

# 스네이크를 이용한 영역기반 물체추적 알고리즘

김영섭\* · 한규범\* · 백윤수\*\*

## Region Based Object Tracking with Snakes

Young Sub Kim , Kyu Bum Han and Yoon Su Baek

**Key Words :** Vision System(비전 시스템), Object Tracking(물체추적), Snake(스네이크), Region Segmentation(영역 분할), 대응문제(Correspondence Problem)

### Abstract

In this paper, we proposed the object-tracking algorithm that recognizes and estimates the any shaped and size objects using vision system. For the extraction of the object from the background of the acquired images, spatio-temporal filter and signature parsing algorithm are used. Specially, for the solution of correspondence problem of the multiple objects tracking, we compute snake energy and position information of the target objects. Through the real-time tracking experiment, we verified the effectiveness of the suggested tracking algorithm.

### 1. 서론

일반적으로 머신 비전을 영상 추적의 관점에서 보면 추적체를 인식하고 영상 시퀀스로부터 추적체의 위치 정보를 연속적으로 얻어내는 것이 중요하다. 이러한 관점에서 인간의 시각 시스템과 머신 비전 시스템의 추적 메커니즘은 추적 대상 물체의 감지 및 분리, 연속적인 추적이라는 두가지 측면에서 그 유사성을 찾을 수 있다. 이러한 관점에서 머신 비전 분야의 물체 추적 시스템을 분리하면 모델 기반 방법(model-based method), 영상 기반 방법(image-based method), 동적 형상 기반 방법(active-contour based method)으로 구분할 수 있다. 모델에 기반한 방법<sup>(1,2)</sup>은 모델의 구축이 필수적이며 3D 물체의 간략화를 통한 모델과의 매칭을 통해 안정적인 추적이 가능하다는 강점을 가지는 반면 모델화 되지 않은 임의의 형상에 대한 추적은 수행하기 힘들다는 한계점을 가지고 있다. 또한 3차원 공간 내에서의 변형을 수반하거나 복잡한 형

상의 물체인 경우 인식 및 추적이 어려운 단점을 가진다. 영상에 기반한 방법<sup>(3,4)</sup>은 영상으로부터 얻을 수 있는 특징을 이용하여 운동의 연속성을 가정하고 이들간의 대응을 구함으로써 물체의 동특성을 파악하고 추적하는 방법으로 추적체의 크기나 형상에 관계없이 추적이 가능하다는 강점을 가지나 노이즈 요소에 민감하고, 매칭 문제를 해결해야 한다는 한계점을 가진다. 동적 형상 기반 방법은 Kass<sup>(5)</sup>의 기본 모델(snake)로부터 기인된 추적 이론으로 기존의 점이나 선과 같은 영상 특징치를 추적하는 방법보다 물체의 형상이나 경계가 변화하는 훨씬 더 복잡한 운동 요소를 추적할 수 있다는 점에서 강점을 가지는 방법이다<sup>(6)</sup>. 그러나 이 방법은 추적 특징치가 물체의 윤곽에 한정되는 단점을 가지고 있다. 사용자가 적절한 초기 위치의 설정 작업이 필요하고(semi automatic), 물체가 영상에서 느린 속도의 운동을 해야 한다는 한계점이 있다<sup>(7)</sup>. 또한 기존의 스네이크로는 3차원 물체의 운동 및 자세 추정이 어렵고 매칭 문제의 해결이 어렵다.

본 연구에서는 기존의 추적 방법들을 분석하고 기존 방법들의 한계점을 보완할 수 있는 알고리즘을 제시한다. 먼저, 영상 기반 방법 중 영역 기반의 방법을 사용하여 전처리 과정을 수행한다. 영

\* 연세대학교 기계공학과 대학원

\*\* 연세대학교 기전공학부

상 내의 추적체 영역 추출을 위해 Murray<sup>(8)</sup>가 제시한 spatio-temporal 필터를 통해 영상 내 운동 에너지를 구하고 이로부터 영상 내 추적체의 위치 영역을 추출한다. 다음으로 추적체의 위치 정보를 바탕으로 스네이크를 구성하여 추적체의 윤곽을 추출하고 영상 내의 무게 중심을 계산한다. 스네이크의 인위적 초기화를 피하기 위해 물체 영역 분할 정보에 의해 제어점을 설정하였으며, 이러한 방법은 스네이크가 영상 노이즈에 의한 국부 최소값(local minimum)에 수렴해버리는 경우를 배제시키며 따라서 추적체의 대략적 형상을 강건하게 추출할 수 있는 도구가 된다. 또한 속도 향상을 위해 Greedy method<sup>(9)</sup>에 의한 8 근방 탐색을 적용한다. 스네이크 알고리즘의 적용으로부터 얻을 수 있는 또 하나의 장점은 이를 이용하여 복수 추적체가 존재하는 경우, 대응성 문제(correspondence problem)의 해결 도구가 될 수 있다는 점이다. 기존의 대응 문제 해결 방법으로 SSD (Sum of Squared Difference) 방법 또는 유사성 검토(likelihood test)를 사용하는 방법 등이 있으나, 이 경우 계산 시간이 많이 걸리고 영상 노이즈에 민감하다는 단점이 있으며 일반적으로 대응 문제의 해결을 전제로 추적 시스템을 구성한다. 본 연구에서는 각 추적체의 에너지 준위를 분석함으로써 영상내의 위치 구조와 함께 대응 문제를 해결하는 방법을 제시한다. 본 연구에서 제시하는 추적 알고리즘의 특징은 다음과 같다.

- (1) 영상내의 추적체의 추출은 영상 기반의 방법 중 추적체 영역 추출을 기반으로 하며 따라서 특정 물체의 모델 구축이 필요하지 않은 임의의 형상을 가진 물체의 추적을 수행한다.
- (2) 스네이크를 사용하여 추적체의 윤곽을 추출함으로써 강제 뿐 아니라 변형이 가능한 비강체 추적에도 적용될 수 있는 알고리즘을 구성한다. 스네이크 적용시 문제가 되는 초기 제어점의 위치 설정은 spatio-temporal 필터링을 통해 분할된 추적체 영역의 경계상으로 설정함으로써 인위적 조작을 피함과 동시에 스네이크 제어점이 국부 최소값으로 수렴할 수 있는 가능성을 최소화하는 강건한 추출을 도모한다.
- (3) 다중 물체 추적의 경우, 대응문제의 해결을 위한 방안으로써 영상 내 물체의 위치 구조와 스네이크의 에너지 분석을 통해 이를 해결하는 알고리즘을 제시한다.

## 2. 추적체 영역 추출

### 2.1 전처리(Preprocess)

본 연구에서는 spatio-temporal 필터를 이용하여 영상 내의 추적체 영역을 추출한다. Spatio-temporal 필터는 연속되는 차영상을 계산하는 방법으로 이로부터 다음 식(1)과 같이 영상의 운동 에너지를 계산한다.

$$\frac{df(x, y, t)}{dt} \approx \frac{f(x, y, t) - f(x, y, t - \Delta)}{\Delta} \quad (1)$$

Fig. 1 은 전처리 영상의 알고리즘을 보여준다.

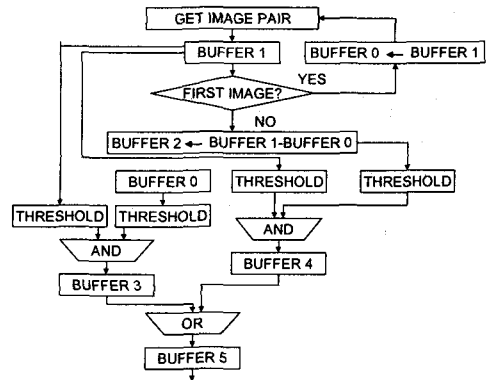


Fig. 1 Preprocess algorithm

### 2.2 추적체의 추출 및 결정

전처리 과정을 사용하여 얻은 이진 영상(binary image)에는 배경이 가지는 효과는 사라지고 추적 물체의 후보군으로 간주할 수 있는 영역들만 남게 된다. 추적체 영역의 검출은 선행된 전처리 과정을 통해 얻은 영상에 대해 signature parsing 과정을 거쳐 영역 분할을 수행한다. 각각 수평 방향과 수직 방향에 대한 signature는 다음 식(2)와 (3)과 같이 정의된다.

$$S_h(i) = \sum_j^M f[i, j] \quad (2)$$

$$S_v(j) = \sum_i^N f[i, j] \quad (3)$$

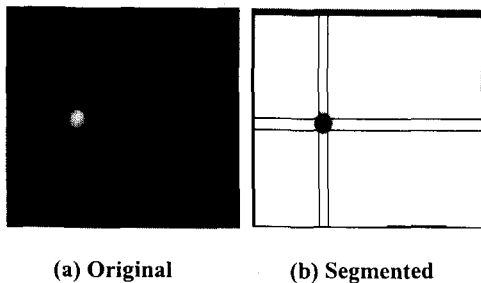
위 식들에서  $S_h(i)$ 는 각 행을 가로질러서 합산한

영상의 강도합(intensity summation)을 나타내고,  $S_i(j)$ 는 각 열을 가로질러서 합산한 영상의 강도합을 나타낸다. 추적체가 흰색이고 배경이 검정색으로 전처리된 이진 영상의 경우, 강도합이 0에서 양의 값으로, 혹은 양의 값에서 0으로 전이 하는 점을 찾아 밴드(band)를 형성하고, 이 밴드를 이용해 사각형의 부영역(sub image)들을 형성한다. 이 부영역들은 각각 추적체를 포함할 수 있는 후보군이 된다. 분할된 부영역은 노이즈 효과에 의해 분할된 경우가 발생할 수 있고, 하나의 부영역에 두 개 이상의 추적체가 포함될 수 있다는 문제가 생긴다.

본 연구에서는 이러한 문제점들의 해결을 위해 전자의 경우는 크기에 제한을 주는 사이즈 필터링을 수행하였고, 후자의 경우 분할된 영역에 대해 물체를 포함하는 영역인지 판별한 다음 물체를 포함할 경우 단일 물체를 포함할 때까지 반복적 분해를 수행하였다.

이러한 알고리즘의 적용으로부터 물체와 물체 그룹 사이의 공간적 상호 관계(spatial interrelation)의 계층적 구조를 포함하는 2차원 SQRT(Spatial Object Relation Tree)를 얻을 수 있다. SQRT는 매칭에 대해 강력한 도구가 되고 분산 영상을 멀티 프로세서로 병렬 처리하는데 사용된다. 추적체 영역 분할 과정을 보면 주변 환경이 복잡할수록 전이점이 형성되지 않는 경우가 많다. 이것은 여러 개의 물체가 겹쳐져서 화면에 나타나는 경우 또는 노이즈 효과가 전처리 과정에서 제거되지 않고 남아있음으로 해서 영상 상에서 밴드를 형성 시킬 수 있는 signature가 형성되지 않는 경우에 발생할 수 있다. 그러므로 추가적으로 탐색창의 크기를 줄여서 물체를 검출 하는데 있어 계산량을 줄이고 검출 확률을 높이는 것이 바람직하다.

Fig. 2는 signature parsing을 이용한 영상 내 물체 영역의 추출 예를 보여준다.



(a) Original (b) Segmented

Fig. 2 Example of region segmentation

### 3. 스네이크

#### 3.1 스네이크의 정의

스네이크는 1988년 Kass 등이 제안한 알고리즘으로써, 특정 영역이나 영상 해석 등을 독창적인 방법을 이용하여 접근한 방법으로, 그 거동의 특성에 의해 스네이크로 불리며 선분이나 경계추출, 물체 추적 등의 경우에 널리 이용되고 있다. 스네이크는 영상에 작용하는 힘(image force)의 영향 아래 움직이는 planar deformable contour로 적당한 에너지 함수  $E_{snake}(v)$ 를 부여해 줌으로써 정의된다. 컨투어  $v$ 는 스네이크의 형태를 표현하는 스플라인 곡선에 해당되며 일반적으로  $v$ 는 스플라인 곡선 위의 점들 또는 스플라인 곡선의 제어점들로 표현된다. Kass의 기본 모델은 파라메타로써 경계곡선 길이를 가지는 벡터  $v(s)=(x(s),y(s))$ 에 의해 표현된다. 스네이크의 경계를 따라 제안된 컨투어에 대한 에너지 함수는 다음 식(4)와 같이 정의된다.

$$E_{snake}(v) = \int E_{snake}(v(s))ds \quad (4)$$

$$= \int E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))ds$$

$E_{int}(v(s))$ 는 스네이크 경계의 구부러짐이나 불연속에 의한 경계의 내부 에너지를 표현한다. 또한  $E_{image}(v(s))$ 는 이미지 작용력으로서 선(line), 에지(edge)등의 여러 영상 특징에 의한 것이다. 마지막으로  $E_{con}(v(s))$ 는 외부에서 가해지는 제어 에너지이다. 이들 에너지는 다음 식들과 같이 정의된다.

$$E_{int}(v(s)) = (\alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2)/2 \quad (5)$$

$$E_{image}(v(s)) = -\gamma(s)|\nabla I(v(s))|^2 \quad (6)$$

식(5)의 첫번째 항은 컨투어 방향에 따라  $s$ 에 대한 1차 미분을 나타내고 경계 커브에 벌어짐이 있을 때 더 큰 값을 가지게 되어 제어점들을 뭉치게 하는 경향이 있다. 두번째 항은 컨투어 방향에 따라  $s$ 에 대한 2차 미분을 나타내며 컨투어의 곡률이 커질수록 더 큰 값을 가진다.  $\alpha(s)$ ,  $\beta(s)$ 는 각 에너지에 대한 가중치이며, 각각의 제어점 상에서 서로 다른 값을 가질 수 있고 컨투어가 특정 제어점에서 멀어지는 정도나 구부러지는 정도를 제어하는 역할을 한다.

### 3.2 Greedy algorithm

스네이크 에너지  $E_{snake}(v)$ 를 최소화 하는 방법에 있어서 variational calculus 방법은 수렴성의 측면에서 수치적인 불안정성을 가지므로 Amini<sup>(7)</sup>등은 동적 계획법(Dynamic programming)을 이용하여 수렴 안정성을 증가 시켰다. 그러나 동적 계획법은 계산량이 많다는 단점이 있어서 수렴 속도가 빠르지 못한 단점을 가지고 있다. Greedy algorithm<sup>(9)</sup>은 수렴의 안정성을 보장함과 동시에 프로세스가 진행되는 각각의 iteration에서 한 제어점에서 다음 단계의 위치를 결정하기 위하여 제어점 근방의 크기  $m$  만큼의 이웃 영역을 탐색하므로  $O(nm^2)$ 의 계산 복잡도를 가지는 동적 계획법 보다 빠른 수렴 특성을 가진다. Greedy algorithm은 수렴한 제어점이 전역 최소값(global minimum)이라는 확신을 주지는 못하나 실험적인 결과는 기존의 방법에 대응 된다. 그러나 이 역시 기존 스네이크 알고리즘이 가지고 있는 여러가지 단점을 포함 하므로 물체 추적에 있어 빠른 속도로 움직이는 물체에서는 스네이크가 경계점을 잃게 되어 추적에 실패할 경우가 생길 수 있다. Fig. 3은 제어점의 8 근방에서의 새로운 점의 탐색을 설명 한다.

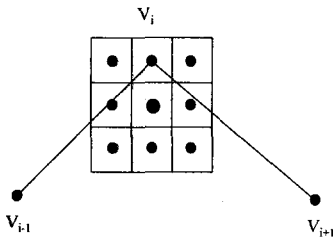


Fig. 3 8 neighbor searching

에너지 함수는 특정 시간에서 경계상의 한 제어점  $v_i$ 와 이의 8 근방으로부터 계산된다. 이때 가장 작은 에너지 값을 가지는 위치는  $v_i$ 의 새로운 위치점으로 정해 진다.

$$E(v_i) = \text{Min}(E_{int}(v_i) + E_{image}(v_i)) \quad (7)$$

내부 에너지 식(5)을 유한 차분(finite difference)의 형태로 나타내면 식 (8)과 같이 결정 된다.

$$E(v_i) = \alpha_i(v_{i+1} - v_i) + \beta_i(v_{i-1} - 2v_i + v_{i+1}) \quad (8)$$

식 (8)에서  $(v_{i+1} - v_i)$ 는 곡선이 축소(shrink) 하도

록 점들 사이의 거리를 최소화 시키며 이것은 경계상의 strong portion으로 뭉치도록 하는 특징이 있다. William<sup>(9)</sup>은 shrink 효과에 영향을 주는 일 없이 점들의 균일한 분포를 유도 하기 위하여 점들 사이의 평균거리  $d$ 와 식 (9)로 표현되는, 고려되는 두 점 사이의 거리인  $|v_{i+1} - v_i|$ 값을 최소화 시켰다.

$$d - |v_{i+1} - v_i| \quad (9)$$

이 방법에 의해 평균에 근사한 거리를 가지는 점들은 최소 에너지 값을 가지며, 각 iteration의 끝에서  $d$ 는 새로운 값으로 갱신 된다. Curvature 항은 현재의 제어점을 포함한 9 근방의 가장 큰 값으로 나누어 정규화 시켜 사용하며 이 값의 범위는 0과 1 사이의 값을 가지게 된다. 이미지 작용력  $E_{image}$ 는 gradient의 작용력으로 9 근방의 최대 최소값을 선택한 후 식(10)과 같이 정규화 시킨다. 식 (10)에서  $\nabla I_{cur}$ 는 9 근방 각 점에서의 gradient 값이다.

$$E_{image} = \frac{\nabla I_{\min} - \nabla I_{cur}}{\nabla I_{\max} - \nabla I_{cur}} \quad (10)$$

식(10)의 결과는 항상 음의 값이고 큰 gradient에서 더 작은 값을 가지게 되며 제어점을 그 위치로 끌게 된다.

### 4. 대응 문제(Correspondence Problem)

복수의 물체들 또는 특징점들을 포함하는 두개 이상의 영상으로부터 동일한 물체 또는 특징점(feature)을 판정하는(identify) 과정을 대응 문제(correspondence problem)이라 한다<sup>(10-12)</sup>. 본 연구에서는 이러한 대응 문제를 해결하기 위해 영상 내 물체간의 위치구조와 스네이크 에너지를 이용한 방법을 제시한다. N개의 물체를 포함하고 있는 경우 각각의 물체가 다음 영상에서는 각각 어떤 물체에 포함되는가를 풀기 위하여 에너지와 각 물체간 거리 정보를 이용하여 최소화되는 물체로 선정한다. 각 영상 시퀀스에서 물체간의 에너지 변화량을 다음과 같이 정의한다.

$$\Delta \text{Energy}(i, j) = \text{Energy}'(i) - \text{Energy}'^{-1}(j) \quad (11)$$

$(i = 1 \dots N, j = 1 \dots N)$

위 식에서 상첨자  $t$ 는 시간  $t$ 를 의미한다. 한편 물체간의 거리 변화량은 다음과 같이 정의한다.

$$\Delta Dis(i, j) = DIST(i^t, j^{t-1}) \quad (12)$$

$(i=1 \dots N, j=1 \dots N)$

위 식에서 DIST는 각 물체의 영상내 중심 위치간의 유클리드 거리를 의미하며 다음식과 같이 정의된다.

$$DIST(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (13)$$

매칭을 위해 다음과 같은 함수를 정의한다.

$$G(i, j) = w1 \times \Delta Energy(i, j) + w2 \times \Delta Dis(i, j) \quad (14)$$

위 식에서  $w1$ 과  $w2$ 는 각 변화량의 가중치를 의미한다. 위의 방법에 의하여 시간  $t$ 의 물체  $i$ 는 위의 함수  $G$ 를 최소화 하는 시간  $t-1$ 의 물체  $j$ 에 대응된다.

## 5. 실험

제안된 알고리즘의 타당성을 검토하기 위해 서로 다른 형상과 크기를 가진 두개의 추적 물체를 포함하는 영상 시퀀스에 대한 추적 결과를 검토한다. Fig. 4는 200x200 사이즈의, 궤도를 따라 운동하는 두개의 이동 물체를 포함한 영상 시퀀스를 보여준다. 각각의 물체의 평균속도는 95mm/s이다.

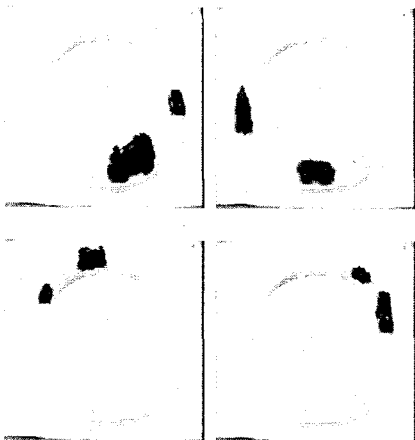


Fig. 4 Image sequence (Every 50 frames)

Fig. 5는 전처리와 영역 추출 알고리즘을 이용하여 추출된 영역 이미지로부터 Greedy 알고리즘을 적용한 스네이크의 형상을 보여준다. Fig. 5에서 추출된 부영역은 물체 주위의 사각 영역으로 나타내었으며 ■는 물체 1의 중심을, +는 물체 2의 중심을 나타내고 물체 주위의 점들은 수렴된 스네이크의 제어점을 표시한다. 앞 절에서 언급한 바와 같이 스네이크의 초기 제어점의 위치는 영역 추출 알고리즘으로부터 추출된 부영역의 경계상의 점들로 설정되었으며, 본 연구에서 적용한 스네이크의 제어점의 개수는 각 물체에 대하여 각각 12개로 설정하였다.

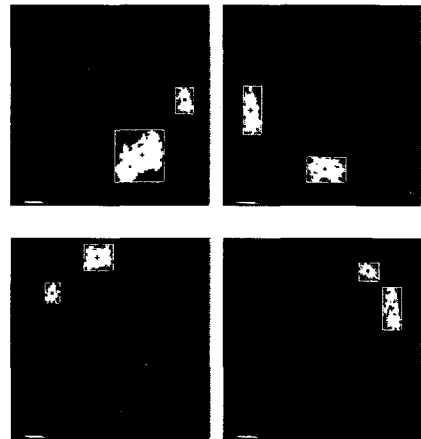


Fig. 5 Objects tracking(Every 50 frames)

Fig. 6과 Fig. 7은 물체 1과 물체 2에 대해 영상 좌표계에서의 수평 방향과 수직 방향의 추적된 물체 중심 위치 추이를 도시한 결과를 보여 준다. 그림에서 점선은 물체 1의 추이이며 실선은 물체 2의 추이를 나타낸다.

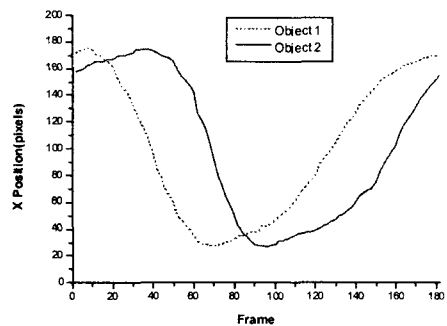


Fig. 6 Center position in image coordinate(1)

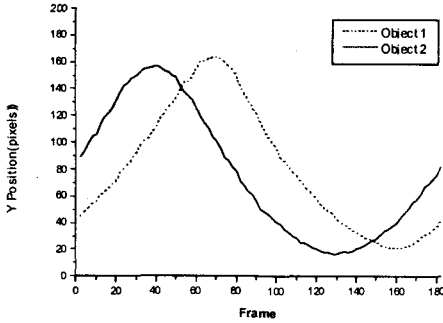


Fig. 7 Center position in image coordinate(2)

Fig. 8은 스톱 에너지 추이 곡선을 나타낸다. 75~110 frame에서의 두 물체간의 스톱 에너지 준위차는 많이 생기지 않는다. 그 외의 구간에서는 두 물체간의 식별이 가능케 하는 에너지의 차이가 생긴다. 이러한 에너지의 차이는 영상내의 gradient의 크기와 조명조건, 추적체의 형상, 추적체의 크기등과 관련이 있으며, 이러한 에너지 추이와 영상 내 위치 구조의 추이에 의해 두 물체의 대응 문제를 해결하였다.

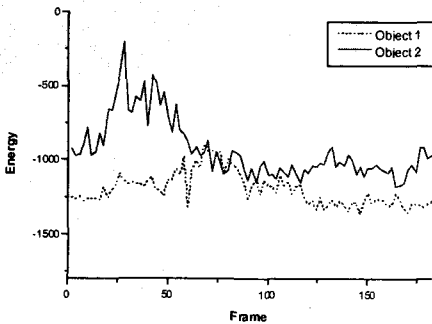


Fig. 8 Snake energy

## 6. 결론

본 연구에서는 비전 시스템을 이용하여 임의의 형상을 가진 이동 물체의 추적 시스템을 구성하였으며, 특히 스톱 에너지의 준위와 영상 내 물체들의 위치 구조를 이용하여 다중 물체의 대응 문제의 해결 방안을 제시하였다.

영상으로부터 추적하고자 하는 물체를 추출하기 위하여 spatio-temporal 필터와 signature parsing 방법을 사용하였으며 스톱 에너지의 초기 제어점을

추출된 영역의 경계상으로 설정하여 인위적인 조작을 피했다. 제시된 알고리즘의 타당성을 검토하기 위해 실시간으로 이동하는 두개의 서로 다른 물체 추적 실험을 수행하였으며 그 결과 안정적으로 서로 다른 물체를 구별하며 추적할 수 있었다. 두 물체의 평균 이동 속도는 75 mm/sec 이고 본 알고리즘의 처리 속도는 14 frame/sec의 결과를 얻었다.

## 참고문헌

- (1) I. Weiss and M. Ray, 2001, "Model Based Recognition of 3D Object from Single Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No.3, PP.116-28.
- (2) G. Jones, and B. Bhaun, 1999, "Recognition of Articulate and Occluded Objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No.7, PP.603-613.
- (3) F. Chaumette, 1994, "Visual Servoing Using Image Features Defined Upon Geometric Primitives", *Proc. of the 32<sup>nd</sup> Conference on Decision and Control*, Vol. 4, PP. 3782-3787.
- (4) J. R. Jain and A. K. Jain, 1981, "Displace Measurement and Its Application in Image Coding", *IEEE Transaction in Communication*, Vol. 29, No. 12, PP. 1799-1808.
- (5) M. Kass, A. Witkin and D. Terzopoulos, 1988, "Snakes : Active Contour Models", *International Journal of Computer Vision*, Vol.1, No.4, PP.321-331.
- (6) F. Leymarie and M. D. Levine, 1993, "Tracking Deformable Objects in the Plane Using an Active Contour Model", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14, No.6, PP. 617-634.
- (7) A. A. Amini, T. E. Weyouth and R. C. Jain, 1990, "Using Dynamic Programming for Solving Variational Problems in Vision", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 6, PP. 564-569.
- (8) D. Murray and A. Basu, 1993, "Motion Tracking with an Active Vision", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 5, PP.449-459.
- (9) D. J Williams M. Shar, 1990, " A Fast Algorithm for Active Contours", *Proc. of International conference on Computer Vision*, PP. 592-595.
- (10) R. Jain, R. kasturi and B. Schunck, 1995, *Machine Vision*, McGraw Hill.
- (11) H. S. Lim and T. O. Binford, 1988, "Structural Correspondence in Stereo Vision", *Image Understanding Workshop*, PP. 798-808.
- (12) D. H. Ballard and C. M. Brown, 1986, *Computer Vision*, Prentice Hall.