2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

A Fast Multi-camera Tracking System with Heterogeneous Lenses

Xiaorong Zhao, Qingyi Gu, Tadayoshi Aoyama, Takeshi Takaki, and Idaku Ishii

***Abstract*— We have developed a fast target tracking system that utilizes four cameras with lenses of different focal lengths to track an object without blurring images, even when the object moves in the depth direction away from the cameras in three-dimensional (3-D) space. This system can maintain a well-focused camera view by switching among the four input images, instead of using lens motor control. The multi-camera system was mounted on a two-axis mechanical active vision platform. The active vision control and camera-view switching are executed by processing color 512 × 512 images from the four camera inputs at 500 fps in real time on a high-speed vision platform. The performance of our system was verified by its tracking results for objects moving rapidly in 3-D space.**

I. Introduction

Recently, real-time high-frame-rate (HFR) vision systems have been developed that can simultaneously process video at higher frame rates than conventional video cameras (e.g., NTSC 30 fps) [1]–[3]. Such HFR vision systems enable fast target tracking at hundreds of frames per second in real time for recognition of high-speed moving objects. Its effectiveness has already been demonstrated in many applications such as robot manipulation [4], virtual stillness for beating heart surgery [5], and microbe tracking [6]. Several tracking systems with pattern recognition have been developed by implementing image recognition algorithms with parallel logic [7]–[9]. Fast stereo tracking systems using multiple HFR cameras to track an object moving in three-dimensional (3-D) space have also been reported [10], [11]. Most stereo tracking systems have used lenses with focal lengths and apertures that are fixed. These systems have a weakness when tracking an object moving in the depth direction, because the depth of field (DOF), which depends on the focal length and aperture of the lens, is also constrained.

Autofocus (AF) technology that can automatically control a motor to position a lens for well-focused images is widely used in consumer digital cameras, and many AF algorithms have been proposed [12]. In most AF systems, the focusing speeds are low because the response times of the motors that physically move the variable-focus lenses are long. To shorten the response time of a variable lens for faster imaging, several liquid-filled active lenses have been developed [13], [14]. However, AF systems that use a single camera with a variable lens have a limited DOF, because it is difficult for a single variable lens to realize a wide DOF with high image intensity. Furthermore, motor mechanisms should be integrated with the AF system to control the variable lens, but their sizes and weights impose constraints when the AF system is mounted on a movable platform such as a manipulator.

In this study, we develop a multi-camera tracking system that can track a moving object in 3-D space at a high frame rate without image blurring. Our system has four cameras with lenses of different focal lengths to cover a wide DOF for 3-D tracking. This can function as an AF system with no lens motor control by selecting well-focused images from the four cameras. Section III provides an outline of our multi-camera tracking system. Section III-B describes a 3-D tracking algorithm implemented by our system. This algorithm can be executed for 512×512 images from four cameras at 500 fps in real time. Section IV verifies the real-time 3-D tracking with several experimental results.

II. Concept

In our concept of multi-camera tracking, a well-focused camera view is selected from the views of multiple cameras with heterogeneous lenses. This minimizes the measurement error caused by lens blur and image quantization when an object is mechanically tracked in a selected camera view without protruding from the image sensor area.

Figure 1 illustrates our concept of multi-camera tracking with heterogeneous lenses when camera DOFs are layered in a mechanical tracking system. The optical center of the lens of the i-th camera $(i = 1, \ldots, I)$ lies on the plane $z = 0$, and the optical axis is parallel to the z-axis. The (x, y, z) coordinate system is defined as the world coordinate system, and its z-coordinate represents the depth from the cameras. The lens of the i-th camera can be described with a thin lens model as shown in Figure 2, where ${}^{i}f$ is the focal length, ${}^{i}F$ is the f-number, ${}^{i}a$ is the distance from the object to the lens, and ${}^{i}b$ is the distance from the lens to the image sensor.

A. Measurement error

First, we consider the image degradation when an object is observed with the i-th camera. Depending on the location

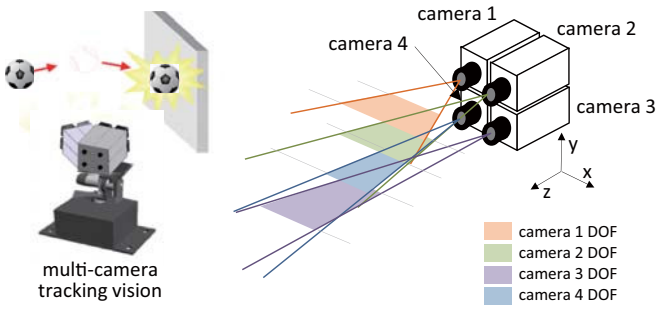


Fig. 1. Multi-camera tracking with heterogeneous lenses.

978-1-4673-6357-0/13/$31.00 ©2013 IEEE

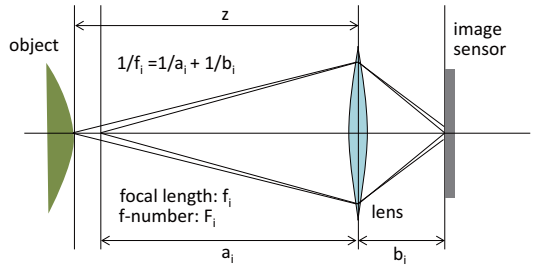


Fig. 2. Thin lens model.

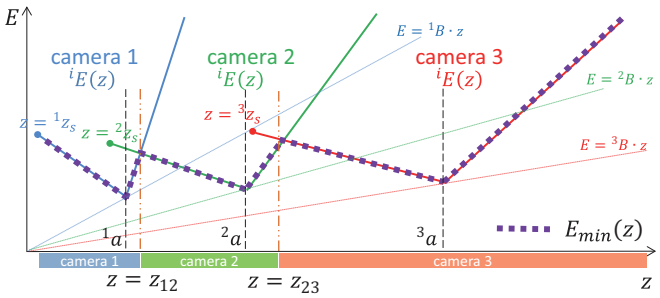


Fig. 3. Relationship between measurement error and object location.

of the object in the z-direction, the measurement error by the image sensor of the i-th camera can be described as follows:

$$ {}^{i}\varepsilon(z) = \mathrm{sgn}(z - {}^{i}a)\frac{{}^{i}f^2}{{}^{i}F}\left(\frac{1}{{}^{i}a} - \frac{1}{z}\right) + \frac{2}{\sqrt{\pi}}\delta \tag{1} $$

where the first term is the diameter of the circle of confusion for the i-th camera, which indicates the lens blur error, and the second term is the quantization error for an image sensor of pixel pitch δ. By multiplying ${}^{i}\varepsilon(z)$ with the ratio

$$ {}^{i}Z(z) = \frac{z}{{}^{i}b} = z\left(\frac{1}{{}^{i}f} - \frac{1}{{}^{i}a}\right), \tag{2} $$

the measurement error ${}^{i}E(z)$ of the object can be determined for the i-th camera as follows:

$$ {}^{i}E(z) = {}^{i}\varepsilon(z){}^{i}Z(z) = \mathrm{sgn}(z - {}^{i}a)\cdot {}^{i}A(z - {}^{i}a) + {}^{i}B, \tag{3} $$

where ${}^{i}A$ and ${}^{i}B$ are constants given by

$$ {}^{i}A = \left(\frac{1}{{}^{i}f} - \frac{1}{{}^{i}a}\right)\frac{{}^{i}f^2}{{}^{i}F\cdot{}^{i}a}, \quad {}^{i}B = \left(\frac{1}{{}^{i}f} - \frac{1}{{}^{i}a}\right)\frac{2}{\sqrt{\pi}}\delta. \tag{4} $$

As expressed in Eq. (3), the measurement error ${}^{i}E(z)$ describes a polygonal line that consists of two line segments whose vertex is located at $z = {}^{i}a$.

B. Tracking condition without any protrusion

To track a whole object with no protrusion from the image, the measurable range in the z-direction must satisfy the following condition when an object of width L is observed with the i-th camera:

$$ z > \frac{2L}{\sqrt{\pi}N\cdot{}^{i}B} (= {}^{i}z_s), \tag{5} $$

where N is the number of pixels the image sensor has in the width direction.

C. Selection of a well-focused camera

Under the condition of Eq. (5), the number $i_o(z)$ of the well-focused camera is determined as follows by searching for the camera whose measurement error is the smallest:

$$ i_o(z) = \arg\min_{i} {}^{i}E(z), \quad E_{min}(z) = {}^{i_o(z)}E(z), \tag{6} $$

where $E_{min}(z)$ is the measurement error of the well-focused camera. Figure 3 shows the relationship between the measurement error and the z-coordinate of the object. Here, $i_o(z)$ and $E_{min}(z)$ are given as functions of the z-coordinate. Considering that the measurement error must be less than a certain value E_0 for accurate observation, the measurable range in the z-direction where $E_{min}(z) \le E_0$ holds becomes much wider than the range of z where ${}^{i}E(z) \le E_0$ holds for the i-th camera. Thus, compared with single-camera tracking, multi-camera tracking with heterogeneous lenses covers a wider DOF by selecting a well-focused camera depending on the z-coordinate of a tracked object.

D. Advantages and disadvantages

In comparison with conventional AF systems that use motor-controlled variable-focus lenses, our concept has two advantages for fast 3-D tracking: (1) it can expand the DOF for 3-D tracking by increasing the number of cameras, and (2) it can emulate a fast AF function without any slow-response motor control. However, our concept also has two disadvantages for 3-D tracking: (1) a multi-camera system becomes too large and heavy for mounting on a movable platform when the number of cameras becomes too large, and (2) the tracked object is divergently observed when switching between selected camera views, because the multiple cameras are located at different positions owing to their physical sizes. In this study, we resolve these disadvantages by introducing a mechanical active vision platform with a lightweight multi-camera head.

III. Fast Multi-camera Tracking System

A. System configuration

Based on our concept, we developed a fast multi-camera tracking system. Figure 4 shows its configuration, which consists of a two-axis active vision platform, four camera heads with lenses of different focal lengths, two FPGA-based image processing boards (IDP Express board), and a PC.

The camera heads and IDP Express boards used in our system were developed for real-time high-speed image processing [11]. The camera head was compactly designed for mounting on movable objects. Its dimensions and mass are 23 × 23 × 77 mm and 145 g, respectively, when no lens is mounted. The camera head captures 8-bit RGB color 512×512 images with the Bayer filter of its image sensor and transfers images to the IDP Express board at 2000 fps. The IDP Express board has two camera inputs and a user-specific FPGA (Xilinx XC3S5000) for hardware implementation of high-speed image processing algorithms. The 512×512 input

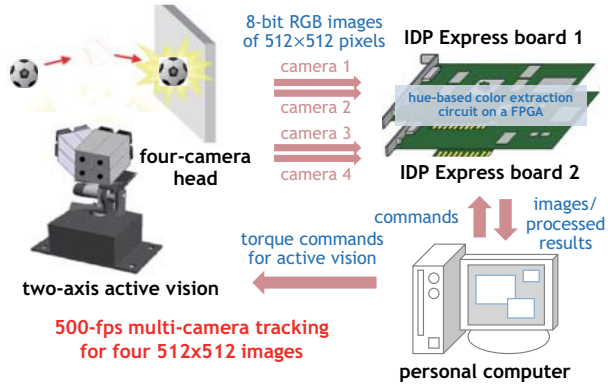


Fig. 4. System configuration.

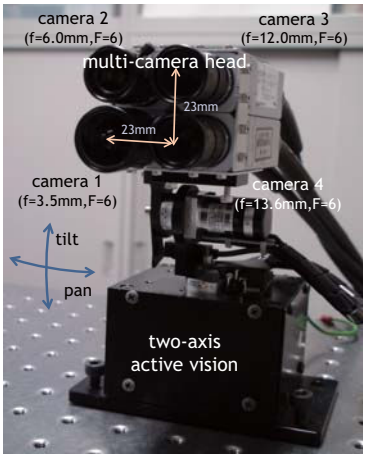


Fig. 5. A multi-camera head mounted on a two-axis active vision platform.

images and processing results from two camera heads are transferred via a PCIe 2.0 bus from the IDP Express board to PC memory at 2000 fps. For acceleration of the multi-camera tracking process, we implemented a hue-based color extraction circuit in the hardware of the user-specific FPGA on the IDP Express board. This circuit extracts specified hue regions for two 512×512 color images and calculates the centroid positions in real time at 2000 fps. In our system, two IDP Express boards were used for the four camera heads, and these two IDP Express boards were mounted on a single PC. Through various API functions associated with board control and access to memory-mapped data, arbitrary processes can be programmed as software for real-time execution on the PC. In this study, we used a PC with the following specifications: ASUSTeK P5Q-E mainboard, Intel Core 2 Quad 9505 2.83 GHz CPU, 4 GB memory, Windows 7 (32-bit), and two 16-lane PCIe 2.0 buses.

The four camera heads were mounted on the active vision platform as a 2×2 multi-camera head. Figure 5 shows an overview of the multi-camera head and the active vision platform. The camera heads are arranged in a 2×2 array with an interval of 23 mm, corresponding to the height and width of the camera head. The camera heads were assigned numbers from one to four, proceeding clockwise from the lower left. On the camera heads were mounted Sony

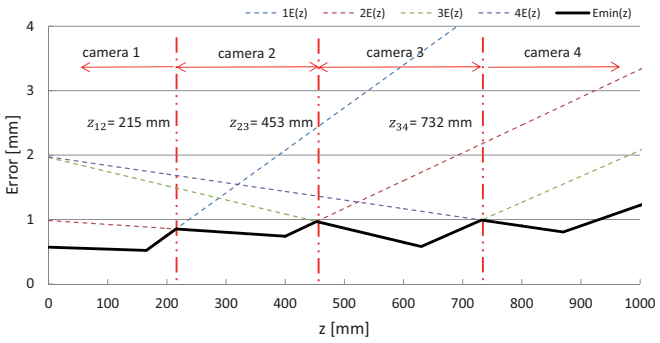


Fig. 6. Relationship between measurement errors and z-coordinate of object for our multi-camera tracking system.

NF-mount lenses of different focal lengths: VCL-03S12XM for Camera 1, VCL-06S12XM for Camera 2, and VCL-03S12XM for Cameras 3 and 4. The dimensions of the multi-camera head are $46 \times 46 \times 77$ mm, and its mass is 695 g, including the masses of the lenses mounted on the camera heads. For the i-th camera $(i = 1, \ldots, 4)$, the f-number was set as ${}^{i}F = 6$. The focal lengths were set as ${}^{1}f = 3.5$ mm, ${}^{2}f = 6.0$ mm, and ${}^{3}f = {}^{4}f = 12.0$ mm. The distances from the object to the lens were set as ${}^{1}a = 165$ mm, ${}^{2}a = 400$ mm, ${}^{3}a = 630$ mm, and ${}^{4}a = 870$ mm. For the multi-camera head of our system, Figure 6 shows the relationship between the measurement errors and the z-coordinate of the object. The number i_o of a well-focused camera is determined as follows:

$$i_o = \begin{cases} 1 & (z \le z_{12}) \\ 2 & (z_{12} < z \le z_{23}) \\ 3 & (z_{23} < z \le z_{34}) \\ 4 & (z_{34} > z) \end{cases}, \tag{7}$$

where the z-coordinates at which to switch camera views are $z_{12} = 215$ mm, $z_{23} = 453$ mm, and $z_{34} = 732$ mm.

The two-axis active vision platform is moved by pan and tilt motors, which are compact high-speed RSF-5A-50 motors (Harmonic Drive Systems). The dimensions of the active vision platform are $12 \times 12 \times 7$ cm without the multi-camera head. The orientation of the multi-camera head was controlled using the visual information from the camera heads in order to track a moving object in 3-D space.

B. The implemented algorithm

For fast tracking of a colored object moving in 3-D space, our system implements a multi-camera tracking algorithm, which has the six steps described below. Figure 7 shows the flowchart of the algorithm implemented.

(1) Hue-based color region extraction

To extract a specified color region from an image as an object to be tracked, each color input image from the i-th camera $(i = 1, \cdots, 4)$ is assigned thresholds in the HSV color space [15] as follows:

$${}^{i}B(X,Y) = \begin{cases} 1 & \left(|{}^{i}H - H_0| < \theta_H, {}^{i}S > \theta_S, {}^{i}V > \theta_V\right) \\ 0 & (\text{otherwise}) \end{cases}. \tag{8}$$

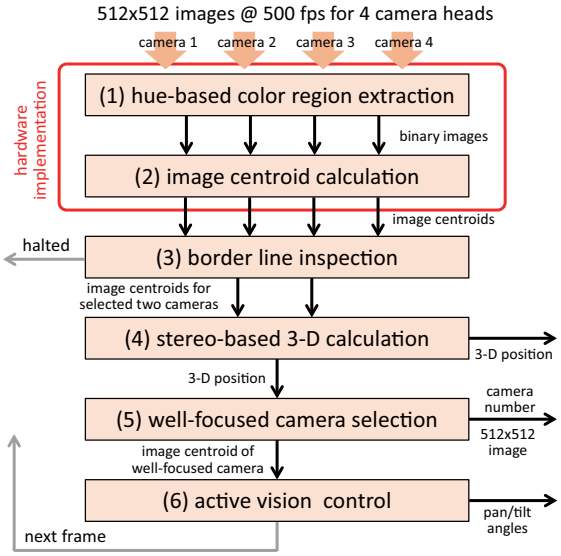


Fig. 7. Flowchart of multi-camera tracking algorithm.

where ${}^iH(X,Y)$, ${}^iS(X,Y)$, and ${}^iV(X,Y)$ are the hue, saturation, and value components of the input image of the i-th camera; H_0, θ_H, θ_S, and θ_V are parameters to specify a vivid color that corresponds to the object to be tracked. The observational ranges of hue, saturation, and value are set to $0° \le {}^iH < 360°$, $0 \le {}^iS \le 1$, and $0 \le {}^iV \le 255$, respectively. The HSV color images are converted into 8-bit images after RGB conversion of the input images with a Bayer filter.

(2) Image centroid calculation

For the i-th camera, the zeroth-order and first-order moment features of ${}^iB(X,Y)$ are computed as follows:

$$ {}^iM_0 = \sum_{X,Y} {}^iB(X,Y), \tag{9} $$

$$ {}^iM_X = \sum_{X,Y} X\,{}^iB(X,Y),\ {}^iM_Y = \sum_{X,Y} Y\,{}^iB(X,Y). \tag{10} $$

The position of a tracked object in the input image of the i-th camera is calculated as the following image centroid:

$$ {}^i\boldsymbol{c} = ({}^ic_X, {}^ic_Y) = \left(\frac{{}^iM_X}{{}^iM_0}, \frac{{}^iM_Y}{{}^iM_0} \right). \tag{11} $$

(3) Borderline inspection

On the borderlines of a 512×512 image, the number of pixels where ${}^iB(X,Y) = 1$ is counted for the i-th camera as follows:

$$ {}^in_b = \sum_{(X,Y)\in X_b} {}^iB(X,Y), \tag{12} $$

where $(X,Y) \in X_b$ indicates that (X,Y) is located on $X = 0$, $X = 511$, $Y = 0$, or $Y = 511$. If N_b is defined as the set of the numbers of the camera heads for which no pixel of the specified color lies on the borderlines of the images, then Camera i_s and Camera $i_s - 1$ are selected as a pair of camera heads for a stereo-based 3-D calculation, as follows:

$$ i_s = \max_{i\in N_b} i, \quad N_b = \{\ i : {}^in_b = 0\ \}. \tag{13} $$

The tracking process is halted when $N_b = \emptyset$ or $i_s = 1$ but is restarted when the object to be tracked is observed with two or more camera heads.

(4) Stereo-based 3-D calculation

The 3-D position of the tracked object is calculated by triangulation as follows:

$$ (x,y,z) = \mathrm{Stereo}_{i_s i_s-1}({}^{i_s}\boldsymbol{c}, {}^{i_s-1}\boldsymbol{c}). \tag{14} $$

where $\mathrm{Stereo}_{ij}(\cdot)$ is a triangulation function to calculate the 3-D position of the object from corresponding points in a stereo pair of Camera i and Camera j. All the camera parameters are initially obtained from a prior calibration.

(5) Well-focused camera selection

Depending on the z-coordinate, the number i_o of a well-focused camera is selected in accordance with Eq. (7).

(6) Active vision control

The pan and tilt motors of the two-axis active vision platform are controlled to minimize the following evaluation function:

$$ F = \sum_{i=1}^{4} w_i \|{}^i\boldsymbol{c} - \boldsymbol{x}\|^2, \tag{15} $$

where F is the weighted sum of the squared errors between the image centroids and the center pixels $(256, 256)$ of the 512×512 images from the four camera heads. The weight w_i $(i = 1, \cdots, 4)$ is determined using the z-coordinate of the tracked object as follows:

$$ (w_1,w_2,w_3,w_4) = \begin{cases} (1,0,0,0) & (z \le {}^1D) \\ \left(\frac{{}^2D-z}{{}^2D-{}^1D}, \frac{z-{}^1D}{{}^2D-{}^1D}, 0, 0\right) & ({}^1D < z \le {}^2D) \\ \left(0, \frac{{}^3D-z}{{}^3D-{}^2D}, \frac{z-{}^2D}{{}^3D-{}^2D}, 0\right) & ({}^2D < z \le {}^3D) \\ \left(0, 0, \frac{{}^4D-z}{{}^4D-{}^3D}, \frac{z-{}^3D}{{}^4D-{}^3D}\right) & ({}^3D < z \le {}^4D) \\ (0,0,0,1) & ({}^4D < z) \end{cases}, \tag{16} $$

where the ${}^iD$ $(i = 1, \cdots, 4)$ are determined by using z_{12}, z_{23}, and z_{34} in the formulas

$$ {}^1D = (3z_{12} - z_{23})/2 = 96 \text{ mm}, \tag{17} $$

$$ {}^2D = (z_{12} + z_{23})/2 = 334 \text{ mm}, \tag{18} $$

$$ {}^3D = (z_{23} + z_{34})/2 = 592.5 \text{ mm}, \tag{19} $$

$$ {}^4D = (-z_{23} + 3z_{34})/2 = 871.5 \text{ mm}. \tag{20} $$

In this study, Steps 1 and 2 were implemented by parallel hardware logic for 8-bit color images on the user-specific FPGA of the IDP Express, and Steps 3–6 were implemented by software on the PC. We have confirmed that all six steps can be executed for 512×512 images in real time at 500 fps, and so the motors of the active vision platform were controlled using visual feedback at 500 Hz.

IV. Experiments

We present tracking results for an object moved by a human hand. Figure 8 shows the experimental setup. In the experiments, an artificial red rose flower of $6 \times 6 \times 7$ mm size was moved with respect to a color-patterned background. To extract the red region, the parameters for binarization were set to θ_S = 0.30, θ_V = 76, θ_H = 16.9, and H_0 = 61.8.

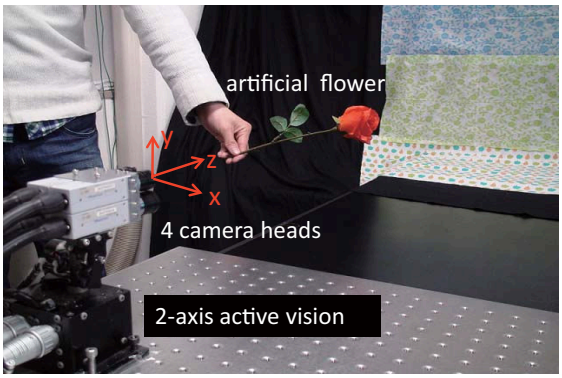
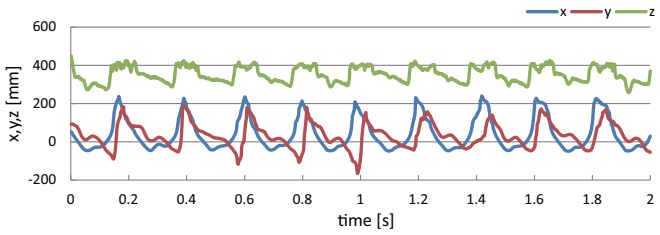
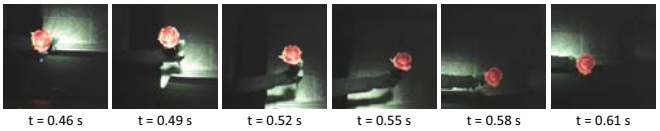


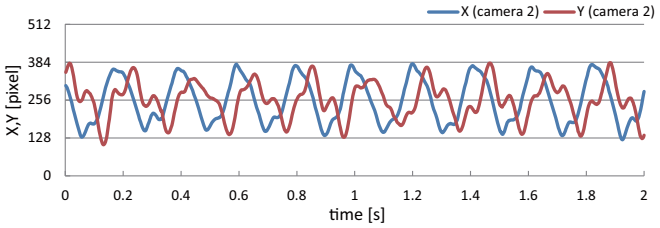
Fig. 8. Experimental setup.



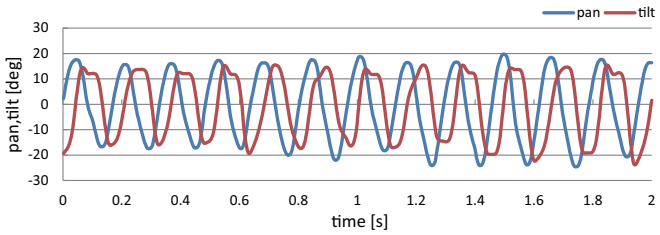
(a) 3-D position



(b) images from Camera 2



(c) image centroid of Camera 2



(d) pan and tilt angles of active vision platform

Fig. 9. Results obtained when a flower was rotated at fixed depth.

A. *Rotation at fixed depth*

To demonstrate the tracking performance of our multi-camera system with the four camera heads mounted, we conducted a tracking experiment in which an artificial flower was manually rotated 35 cm in front of the multi-camera head at five rotations per second.

Figure 9(a) shows the observed (x, y, z) coordinates of the tracked object in the camera coordinate system over an interval of 2 s starting at $t = 0$. The z-coordinate varied slightly around 350 mm within a range of $z_{12} = 215$ mm and $z_{23} = 453$ mm, so Camera 2 was always selected as the well-focused camera among the four camera heads. Figure 9(b) shows the input sequence of six images captured by Camera 2 at intervals of 0.03 s for $t = 0.46$–0.61 s. Figure 9(c) shows the (X, Y) coordinates of the image centroid of Camera 2 over the interval of 2 s starting at $t = 0$. It can be observed in Figures 9(b) and 9(c) that the image centroid of Camera 2 was always in the camera view to be controlled near the center pixel $(256, 256)$, but there were deviations caused by the limitations in the abilities of the motors of the active vision platform. Owing on these deviations, the (x, y) coordinates of the tracked object in the camera coordinate system were also deviated in Figure 9 (a). Figure 9(d) shows the pan and tilt angles of the active vision platform over the interval of 2 s starting at $t = 0$. It can be observed that the pan and tilt angles varied periodically at five cycles per second, corresponding to the rotation of the artificial flower.

B. *Back-and-forth motion in the depth direction*

To demonstrate the camera selection performance of our multi-camera system, we conducted a tracking experiment in which an artificial flower was manually moved back and forth between 20 cm and 120 cm in front of the multi-camera head at two cycles per second.

Figure 10(a) shows the (x, y, z) coordinates of the tracked object and the number of the well-focused camera over an interval of 2 s starting at $t = 0$. It can be observed that the z-coordinate varied periodically at 2 cycles per second, whereas the (x, y) coordinates were relatively constant. The number of well-focused camera was selected according to the z-coordinate and became smaller when the tracked object became nearer to the camera head. Figure 10(b) shows the input sequences of six images captured by the four camera heads at intervals of 0.03 s for $t = 1.59$–1.74 s. The images with red borders were selected as the well-focused camera views without any protrusion of the artificial flower from the image. Figure 10(b) also shows the size-normalized images that were generated by resizing and cropping the selected well-focused images so that the artificial flower was located at the center of each image and was adjusted in size to 35% of the 512×512 pixels. Thus, the tracked object was always located at the center of the size-normalized image. Figure 10(c) shows the (X, Y) coordinates of the image centroids of the four camera heads over the interval of 2 s starting at $t = 0$. Here the thick line indicates the image centroid of the well-focused camera. It can be observed in Figures 10(b) and 10(c) that the image centroid of the well-focused camera was near the center pixel $(256, 256)$, but there were deviations when the well-focused camera was switched according to the z-coordinate of the tracked object. Figure 10(d) shows the pan and tilt angles of the active vision platform over the interval of 2 s starting at $t = 0$. The

pan and tilt angles changed at the moment that the well-focused camera was switched to reflect the motion of the artficial flower in the depth direction, but otherwise these angles varied little. This is because the pan and tilt angles were controlled to reduce the deviation of the image centroid when the well-focused camera was switched.

The experimental results show that our system can correctly track an object in a well-focused image, even when the object moves rapidly in 3-D space. Our system achieves this by processing the 512×512 images from the four camera heads at 500 fps for active vision control and selection of the well-focused camera.

V. Conclusion

We have developed a fast multi-camera tracking system using four cameras with heterogeneous lenses for mechanically tracking an object moving in 3-D space. By switching among the four cameras to maintain a well-focused camera view, this system can track an object in 512×512 images at 500 fps without blurring. Its effectiveness was verified by its tracking results for objects moving rapidly in 3-D space. On the basis of these results, we will improve our system for high-speed 3-D robot control in the real world by expanding the image feature extraction for heterogeneous stereo cameras.

References

[1] Y. Watanabe, T. Komura, and M. Ishikawa, "955-fps real-time shape measurement of a moving/deforming object using high-speed vision for numerous-point analysis," *Proc. IEEE Int. Conf. Robot. Automat.*, 3192–3197, 2007.

[2] S. Hirai, M. Zakoji, A. Masubuchi, and T. Tsuboi, "Realtime FPGA-based vision system," *J. Robot. Mechat.*, **17**, 401–409, 2005.

[3] I. Ishii, T. Taniguchi, R. Sukenobe, and K. Yamamoto, "Development of high-speed and real-time vision platform, H^3 vision," *Proc. IEEE/RSJ Int. Conf. Intelli. Rob. Sys.*, 3671–3678, 2009.

[4] A. Namiki, Y. Imai, and M. Ishikawa, "Development of a high-speed multifingered hand system and its application to catching," *Proc. IEEE/RSJ Int. Conf. Intelli. Rob. Sys.*, 2666–2671, 2003.

[5] Y. Nakamura, K. Kishi, and H. Kawakami, "Heartbeat synchronization for robotic cardiac surgery," *Proc. IEEE Int. Conf. Robot. Automat.*, 2014–2019, 2001.

[6] H. Oku, I. Ishii, and M. Ishikawa, "Tracking a protozoon using high-speed visual feedback," *Proc. Annu. Int. IEEE-EMBS Special Topic Conf. Microtechnologies in Medicine and Biology*, 2014–2019, 2001.

[7] I. Ishii, T. Ichida, Q. Gu, and T. Takaki, "500-fps face tracking system," *J. Real-Time Image Proc.*, doi: 10.1007/s11554-012-0255-8, 2012.

[8] I. Ishii, T. Tatebe, Q. Gu, and T. Takaki, "Color-histogram-based tracking at 2000 fps", *J. Electron. Imaging*, **21**, 013010, 2012.

[9] Q. Gu, T. Takaki, and I. Ishii, "Fast FPGA-based multi-object feature extraction," *IEEE Trans. Circuits and Syst. Video Technol.*, **23**, 30–45, 2013.

[10] Y. Nakabo, I. Ishii, and M. Ishikawa, "3D tracking using two high-speed vision systems," *Proc. IEEE/RSJ Int. Conf. Intelli. Rob. Sys.*, 360–365, 2002.

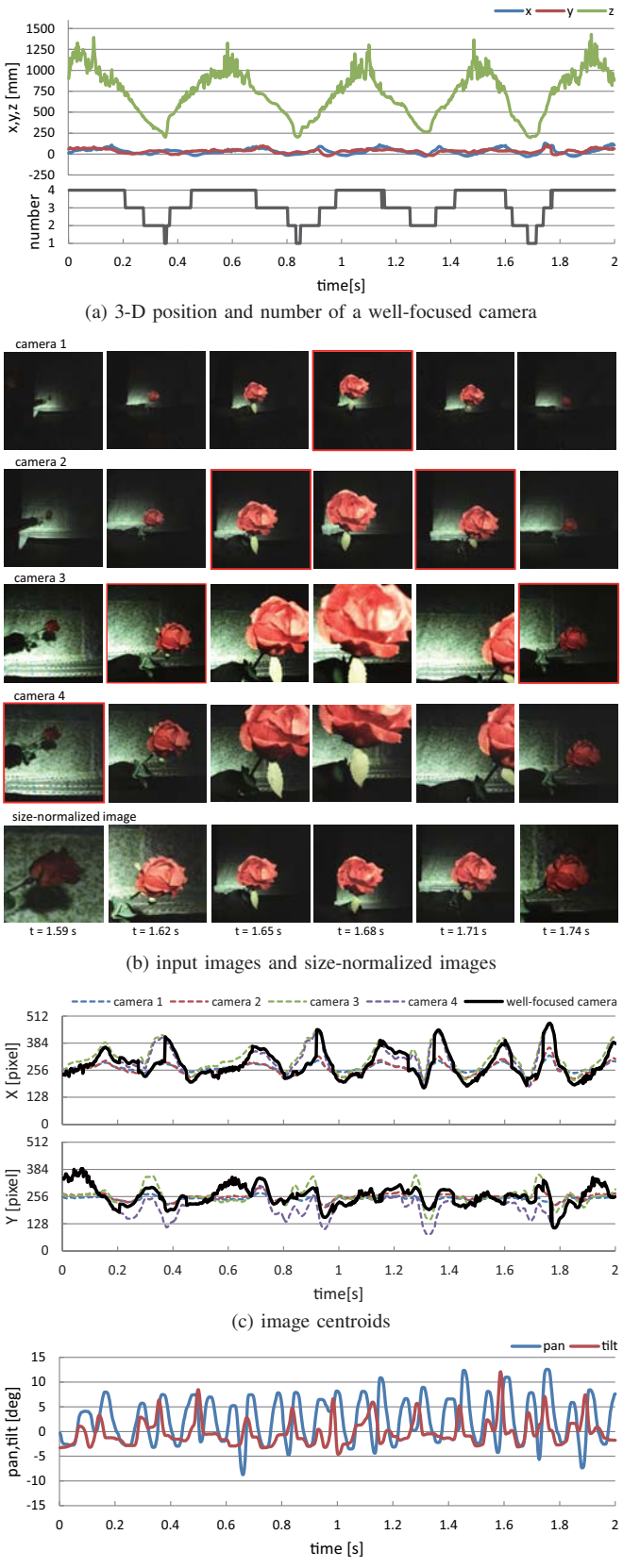
[11] I. Ishii, T. Tatebe, Q. Gu, Y. Moriue, T. Takaki, and K. Tajima, "2000 fps real-time vision system with high-frame-rate video recording," *Proc. IEEE Int. Conf. Robot. Automat.*, 1536–1541, 2010.

[12] L. Shih, "Autofocus survey: a comparison of algorithms," *Proc. SPIE Electron. Imaging*, 650208, 2007.

[13] V. Laude and C. Dirson, "Liquid-crystal active lens: application to image resolution enhancement," *Opt. Comm.*, **163**, 72–78, 1999.

[14] H. Oku and M. Ishikawa, "High-speed liquid lens with 2 ms response and 80.3 nm root-mean-square wavefront error," *Appl. Phys. Lett.*, **94**, 221108, 2009.

[15] A. R. Smith, "Color gamut transform pairs," *Proc. SIGGRAPH Comput. Graph.*, **3**, 12–19, 1978.



(a) 3-D position and number of a well-focused camera

(b) input images and size-normalized images

(c) image centroids

(d) pan and tilt angles of the active vision platform

Fig. 10. Results obtained when a flower was moved back-and-forth in the depth direction.