanding Workshop*, pages 1431–1437, 1997.

[16] K. Ng, H. Ishiguro, M. Trivedi, and T. Sogo. Monitoring dynamically changing environments by ubiquitous vision system. In *Second IEEE Workshop on Visual Surveillance*, pages 67–73, 1999.

[17] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Trans. PAMI*, 15(4):353–363, 1993.

[18] K. Pattipati, S. Deb, Y. Bar-Shalom, and R. Washburn. A new relaxation algorithm and passive sensor data association. *IEEE Trans. Automatic Control*, 37(2):198–213, 1992.

[19] D. Rees. Panoramic television viewing system. United States Patent No. 3,505,465, 1970.

[20] C. Sastry, E. Kamen, and M. Simaan. An efficient algorithm for tracking the angles of arrival of moving targets. *IEEE Trans. Signal Processing*, 39(1):242–246, 1991.

[21] J. Segen and S. Pingali. A camera-based system for tracking people in real time. In *Proc. ICPR*, volume 3, pages 63–67, 1996.

[22] S. Shams. Neural network optimization for multi-target multi-sensor passive tracking. *Proc. IEEE*, 84(10):1442–1457, 1996.

[23] M. Yachida. 3-D data acquisition by multiple views. In *3rd International Symposium on Robotics Research (ISRR'85)*, pages 11–18, London, 1986. The MIT Press.

[24] Y. Yagi and S. Kawato. Panorama scene analysis with conic projection. In *Proc. IROS*, pages 181–187, 1990.

[25] J. Yamamoto, J. Ohya, and K. Ishii. Recognizing human action in time-seqential images using hidden markov models. In *Proc. CVPR*, pages 379–387, 1992.

[26] K. Yamazawa, Y. Yagi, and M. Yachida. Omnidirectional imaging with hyperboloidal projection. In *Proc. IROS*, pages 1029–1034, 1993.