

## 칼만필터를 이용한 3-D 이동물체의 강건한 시각추적

조지승\*(영남대 대학원 기계공학과), 정병목(영남대 기계공학과)

### Robust Visual Tracking for 3-D Moving Object using Kalman Filter

C. S. Cho(Mechanical Eng. Dept. YNU), B. M. Chung(Mechanical Eng. Dept., YNU)

#### ABSTRACT

The robustness and reliability of vision algorithms is the key issue in robotic research and industrial applications. In this paper robust real time visual tracking in complex scene is considered. A common approach to increase robustness of a tracking system is the use of different model (CAD model etc.) known a priori. Also fusion of multiple features facilitates robust detection and tracking of objects in scenes of realistic complexity. Voting-based fusion of cues is adapted. In voting, a very simple or no model is used for fusion. The approach for this algorithm is tested in a 3D Cartesian robot which tracks a toy vehicle moving along 3D rail, and the Kalman filter is used to estimate the motion parameters, namely the system state vector of moving object with unknown dynamics. 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), Visual Cue(시각 큐), Cue Integration(큐 융합)

#### 1. 서론

대상물 추적은 이동로봇의 항법, 물체 인식, 속도 추정등과 같은 많은 로봇의 작업에 있어서 중요한 부분을 차지하기 때문에 최근 시각추적(Visual tracking)과 시각 제어(Visual servoing)에 관한 연구가 많이 이루어지고 있다.

비전시스템을 이용한 이동 물체에 대한 추적은 동적환경에 대한 분석에 있어 유용한 정보를 제공한다. 시각제어는 움직이는 물체의 추적과 제어의 상호 연관된 과정으로 구성된다. 움직이는 물체에 대한 추적은 배경과 관심물체를 분리하는 것이라 볼 수 있으며 다음 세가지 과정으로 나눈다. 1)대상물 감지, 2)영상 매칭, 3)물체의 움직임에 대한 예측이다. 추적 대상물의 위치와 속도를 찾으며 그 값을 제어 루프에 피드백 함으로써 대상물을 추적할 수 있다.

강건한 시각추적을 하기 위한 방법으로는 인공적인 marker 를 이용하여 추적 대상물을 찾아내는 방법과 CAD 데이터 같은 잘 정의된 사전 모델을 이용하는 방법이 있다. 하지만 정의되지 않은 다양한 형상의 물체를 추적하는 경우에는 여러 가지 시

각정보의 조합을 이용함으로써 보다 강건한 추적을 할 수 있다. 예를 들면, 로봇 항법과 같은 다이나믹한 환경에서 작업 수행 시 멀티 센서나 큐를 융합하면 보다 강건한 작업을 수행할 수 있다. 영상처리를 통한 추출된 시각정보의 융합은 신경망, 퍼지 로직, 확률적인 융합, voting 등 다양한 방법이 있다[1-2]. 관련 연구를 보면, I.Boch[3]는 확률적 기반을 이용하여 패턴인식을 하였으며, Danica K. et al.[4] 는 로봇 작업 시 시각 큐를 voting 과 퍼지 로직을 이용하여 융합하였다.

추적물의 영상으로부터 대상물의 위치, 속도 및 자세등을 추정하는 연구도 많이 이루어졌다[5-6]. P. Allen[7]은 2D 레일 위의 움직이는 기차의 위치를 칼만 필터를 이용하여 추적하고 로봇을 이용하여 잡았으며, W. Y. Jae[8]는 칼만필터를 이용하여 스테레오 영상으로부터 움직이는 3D 물체의 위치 및 깊이 정보를 추정하였다. 본 연구에서는 시각제어 시스템을 이용하여 3D 레일을 따라 움직이는 기차를 3D 적교로봇을 이용하여 실시간 추적하는 알고리즘을 제안한다. 복잡한 배경으로부터 색상, 형상모멘트, disparity 와 같은 시각 큐(visual cue)를 voting 을 통해 융합하여 추적 대상물을 추출한다. Voting 의

경우 융합을 위한 모델을 사용하지 않고 간단하게 구성할 수 있는 장점이 있다. 실시간 추적을 위해 추적물의 위치 및 속도는 칼만 필터로 예측하고 3D 직교로봇을 이용하여 추적하고 그 결과를 보여준다.

## 2. 시각 큐

다음은 실험에 사용된 시각 큐를 나타내고 있다.

### 2.1 색상 분석

색상에 대한 검출은 hue(H), saturation(S)성분을 이용한다. 색상에 대한 학습은 off-line 이루어진다. 즉 사전에 움직이는 대상물의 색상에 대한 정보는 주어지고 이로부터 H-S 평면에서의 분포를 계산한다.

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

$$S = 1 - \frac{3}{(R+G+B)} \min(R+G+B) \quad (2)$$

### 2.2 형상모멘트

도형의 형상을 구분하는데 중요하게 적용할 수 있는 모멘트를 이용한다. 모멘트 특징량은 대상체의 확대, 축소 및 회전과 상관없이 일정한 값을 갖는 특징이 있다. 모멘트  $m_{ij}$ 는 다음과 같이 정의된다.

$$m_{ij} = \sum_x \sum_y x^i y^j f(x,y) \quad (3)$$

여기서  $i, j$ 는 모멘트 차수이고,  $x, y$ 는 영상의 수평 및 수직 좌표이다.

$x, y$ 는 무게 중심(center of the gravity),  $\mu_{ij}$ 는 중심 모멘트(central moment)를 나타낸다.

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

$$\mu_{ij} = \sum_x \sum_y (x-\bar{x})^i (y-\bar{y})^j f(x,y) \quad (4)$$

컬러 바의 장,단축 길이의 근사값은 다음과 같고, 길이의 비는 다음과 같다.[9]

$$a \approx \sqrt{2} \sqrt{(\mu_{20} + \mu_{02} + \sqrt{((\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2)}) / m_{00}}$$

$$b \approx \sqrt{2} \sqrt{(\mu_{20} + \mu_{02} - \sqrt{((\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2)}) / m_{00}}$$

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

### 2.3 Disparity

Disparity는  $m_1$ 과  $m_2$ 간의 수평 방향의 좌표차이로 정의된다. Fig 1은 Disparity와 3차원 공간상의 점 M과의 거리와의 상관관계를 나타내며 다음과 같다.

$$d = v_2 - v_1 = \frac{bf}{z} \quad (6)$$

여기서,  $b$ 는 disparity,  $f$ 는 유효초점거리,  $z$ 는 물체와 카메라사이의 거리를 나타낸다.

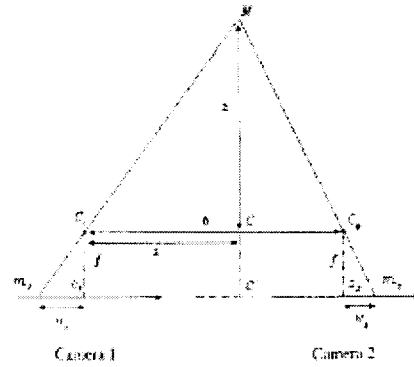


Fig. 1 Parallel stereo camera setup

스테레오 정합은 좌우 영상의 동일점을 찾는 작업으로 정합 시 탐색영역의 범위를 줄임으로서 계산오차와 시간을 줄이는 것이 필요하다. 탐색영역을 줄이기 위하여 극상선(epipolar line)을 수평 주사선과 평행하게 두 카메라를 설치하였다.

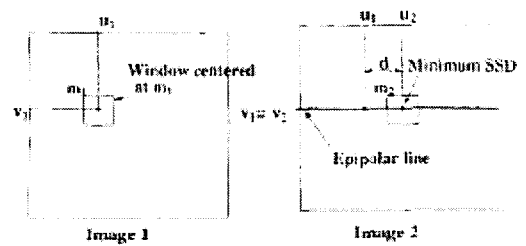


Fig. 2 Intensity based stereo matching

스테레오 정합은 각각의 화소에 대해 상응하는 다른 쪽 영상에서의 화소를 결정하는 것이지만 정합의 오류를 줄이기 위해 대응점의 길쭉은 화소 주위의 영역에 창(window)을 씌워 SSD(Sum of Squared Differences)가 최소값을 가지는 좌표 값을 찾고 두 영상의 x 좌표 차이를 disparity로 구하였다. SSD 오차는 다음과 같다.

$$e(u,v,d) = \sum_{m,n} [I_1(u+m,v+n) - I_2(u+m,v+n)] \quad (7)$$

### 3. 시각 큐의 융합

Voting 은 머신 비전에서 다양한 형태로 이용되며 확률적 융합보다 개별 큐에 대한 모델이 필요하지 않으므로 간단한 분류기로 구성할 수 있다. 일반적인 분류(classification)공간(Θ)는 큐 추정기(v<sub>i</sub>)로 다음과 같이 매핑된다.

$$v_i : \Theta \rightarrow [0,1] \quad (8)$$

일반적인 잘 알려져 있는 Weighted Consensus Voting의 정의는 다음과 같다.

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

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

여기서, v<sub>i</sub>(θ) 는 voting 함수이고, Λ: [0;1]<sup>n</sup> → [0;1] 는 각 추정기의 값을 융합하는 함수, n 은 큐 추정기의 개수이다.

### 4. 칼만 필터

움직이는 물체의 위치를 추정하기 위한 시스템과 출력모델은 식(1), (2)의 선형운동방정식과 같다.

$$x_{k+1} = \Phi_k x_k + \omega_k \quad (10)$$

$$z_k = H_k x_k + v_k \quad (11)$$

Φ<sub>k</sub> 는 시스템행렬, x<sub>k</sub> 는 시스템상태 벡터, H<sub>k</sub> 는 출력행렬을 나타낸다. ω<sub>k</sub> 와 v<sub>k</sub> 는 프로세스 노이즈 공분산 Q<sub>k</sub> 와 측정노이즈 공분산 R<sub>k</sub> 을 가지는 시스템 노이즈와 출력 노이즈이다.

$$E[\omega_k \omega_i^T] = \begin{cases} Q_k, & i = k \\ 0, & i \neq k \end{cases} \quad (12)$$

$$E[v_k v_i^T] = \begin{cases} R_k, & i = k \\ 0, & i \neq k \end{cases} \quad (13)$$

칼만필터 알고리즘은 다음 두 식으로 구성된다.

1) 시간갱신식

$$\hat{x}_{k+1}^- = \Phi_k \hat{x}_k^- \quad (14)$$

$$P_{k+1}^- = \Phi_k P_k \Phi_k^T + Q_k \quad (15)$$

2) 측정갱신식

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \quad (16)$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H_k \hat{x}_k^-) \quad (17)$$

$$P_k = (I - K_k H_k) P_k^- \quad (18)$$

$\hat{x}_k, K_k, P_k$  는 추정상태, 칼만게인행렬, 추정오차 공분산 행렬을 나타낸다.  $H_k \hat{x}_k^-$  은 예측 값  $\hat{z}_k$  를 나타낸다.  $(z_k - H_k \hat{x}_k^-)$  은 예측 오차로 innovation 또는 residual  $r$  로 다음과 같이 정의된다.

$$r_k = (z_k - H_k \hat{x}_k^-) \quad (19)$$

$K_k (z_k - H_k \hat{x}_k^-)$  는 추정값  $\hat{x}_k^-$  을  $\hat{x}_k$  로 갱신한다.

공분산 행렬 Q<sub>k</sub> 와 R<sub>k</sub> 는 각각 60, 0.1 로 설정하였고, 시스템 상태변수 다음과 같다.

$$x_k = [x(t), \dot{x}(t), y(t), \dot{y}(t), z(t), \dot{z}(t)] \quad (20)$$

### 5. 실험

Fig.3 은 전체 시스템을 나타내며 3 축 직교로봇, 서보 드라이브, 컨트롤러, 스테레오 카메라, PC(Pentium-3)로 구성되어 있다. 적색의 컬러바(마커)가 부착된 기차가 3D 레일 위로 움직이면 카메라 영상으로부터 시각 큐를 융합하여 기차를 인식하고 추출한다. 컬러바의 장단축의 비는 0.6(1.5/2.5)이다.

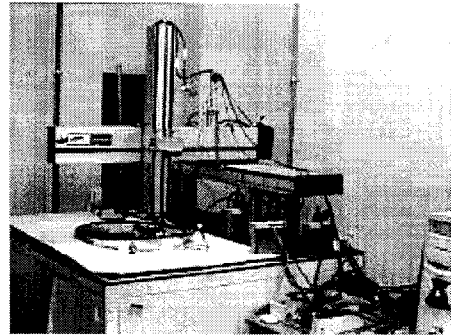


Fig. 3 Experimental Setup

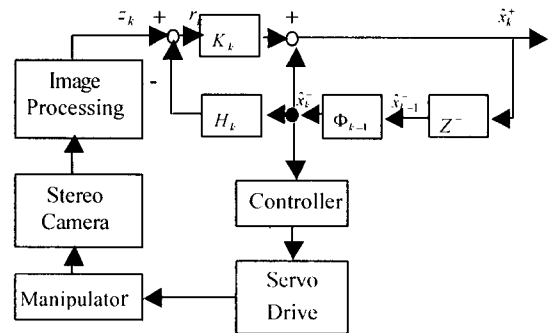


Fig 4 Block diagram for visual servo

Fig.4 는 전체 제어기 구성을 나타내며 로봇은

칼만필터로부터 추정된 위치를 따라 기차를 추적한다. Fig.5(a)는 레일 위의 이동 물체를 보여주며 Fig.5(b)는 큐 융합을 통하여 배경으로부터 레일 위의 기차를 추출한 결과를 보여준다. 평균 영상처리시간은 0.5 초 소요되었다.

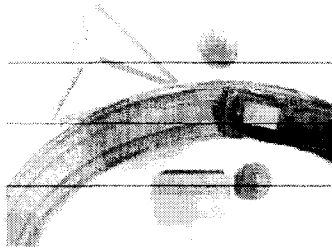


Fig 5(a) Original Image



Fig 5(b) Extracted Target

Fig 5 Extracted target by que integration

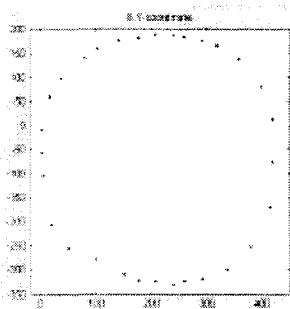


Fig 6(a) X-Y-coordinate

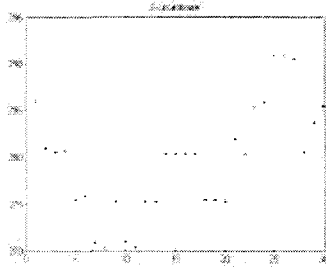


Fig.6 b) Z-coordinate

Fig 6 Result of tracking position

Fig.6 은 전체 추적 영역과 추적결과를 보여준다. Fig.6(a)는 로봇 manipulator 의 추적 위치를 나타내고, Fig.6(b)는 추정깊이를 나타낸다.

## 6. 결론

본 연구에서는 비전 시스템을 이용하여 시각 큐를 획득하고 voting 을 통해 융합하여 추적 대상물을 추출하였다. 3D 레일 위를 따라 움직이는 기차는 실시간 추적을 위해 칼만필터를 이용하여 위치를 예측하고 3D 직교로봇을 이용하여 실시간 추적하였다. 시각 큐의 수와 복잡성은 실시간 추적에 큰 영향을 미치므로 향후 시각 큐의 수와 각각의 신뢰도에 따른 영향을 살펴보아야 할 것이다. 또한 정의되지 않은 다양한 형상의 물체에 대한 추가적인 연구가 필요하다.

## 참고 문헌

1. H. Borotsching, et al. "A new concept for active fusion in image understanding applying fuzzy set theory," in Proc. 5<sup>th</sup> IEEE Int. Conf. Fuzzy System, 1996.
2. B. Parhami, "Voting algorithm," IEEE Trans. Rel., vol.43, no.3, pp. 617-629, 1994.
3. Danica K. et al., "Cue integration for visual servoing", IEEE Transaction on robotics and automation, Vol.17, No.1, 2001.
4. I. Bloch, "Information combination operators for data fusion", IEEE Trans. Sys. Man. Cyber. A, vol.26, no.1, pp.42-52, 1996.
5. N. P. Papanikolopoulos and P. K. Khosla, "Adaptive robotic visual tracking: Theory and experiments," IEEE Trans. Automat. Control, vol.38, pp. 429-445, 1993.
6. H. Kollnig and H. Nagel, "3D pose estimation by directly matching poly-hedral models to gray value gradients," Int. J. Computer Vision., vol. 23, no.3, pp. 282-302, 1997.
7. P. Allen, "Automated tracking and grasping of a moving object with a robotic hand-eye system," IEEE Trans. Robot. Automat, vol. 9, p. 152, 1993.
8. W.Y. Jae et al., "Estimation of depth and 3D motion parameters of moving objects with multiple stereo images by using Kalman filter". IEEE, pp. 1225-1230, 1996
9. F. Heijden, "Image Based Measurement System." Wiley, 1994.