

Pan/Tilt 스테레오 카메라를 이용한 이동 물체의 강건한 시각추적

조지승*, 정병목#, 최인수**, 노상현***, 임윤규****

Robust 3D visual tracking for moving object using pan/tilt stereo cameras

Che Seung Cho*, Byeong Mook Chung#, In Su Choi**,
Sang Hyun Nho*** and Yoon Kyu Lim****

ABSTRACT

In most vision applications, we are frequently confronted with determining the position of object continuously. Generally, intertwined processes are needed for target tracking, composed with tracking and control process. Each of these processes can be studied independently. In case of actual implementation we must consider the interaction between them to achieve robust performance. In this paper, the robust real time visual tracking in complex background is considered. A common approach to increase robustness of a tracking system is to use known geometric models (CAD model etc.) or to attach the marker. In case an object has arbitrary shape or it is difficult to attach the marker to object, we present a method to track the target easily as we set up the color and shape for a part of object previously. Robust detection can be achieved by integrating voting-based visual cues. Kalman filter is used to estimate the motion of moving object in 3D space, and this algorithm is tested in a pan/tilt robot system. Experimental results show that fusion of cues and motion estimation in a tracking system has a robust performance.

Key Words : Kalman Filter(칼만 필터), Visual Tracking(시각 추적), Moving Object(이동 물체), Cue Integration(큐 융합)

1. 서론

많은 비전시스템의 적용분야에서 대상물의 위치를 연속적으로 결정해야 하는 문제에 접하게 된다. 대상물 추적은 이동로봇의 항법, 물체 인식, 로봇 가이드 시각제어 등과 같은 많은 로봇의 작

업에 있어서 중요한 부분을 차지하고 있다. 비전 시스템을 이용한 이동 물체에 대한 추적은 동적환경에 대한 분석에 있어 유용한 정보를 제공한다. 다양한 연구들이 로봇 비전 분야에서 연구되어 왔다.^{1,2} 일반적으로 목표 추적은 시각추적과 제어의 상호 연관된 과정으로 구성되는데 대부분 이러한

**** 접수일: 2004 년 10 월 25 일; 게재승인일: 2005 년 6 월 29 일

* 대구기계부품연구원

교신저자: 영남대학교 기계공학과

E-mail : bmchung@yu.ac.kr Tel. (053) 810-2569

** 상주대학교 자동차공학과

*** 대구공업대학교 자동차과

**** 울산발전진흥재단 전략발전기획단

과정이 독립적으로 연구되어왔다. 하지만 실제 추적 작업에 있어서 강건한 성능을 얻기 위해서는 위의 두 과정이 동시에 고려되어야만 한다. 추적 과정에서 중요한 핵심은 강건성과 신뢰성의 유지에 있다. Fig.1 은 개략적인 시각제어 시스템을 나타내고 있다. 시각 추적과정은 먼저 대상물을 감지한 후 영상 처리를 통하여 대상물을 추출한다. 추출된 대상물로부터 위치와 속도를 구하고 이를 제어 루프 속으로 되먹임함으로써 연속적으로 시각제어를 수행한다. 제어기는 목표점과 현재의 위치 사이의 오차를 줄이는 역할을 한다.

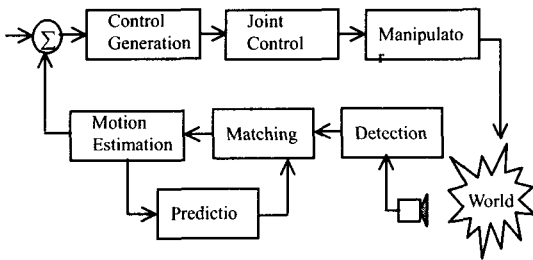


Fig. 1 Major components of a visual servo system

강건한 시각추적을 하기 위한 방법으로는 추적 대상물을 쉽게 분리할 수 있도록 인공적인 식별자(marker)^{3,4} 를 이용하거나 Kollnig, et al.⁵ and Hirzinger, et al.⁶ 의 CAD 모델을 이용한 연구와 같이 잘 정의된 사진 모델을 이용하는 방법이 있다. 정의되지 않은 다양한 형상의 물체 추적의 경우에는 여러 가지 시각정보의 조합을 이용함으로써 보다 강건한 시각 추적을 할 수 있다. 대상물이 임의의 형상을 가지고 있으며 식별자를 부착하기 어렵거나 기하학적 형상 정보가 없을 경우 특정 부위의 색상과 형상 정보들을 사전에 간단히 설정함으로써 대상물을 쉽게 추적할 수 있다. 예를 들면, 이동 로봇 항법과 같은 다이나믹한 환경에서 작업을 수행할 경우 여러가지 센서나 시각 큐를 융합하면 보다 강건한 작업을 수행할 수 있다. 영상처리를 통해 추출된 시각정보의 융합은 신경망, 퍼지 로직, 확률적인 융합, voting 등 다양한 방법이 있다.⁷⁻⁸ 관련 연구를 보면, Bloch⁹는 확률적 기반을 이용하여 패턴인식을 하였으며, Kragic, et al.¹⁰ 는 로봇 작업 시 voting 과 퍼지 로직을 이용하여 시각 큐의 융합을 보여주었다. 하지만 대상물의 운

동에 대한 추정 과정 없이 단순한 look then move 제어만 이루어졌다. 추적물의 영상으로부터 대상물의 위치, 속도 및 자세 등을 추정하는 연구¹¹⁻¹² 도 많이 이루어졌다. Allen¹³ 은 2D 레일 위의 움직이는 기차의 위치를 칼만 필터를 이용하여 추적하였고, Yi¹⁴ 는 칼만 필터를 이용하여 스테레오 영상으로부터 움직이는 3D 물체의 위치 및 깊이 정보를 추정하였다. 하지만 앞의 연구들은 물체와 배경의 분리가 이상적인 경우에 대해서만 이루어져 복잡한 배경인 경우에는 실시간 추적이 어렵다.

본 논문에서는 복잡한 배경에서 임의의 형상을 가지는 대상물에 대하여 강건한 실시간 추적을 위한 방법을 제시한다. 획득된 영상으로부터 대상물의 색상, 형상모멘트, disparity 그리고 SSD(Sum of Squared Differences)의 시각 큐를 추출하고 voting 방법을 이용하여 융합한 후 추적 대상물을 추출한다. Voting 의 경우 확률적 모델을 이용하는 방법 등과 같이 융합을 위한 복잡한 모델을 사용하지 않고도 간단하게 구성할 수 있는 장점이 있다. 실시간 추적을 위해 추적물의 위치와 속도는 칼만 필터를 통하여 추정하고, pan/tilt 시스템을 이용하여 대상물의 연속적인 추적 결과를 보여준다.

2. 시각 큐의 융합

2.1 색상

RGB 색상 공간은 Red(R), Green(G) 그리고 Blue(B)축으로 나타낸다. RGB 색상 공간은 밝기의 변화에 민감하므로 RGB 색상 공간에서 hue(H), saturation(S) 그리고 intensity(I)성분의 HSI 색상 공간으로 변환하여 이용하고 그 변환 관계는 식(1)과 같다.¹⁵

$$H = \arccos \left[\frac{\frac{1}{2}[(R-G) + (R-B)]}{\sqrt{((R-G)^2 + (R-B)(G-B))}} \right]$$

$$S = 1 - \frac{3}{(R+G+B)} \min(R, G, B)$$

$$I = (R+G+B) / 3 \tag{1}$$

2.2 형상 모멘트

모멘트 특징량은 대상물의 확대, 축소 및 회전과 상관없이 일정한 값을 갖는 특징이 있다. 따라

서 도형의 형상을 구분하기 위하여 모멘트 값을 이용한다. 형상 모멘트 m_{ij} 는 다음과 같이 정의된다.

$$m_{ij} = \sum_x \sum_y x^i y^j f(x, y) \quad (2)$$

여기서 i, j 는 모멘트 차수를 나타내고, x, y 는 영상의 수평 및 수직 좌표이다.

도심(center of the geometry), \bar{x}, \bar{y} 와 중심 모멘트(central moment), μ_{ij} 는 식(2)를 이용하여 식(3)으로 표현된다.

$$\begin{aligned} \bar{x} &= \frac{m_{10}}{m_{00}} & \bar{y} &= \frac{m_{01}}{m_{00}} \\ \mu_{ij} &= \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j f(x, y) \end{aligned} \quad (3)$$

추적 형상의 장축(a)과 단축(b)의 근사값은 도심과 중심 모멘트를 이용하여 식(4)와 같이 나타낼 수 있고, 장·단축 길이의 비는 식(5)와 같다.¹⁶

$$\begin{aligned} a &= \sqrt{2} \sqrt{(\mu_{20} + \mu_{02} + \sqrt{((\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2)}) / m_{00}} \\ b &= \sqrt{2} \sqrt{(\mu_{20} + \mu_{02} - \sqrt{((\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2)}) / m_{00}} \end{aligned} \quad (4)$$

$$\text{Axis ratio} = \frac{b}{a} \quad (5)$$

2.3 Disparity

Fig.2 는 스테레오 영상과 3 차원 공간 상의 기하학적 정의를 나타내고 있다.

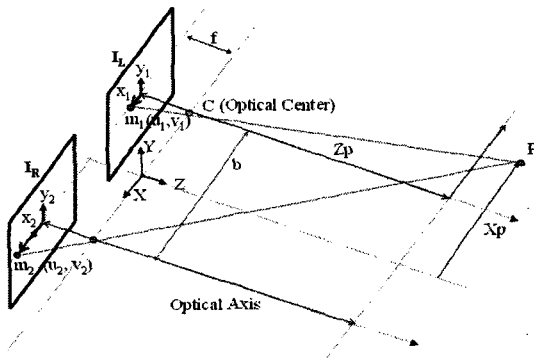


Fig. 2 Parallel stereo camera setup

I_L 과 I_R 은 좌우 영상을 나타낸다. b 는 광축의 중심간의 거리, f 는 유효 초점거리, 점 P 는 공간상의 임의의 점을 나타내며 Z_p 는 물체와 카메라 사이의 거리를 나타낸다.

Fig. 3 은 disparity 의 정의를 나타내는데 m_1 과 m_2 사이의 수평 방향의 좌표 차이(d)를 disparity 라고 정의한다. disparity 와 3 차원 공간 상의 점 P 에 대한 거리의 상관관계는 식(6)과 같다. 연구에서 칼만 필터를 통해 추정된 Z_p 값이 주어지면 disparity 값을 구할 수 있다.

$$d = m_2 - m_1 = \frac{bf}{Z_p} \quad (6)$$

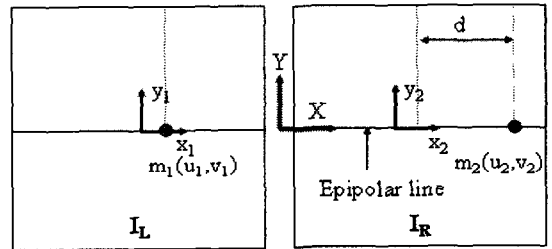


Fig. 3 Definition of disparity

2.4 SSD

스테레오 정합(matching)은 각각의 화소에 대해 상응하는 다른 쪽 영상에서의 화소를 결정하는 것이다. 일반적으로 정합의 오류를 줄이기 위해 좌우 영상에서 대응점을 찾을 때 화소 주위의 영역에 창(window)을 씌워 SSD(Sum of Squared Differences)값을 구한다. 정합 시 탐색 영역의 범위를 줄임으로써 계산 오차와 시간을 줄이는 것이 필요하다. Fig.4 는 intensity 를 기준으로 한 스테레오 정합을 나타내고 있다.

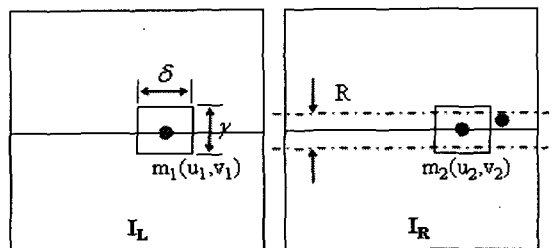


Fig. 4 Intensity based stereo matching

탐색 영역을 줄이기 위하여 극상선(epipolar line)을 수평 주사선과 평행하게 두 카메라를 설치하였으며 정합에 사용한 대응점은 좌우 영상에 대하여 색상에서 구한 시각 큐의 중심점을 이용하였다. 좌측 영상에서 구한 후보점의 위치에서 일정한 검색 영역(R)을 두고 우측 영상에서의 후보점과 매칭하므로 전체 영역을 정합할 때보다 시간과 오차를 줄일 수 있다. 실험에 사용된 매칭 창 크기 δ 와 γ 는 각각 14×14 이다. 식(7)은 SSD를 나타내고 있다.

$$e_{SSD} = \sum_{\delta, \gamma} [I_L(u_1 + \delta, v_1 + \gamma) - I_R(u_2 + \delta, v_2 + \gamma)] \quad (7)$$

2.5 큐의 융합

Voting 방법은 머신 비전에서 다양한 형태로 이용되고 있다. 확률적 융합은 특정 물체와 시각 큐 사이의 관계를 규정하는 P(cue | object) 모델이 필요하지만 Voting 방법은 개별 큐에 대한 모델이 필요하지 않으므로 간단한 분류기로 구성할 수 있다. 일반적으로 분류(classification)공간(Θ)은 큐 추정기(w_i)로부터 다음과 같이 매핑한다.

$$w_i : \Theta \rightarrow [0;1] \quad (8)$$

일반적으로 잘 알려져 있는 Weighted Consensus Voting의 정의는 다음과 같다.

$$w : \Theta \rightarrow [0;1]$$

$$W(\theta) = \begin{cases} \Lambda(C_1(\theta), \dots, C_n(\theta)) & \text{if } \sum_{i=1}^n w_i(\theta) \geq m \\ 0 & \text{otherwise} \end{cases}$$

$$w_i(\theta) = \begin{cases} 1, & \text{if } c_i(\theta) > 0 \\ 0, & \text{otherwise} \end{cases} \text{ for } i=1, \dots, n \quad (9)$$

여기서, $w_i(\theta)$ 는 voting 함수이고, $\Lambda : [0;1]^n \rightarrow [0;1]$ 는 각각의 추정기에 대한 신뢰성 값을 병합하는 함수이다. n 은 큐 추정기의 개수이다. 추정기 값이 m 개 이상으로 양립할 경우 각 추정기에 가중치가 부여된 추정값을 만든다.

본 실험에서는 각각의 시각 큐가 개별 설정 기준을 만족하면 1의 값을 부여하였으며, SSD와 disparity에 대한 가중치는 2배로 두었다. voting

결과 동일한 값을 가지는 물체가 존재할 경우 이전 단계에서 추정된 위치 값을 비교하여 가장 작은 오차 값을 가지는 물체를 추적 대상으로 정하였다.

3. 칼만 필터

추적물은 가속도가 일정한 모델로 가정하고 상태벡터 x_k 는 다음과 같이 정의한다.

$$x_k = [x(k), \dot{x}(k), y(k), \dot{y}(k), z(k), \dot{z}(k)] \quad (10)$$

시스템 모델과 출력 모델은 식(11)과 식(12)로 나타낸다.¹⁷

$$x_k = \Phi_{k-1} x_{k-1} + \omega_{k-1} \quad (11)$$

$$z_k = H_k x_k + v_k \quad (12)$$

여기서, Φ_k 는 6×6 대각 행렬을 나타낸다. $\Phi_k = \text{diag}\{\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \dots, \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}\}$, x_k 는 시스템 상태 벡터, H_k 는 출력행렬을 나타낸다. ω_k 와 v_k 는 프로세스 노이즈 공분산 Q_k 와 측정 노이즈 공분산 R_k 를 가지는 시스템 노이즈와 출력 노이즈이다. T는 샘플링 타임을 나타낸다.

$$E[\omega_k \omega_i^T] = \begin{cases} Q_k, & i=k \\ 0, & i \neq k \end{cases} \quad (13)$$

$$E[v_k v_i^T] = \begin{cases} R_k, & i=k \\ 0, & i \neq k \end{cases} \quad (14)$$

칼만필터 알고리즘은 다음 두 식으로 구성된다.

1) 시간갱신식

$$\hat{x}_k^- = \Phi_{k-1} \hat{x}_{k-1}^+ \quad (15)$$

$$P_k^- = \Phi_{k-1} P_{k-1}^+ \Phi_{k-1}^T + Q_{k-1} \quad (16)$$

2) 측정갱신식

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \quad (17)$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k (z_k - H_k \hat{x}_k^-) \quad (18)$$

$$P_k = (I - K_k H_k) P_k^- \quad (19)$$

\hat{x}_k , K_k , P_k 는 추정상태, 칼만 게인행렬, 추정오차 공분산 행렬을 나타낸다. $H_k \hat{x}_k^-$ 은 예측값 \hat{z}_k 를 나타낸다. $K_k(z_k - H_k \hat{x}_k^-)$ 는 추정값 \hat{x}_k^- 을 \hat{x}_k 로 갱신한다. 실험을 통하여 설정한 노이즈 공분산 행렬 R_k, Q_k 는 다음과 같다.

$$R_k = 0.1I, \quad Q_k = \text{diag}\{20,30,20,30,20,30\} \quad (20)$$

4. 실험

Fig.5 는 전체 시스템을 나타낸다. 스테레오 카메라, pan/tilt 시스템 그리고 PC(Pentium-4)로 구성되어 있다. 실험에서는 추적 대상물로부터 특정 부위에서 4 가지 색상의 구를 추출하고 각 부위의 형상 정보를 추출하였다. 대상물이 움직이면 카메라에서 획득된 영상으로부터 각각의 시각 큐가 추출되고 voting 방법을 이용하여 배경으로부터 대상물을 추출하였다. 영상의 크기는 320×240 이며, 구의 장·단축의 비는 0.6~1.0 로 설정하였다.

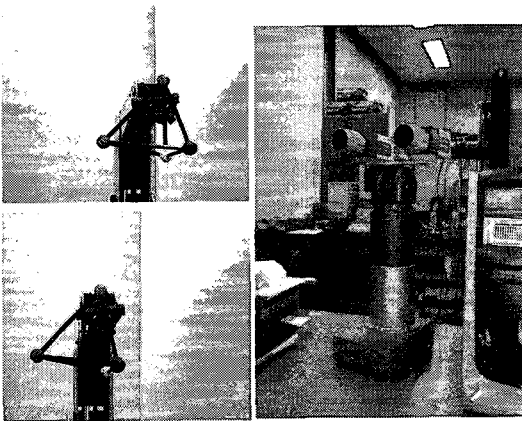


Fig. 5 Experimental Setup

Fig.6 은 본 연구에서 사용된 시각 큐의 융합 과정을 나타낸다. 스테레오 카메라로부터 획득된 영상을 이용하여 4 가지 시각 큐를 구해내고, 이를 voting 방법을 통하여 융합함으로써 대상물을 배경에서 분리한다.

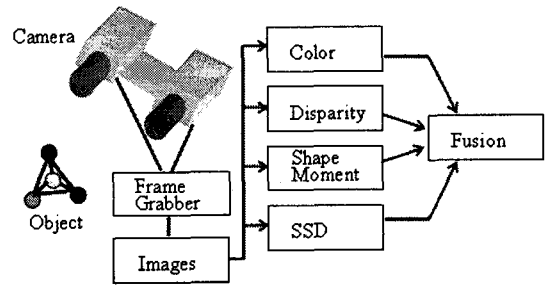


Fig. 6 Overview of fusion of visual cues

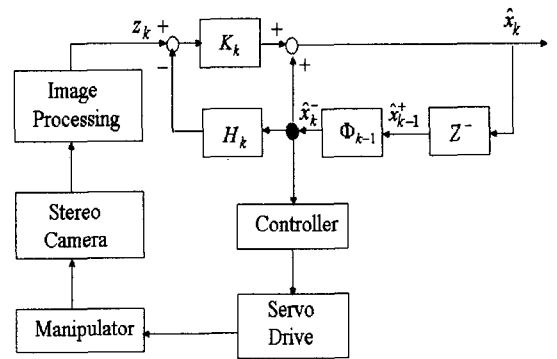
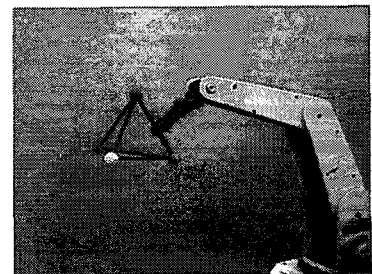


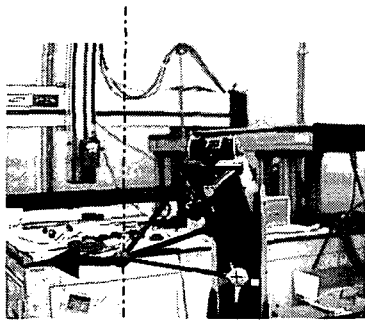
Fig. 7 Block diagram for visual servo

Fig.7 은 전체 제어기 구성을 나타낸다. 영상 처리를 통해 4 가지 부위의 추적물을 분리하고 이들의 도심을 구한다. 칼만 필터로부터 추출된 대상물의 중심 위치와 이동 속도를 구하면 pan/tilt 시스템은 추정된 위치에 따라 물체를 추적한다.

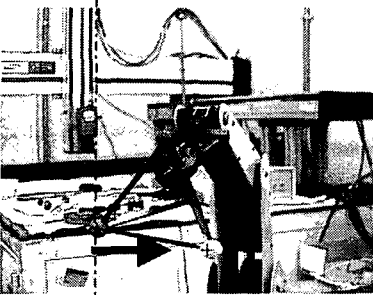
Fig.8 은 복잡한 배경을 가지는 공간에서 시각 큐 융합을 통하여 좌우로 이동하는 물체에 대한 추출 결과를 보여준다. 화살표는 물체의 이동 방향을 나타낸다.



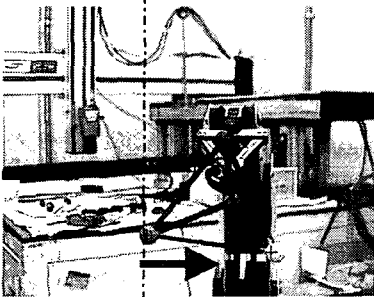
(a) 1th



(b) 1th



(c) 13th

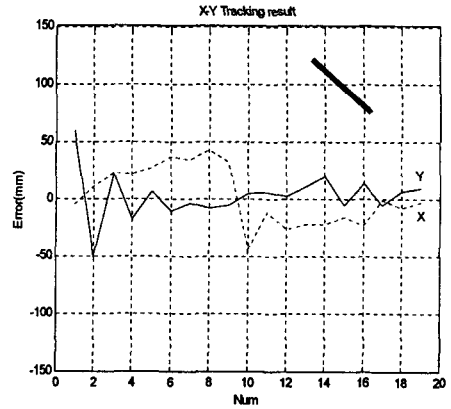


(d) 17th

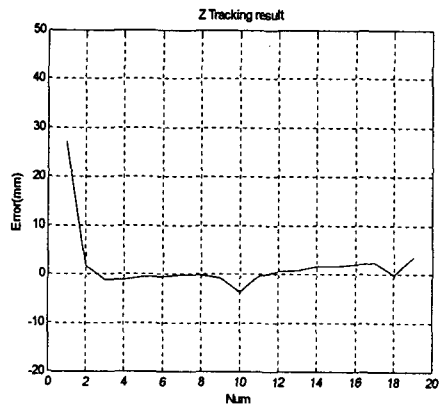
Fig. 8 Extracted target by que integration

Fig.8(a)는 배경이 이상적인 경우를 보여주며 Fig.8(b)-(d)는 배경이 복잡한 경우에 대한 추적과정을 보여준다. Fig.8(b)-(d)에서 보면 실제 배경에서 4 가지 시각 큐를 만족하는 여러 물체가 존재함을 볼 수 있는데 이는 추정된 물체의 도심을 기준으로 일정 범위의 거리를 벗어나는 물체는 제외하여 추적대상물을 추출한다.

Fig.9 와 Fig.10 은 추적물의 직선운동과 사각형 운동에 대하여 각각의 x, y, z 축 위치 추적 오차의 결과를 나타내고 있다. Fig.9 와 Fig.10 의 (a)에서 나타낸 점선은 x 축의 위치 오차를 나타내고, 실선은 y 축의 위치 오차를 나타낸다.

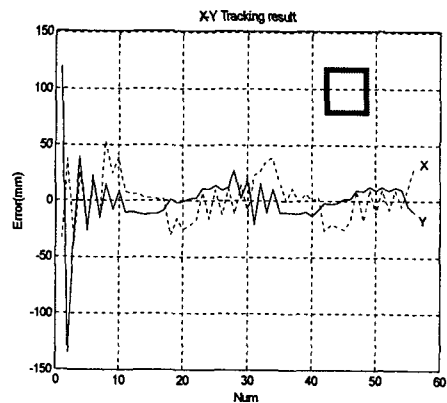


(a) Tracking result of x-y position



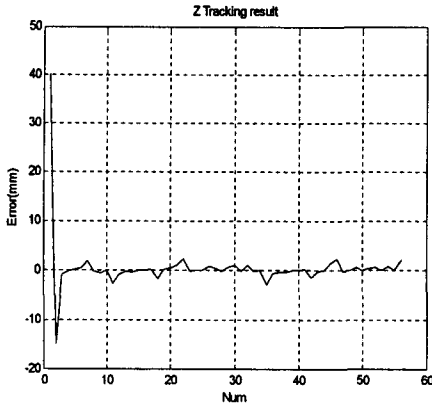
(b) Tracking result of z position

Fig. 9 Estimated tracking result for point to point motion



(a) Tracking result of x-y position

참고문헌



(b) Tracking result of z position

Fig. 10 Estimated tracking result for square motion

5. 결론

본 연구에서는 정의되지 않은 다양한 형상의 물체 추적의 경우에 여러 가지 시각 정보의 조합을 이용함으로써 보다 강건한 시각 추적을 할 수 있음을 보였다. 대상물이 임의의 형상을 가지고 있으며 식별자를 부착하기 어렵거나 기하학적 형상 정보가 없을 경우 특정 부위의 색상과 형상 정보들을 사전에 간단히 설정함으로써 대상물을 쉽게 추적할 수 있었다. 먼저 비전 시스템을 이용하여 시각 큐를 획득하고 voting 방법을 통하여 융합함으로써 배경에서 추적 대상물을 분리하였다. 3D 공간에서 움직이는 물체의 실시간 추적을 위해 칼만 필터를 이용하여 위치를 예측하고 pan/tilt 로봇을 이용하여 실시간 추적하였다. 현재 물체의 자세 추정에 대한 추가적인 연구를 수행 중이며 이를 통하여 공장의 조립 라인 등 다양한 분야에 이용될 수 있을 것이다. 시각 큐의 수와 복잡성은 실시간 추적에 큰 영향을 미치므로 향후 시각 큐의 수와 각각의 신뢰도에 따른 영향을 살펴보아야 할 것이다.

후 기

이 논문은 2004 년도 두뇌한국 21 사업에 의하여 지원되었음.

1. Cretual, A., "Complex object tracking by visual servoing based on 2D image motion," Proceedings 14th International Conference on Pattern Recognition, Vol. 2, pp.1251-1254, 1998.
2. Lim, Y. N. and Lee, S. C., "Real-time Target Tracking System by Extended Kalman Filter," Journal of the Korean Society of Precision Engineering, pp. 175-181, 1998.
3. He, D., Hujic, D., Mills, J. K. and Benhabib, B., "Moving object recognition using premarking and active vision," Proceedings of the IEEE International Conference on robotics and automation, pp. 1980-1985, April, 1996.
4. Maniere, E. C., Couvignou, P. and Khosla, P. K., "Robotic contour following based on visual servoing," Proceedings of the IEEE/RSJ International Conference on intelligent robotics and system, pp. 716-722, July, 1993.
5. Kollnig, H. and Nagel, H., "3D pose estimation by directly matching poly-hedral models to gray value gradients," International Journal of Computer Vision, Vol. 23, No.3, pp. 282-302, 1997.
6. Hirzinger, G., Fisher, M., Brunner, B., Koeppel, R., Otter, M., Grebenstein, M. and Schäfer, I., "Advanced in robotics: The DLR experience," International Journal of robotics research, Vol. 18, pp. 1064-1087, Nov. 1999.
7. Kopp-Borotsching, H. and Pinz, A., "A new concept for active fusion in image understanding applying fuzzy set theory," Fifth IEEE International Conference on Fuzzy Systems, Vol. 2, pp.793-799, sep. 1996.
8. Parhami, B., "Voting algorithm," IEEE Transaction on Reliability, Vol.43, No.4, pp. 617-629, 1994.
9. Bloch, I., "Information combination operators for data fusion," IEEE Transaction on System Man and Cybernatics, Vol.26, No.1, pp.42-52, 1996.
10. Kragic, D. and Christensen, H. I., "Cue integration for visual servoing," IEEE Transaction on robotics and automation, Vol.17, No.1, pp. 18-26, 2001.
11. Wang, J. and Wilson, W. J., "3-D relative position and orientation estimation using Kalman filter for

- robot control," IEEE International conference on Robotics and automation, pp. 2638-2645,1992.
12. Papanikolopoulos, N. P. and Khosla, P. K., "Adaptive robotic visual tracking: Theory and experiments," IEEE Transaction on Automation and Control, Vol.38, pp. 429-445, 1993.
 13. Allen, P., "Automated tracking and grasping of a moving object with a robotic hand-eye system," IEEE Transactions on robotics and automation, Vol. 9, p. 152, 1993.
 14. Yi, J. W., Yang, T. S. and Oh, J. H., "Estimation of depth and 3D motion parameters of moving objects with multiple stereo images by using Kalman filter," IEEE IECON 21st International Conference on Industrial Electronics, Control, and Instrumentation, Vol.2, pp.1225-1230, Nov. 1995.
 15. Gonzalez, R. C. and Wood, R. E., "Digital Image Processing," Addison Wesley, 1992.
 16. Heijden, F., "Image based Measurement system," Wiley, 1994.
 17. Grewal, M. S. and Andrews, A. P., "Kalman filtering theory and practice," Prentice hall, 1993.