

논문 2005-42SP-6-1

# Mean-Shift의 색 수렴성과 모양 기반의 재조정을 이용한 실시간 머리 추적 알고리즘

(A Real-Time Head Tracking Algorithm Using  
Mean-Shift Color Convergence and Shape Based Refinement)

정 동 길\*, 강 동 구\*\*, 양 유 경\*\*\*, 나 종 범\*\*\*\*

(Dong-Gil Jeong, Dong-Goo Kang, Yu Kyung Yang, and Jong Beom Ra)

## 요 약

이 논문에서는 팬-틸트-줌 기능을 가지는 실시간 능동카메라 시스템에 적합한 2단계 머리 추적 알고리즘을 제안한다. 먼저, 색 수렴 단계에서는 머리의 모양을 타원으로 가정하고 모델 색-히스토그램을 얻는다. 그 후, 모델과 후보 타원의 색-히스토그램 간의 유사도를 검사하여 목표 물체의 대략적인 위치를 구하기 위해 mean-shift 방법을 이용한다. 여기에서 영상 내 물체 영역의 색 분포가 카메라의 관찰 방향에 따라 달라지는 것을 고려하기 위하여, 모델 히스토그램 뿐 아니라 이전 프레임에서 얻어진 타원의 색 히스토그램도 함께 고려함으로써 mean-shift의 수렴성을 향상시킨다. 특히, 이전 프레임에서 결정된 타원 내부의 가장자리 영역에 포함되어 있는 배경 색 성분에 의한 오류 누적 문제를 해소하기 위해, 모델 히스토그램을 이용하여 타원의 크기를 적응적으로 축소함으로써 이전 추적 결과중 머리 영역에 해당되는 색 히스토그램을 얻는다. 또한 영상 내의 전역 움직임을 예측하고 이를 보상하여 정확한 초기 위치를 찾음으로써 mean-shift의 색 수렴성을 더욱 향상시킨다. 이 때, 고속 움직임 추정을 위해 1-D 투사 데이터 기반의 방법을 제안한다. 다음 단계에서는, 모양 정보를 이용하여 수렴단계에서 얻어진 타원의 위치와 크기를 보다 정확히 재조정한다. 이를 위해 영상 내 경사도의 방향에 기반한 강건한 모양 유사도 함수를 정의하고 사용한다. 다양한 환경을 고려한 실험을 통하여, 사람의 움직임이 빠른 경우, 영상 내 머리 크기의 변화가 심한 경우, 그리고 배경의 색과 모양이 매우 복잡한 경우에 대해서도 제안한 알고리즘이 비교적 정확히 추적을 수행함을 보였다. 아울러 제안한 알고리즘은 추적을 수행하는데 일반 PC에서 약 30fps의 처리 속도를 보여 실시간 시스템에 적합하다.

## Abstract

In this paper, we propose a two-stage head tracking algorithm adequate for real-time active camera system having pan-tilt-zoom functions. In the color convergence stage, we first assume that the shape of a head is an ellipse and its model color histogram is acquired in advance. Then, the mean-shift method is applied to roughly estimate a target position by examining the histogram similarity of the model and a candidate ellipse. To reflect the temporal change of object color and enhance the reliability of mean-shift based tracking, the target histogram obtained in the previous frame is considered to update the model histogram. In the updating process, to alleviate error-accumulation due to outliers in the target ellipse of the previous frame, the target histogram in the previous frame is obtained within an ellipse adaptively shrunken on the basis of the model histogram. In addition, to enhance tracking reliability further, we set the initial position closer to the true position by compensating the global motion, which is rapidly estimated on the basis of two 1-D projection datasets. In the subsequent stage, we refine the position and size of the ellipse obtained in the first stage by using shape information. Here, we define a robust shape-similarity function based on the gradient direction. Extensive experimental results proved that the proposed algorithm performs head tracking well, even when a person moves fast, the head size changes drastically, or the background has many clusters and distracting colors. Also, the propose algorithm can perform tracking with the processing speed of about 30 fps on a standard PC.

**Keywords:** Head tracking, mean-shift, color histogram similarity, gradient-based shape similarity

\* 정회원, 국방과학연구소

(Agency for Defense Development)

\*\* 학생회원, \*\*\*\* 평생회원, 한국과학기술원 전자전산학과, 영상정보특화연구센터

(Department of EECS and Image Information Research Center, Korea Advanced Institute of Science and Technology)

\*\*\* 정회원, KTF Technologies

(KTF Technologies)

※ 본 연구는 한국과학기술원 영상정보특화연구센터를 통한 국방과학연구소의 연구비 지원으로 수행되었습니다.  
접수일자: 2005년5월9일, 수정완료일: 2005년9월22일

## I. 서 론

객체 추적은 연속한 영상들로부터 관심 객체의 위치와 크기 등을 알아내는 것으로서 사람을 추적대상으로 한 응용 분야는 감시 시스템(surveillance system), HCI(human computer interaction), 가상현실(virtual reality) 등이 있다. 특히 최근에는 신분확인, 표정인식, 행동인식 등의 분야에 대한 연구가 활발해지면서 이를 위한 전처리 과정으로서 머리 영역의 추적이 중요하게 되었다. 특히, 팬-틸트(pan-tilt) 회전 및 배율제어가 가능한 능동 카메라를 사용하면 추적 대상을 영상에서 원하는 크기와 원하는 위치에 오도록 조정할 수 있어 정보량을 극대화 할 수 있다. 능동 카메라를 사용한 추적 시스템에서는 현재의 추적 결과에 의하여 배율 및 팬-틸트 제어가 수행되므로 효율적이고 강건한 제어를 위해서는 추적대상의 위치와 크기를 빠르고 정확하게 검출할 수 있는 추적 알고리즘이 요구된다.

기존의 알고리즘들은 추적에 사용되는 객체의 특징값(feature)에 따라 움직임(motion), 모양(shape) 또는 색(color) 정보를 사용하는 방법으로 나눌 수 있다.

움직임 정보를 이용한 방법<sup>[1-2]</sup>은 이동 물체의 고유한 움직임을 규정하고 이를 추적하는 방법으로서 규칙적인 움직임을 갖는 물체에 대해 강건한 결과를 보인다. 그러나 추적 대상이 정지해 있거나 부분적인 움직임을 가질 경우 적용에 한계가 있고, 팬-틸트 회전 및 배율제어가 빈번히 발생하는 경우에는 이에 따른 영상 내의 전역적 움직임과 이동체의 국부적 움직임을 정확히 구분할 수 있어야 하므로, 움직임 추정에 많은 계산량이 요구되어 별도의 하드웨어의 도움 없이 실시간 시스템에 적용하기 어렵다.

모양 정보를 이용하는 방법<sup>[3]</sup>에서는 주로 객체의 윤곽선 정보를 이용하여 이를 가장 잘 표현할 수 있는 윤곽선 모델을 찾아낸다. 이 방법의 가장 큰 장점은 윤곽선 수준의 매우 정확한 추적 결과를 얻을 수 있다는 것이다. 그러나 배경이 복잡한 경우 탐색영역이 클수록 배경에 존재하는 강한 에지 등에 이끌려 잘못된 결과를 얻기 쉽다. 즉, 복잡한 배경에서는 모양 유사도 곡선에 많은 국부 최대값이 존재하여 전역 최대값을 보장하는 수렴 범위(capture range)가 짧다. 따라서 최적치를 보장하기 위해서는 모든 탐색점을 검색하는 전역 탐색(full search)을 수행하는 것이 바람직하나 이는 많은 계산량을 요구하기 때문에 실시간 처리에 부적합하다.

색 정보를 이용하는 방법에서는 추적 대상이 가지는 색 성분을 예측하여 미리 색 모델을 만들고 이와 가장

유사한 색을 가지는 영역을 검색한다. 색 정보를 이용한 방법의 장점은 모양정보를 이용한 방법에 비해 수렴 범위가 크므로 효율적인 최적화가 가능하다는 것이다. 특히, Comaniciu<sup>[4-6]</sup>는 색 히스토그램의 유사성을 측정하기 위해 Bhattacharyya 계수를 사용하고 이를 최대화하는 위치를 추적하기 위해 mean shift라는 최적화 방법을 도입함으로써 몇 번의 반복(iteration)만으로 추적 대상의 중심 위치로 수렴할 수 있음을 보였다. 하지만 능동카메라 시스템에서는 카메라와 사람의 움직임에 따라 추적 대상의 관찰 각도가 시간에 따라 변하므로 미리 정해놓은 모델이 객체의 색 특성을 제대로 반영하지 못하는 문제가 있다. 또한 모양 기반의 방법과 달리 객체의 정확한 윤곽선을 얻기 어려워 신속하고 정확한 카메라 렌즈의 배율 제어를 어렵게 한다.

Birchfield<sup>[7]</sup>는 색과 모양정보를 동시에 이용하였다. 이 방법에서는 타원의 크기와 위치를 바꿔가면서 타원 내부의 색 히스토그램과 모델 색-히스토그램의 유사한 정도(color similarity)와 타원 경계 위의 경사도(gradient)크기의 합으로 표현되는 모양 유사도(shape similarity)를 합한 새로운 유사도를 일정한 검색 영역에서 모두 계산하여 가장 타당한 후보를 선택하였다. 이 때, 모양과 색 정보를 동시에 고려하므로 보다 정확한 검출 결과를 제공하지만 모양 유사도에 의한 국부 최대값의 영향을 피하기 위한 전역검색(full search)방법을 택하였기 때문에 많은 계산량이 소모된다.

이 논문에서는 색 정보와 모양 정보를 단계적으로 사용하여 정확한 추적을 가능하게 할 뿐 아니라 실시간 추적을 위해 적은 계산량을 가진 두 단계 알고리즘을 제안한다. 첫 번째 단계에서는 mean-shift의 색 수렴성을 이용하여 적은 계산량으로 표적에 가장 가까운 위치로 타원을 수렴시킨다. 이 때, 색 수렴성을 향상시키기 위하여 정확한 초기 파라미터의 설정을 위한 전역움직임 보상기법을 제안하며, 또한 강인한 추적을 위해 시간에 따라 변하는 객체의 색 정보를 mean-shift framework에 반영하기 위한 기법을 제안한다. 두 번째 단계에서는 보다 정확한 예측을 위해 수렴된 위치 주변으로 제한된 구간에 대해 타원의 위치 및 크기의 재조정을 위한 모양기반의 검색을 수행한다. 이를 위해, 경사도의 방향성에 기반을 둔 보다 강인한 모양 유사도 함수를 제안한다.

이 논문의 구성은 다음과 같다. II장에서는 제안하는 알고리즘을 상세히 설명하며 III장에서는 정성적, 정량적 비교 실험을 통해 제안한 알고리즘의 정확성을 보인다. 마지막으로 IV장에서는 결론을 내린다.

## II. 제안하는 알고리즘

### 1. 알고리즘의 개요

머리 모양은 사람마다 그 변화가 크지 않으며 변형되지 않는(rigid) 특징을 가진다. 따라서 이 알고리즘에서는 머리 모양을 타원으로 가정하며, 추적이란 이전 프레임에서 결정된 타원을 기반으로 하여 현재 프레임의 중심좌표와 크기를 결정하는 과정이다.

제안하는 알고리즘은 그림 1에서와 같이 크게 색 기반의 위치 검색 단계와 모양 기반의 재조정 단계로 나뉜다. 색 기반 위치 검색에서는 이전 프레임에서 얻은 결과 타원으로부터, 색 정보에 기반 한 mean shift 방법을 이용하여 현재 프레임에서의 추적대상의 중심 위치로 수렴해 가는 단계이다. 모양 기반 재조정 단계에서는 색 기반 위치 검색 단계에서 수렴된 위치를 초기 위치로 하여 영상의 경사도가 큰 부분이 타원의 모양에 가장 근접할 때의 타원의 위치와 크기를 구한다. 이 때, 이미 수렴된 타원에 대해 재조정하는 것이므로 매우 제한된 공간을 전역 탐색함으로써 계산량을 최소화 할 수 있다.

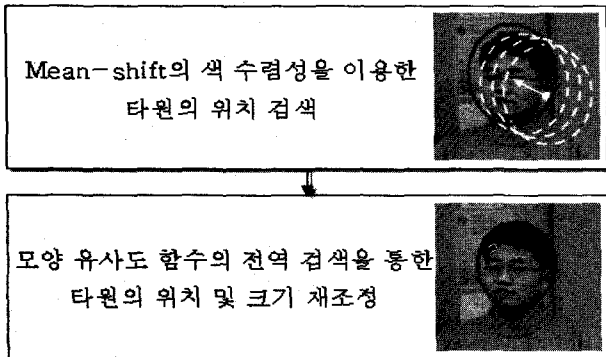


그림 1. 2단계 구조를 갖는 제안한 머리 추적 알고리즘  
Fig. 1. The proposed head tracking algorithm of two-stages.

#### 가. 색 기반 위치 검색

다음 식은 모델 색 히스토그램과 후보영역의 색 히스토그램의 유사도 함수를 Bhattacharyya계수<sup>[8]</sup>를 이용하여 표현한 것이다.

$$\rho_{Model}(\mathbf{y}) = \sum_{u=1}^M \sqrt{\hat{p}_u(\mathbf{y})q_u} \quad (1)$$

이 식에서  $\mathbf{y}$ 는 현재 영상에서 후보 타원의 중심 위치  $(x_c, y_c)$ 를 의미하고  $\hat{p}_u(\mathbf{y})$ 와  $q_u$ 는 각각 후보 타원과 모델의 색 히스토그램에서의  $u$ 번째 bin의 확률을

의미하며  $M$ 은 총 bin 수를 나타낸다. 사용한 색 공간은 B-G, G-R, R+B+G이며, 색 히스토그램의 bin 수는 각각 32, 32, 4를 사용하였다<sup>[7]</sup>. 이때  $\rho_{Model}(\mathbf{y})$ 가 최대가 되는  $\mathbf{y}$ 를 찾기 위해 mean-shift 최적화 기법을 적용한다. 그러나 이를 적용하기 위해서는  $\rho_{Model}(\mathbf{y})$ 의 수식이 커널(kernel)의 가중치의 합으로 표현되어야 한다<sup>[5]</sup>. Comaniciu는  $\hat{p}_u(\mathbf{y})$ 를 식 (2)와 같이 특정 커널의 합의 꼴로 정의하여 이를 식 (1)에 대입한 후 테일러 정리를 적용하여 식을 근사화 및 재정리함으로써  $\rho_{Model}(\mathbf{y})$ 를 mean-shift 최적화에 적용할 수 있는 식, 즉 커널의 가중치 합으로 표현되는 식 (3)으로 유도하였으며<sup>[5]</sup> 이를 색 히스토그램 유사도(CHS)라 한다.

$$\hat{p}_u(\mathbf{y}) = C_h \sum_{i=1}^N k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right) \delta[b(\mathbf{x}_i)-u] \quad (2)$$

$$CHS_{Model} = \sum_{i=1}^N \omega_i k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right) \quad (3)$$

위 식에서,  $k(\cdot)$ 는 convex 하며 단조 감소하는 커널의 프로파일(profile)을 의미하며  $h$ 는 커널의 대역폭(bandwidth)을,  $N$ 은 커널 안에 존재하는 픽셀들의 수를,  $\delta(\cdot)$ 는 Kronecker delta function을,  $b(\mathbf{x}_i)$ 는 위치  $\mathbf{x}_i$ 에서의 색의 bin 인덱스를,  $C_h$ 는 정규화 상수를 각각 의미하고,  $\omega_i$ 는 다음과 같이 쓸 수 있다.

$$\omega_i = \sum_{u=1}^M \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\mathbf{y}_0)}} \delta[b(\mathbf{x}_i)-u] \quad (4)$$

결국, mean-shift 최적화 이론에 의해 식 (3)을 최대화 하는 타원의 위치  $\mathbf{y}$ 는 다음의 식에 따라 수렴되는 것이 알려져 있다<sup>[5],[6]</sup>.

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^N \omega_i \mathbf{x}_i g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^N \omega_i g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\right\|^2\right)} \quad j=1,2,\Delta \quad (5)$$

Mean-shift 최적화 기법은 최소한의 계산량으로 주어진 유사도 함수를 최대화하는 해의 수렴성을 보장한다고 알려져 있지만, 이는 가장 가까운 주변 국부 최대값으로 수렴함을 의미하기 때문에, 초기점이 부정확하거나 유사도 함수가 부적절하게 정의된 경우 올바른 해

로 수렴하지 못할 수 있다. 더구나 mean-shift를 이용한 색 기반 위치 검색의 정확도가 전체 알고리즘의 성능에 큰 영향을 미칠 것이다. 따라서 이 논문에서는 올바른 해를 제공하기 위해 고려해야 할 색 유사도 함수와 초기 위치 설정에 관한 효과적인 기법을 제안한다.

(가) 시간적 색 변화에 강인한 색 유사도 함수의 생성  
일반적으로 사람의 머리색은 머리카락과 얼굴색으로 표현된다. 그러나 카메라가 머리를 관찰하는 각도에 따라 이들 색의 비(ratio), 즉 색 히스토그램의 분포가 점진적으로 변하기 때문에 미리 설정된 모델의 색 히스토그램만을 이용할 경우 이를 반영하지 못하여 강인한 추적을 보장할 수 없다. 따라서 이 논문에서는, 모델 히스토그램 뿐 아니라 이전 프레임에서 추출된 타원 내부의 색-히스토그램도 함께 고려하여 색 유사도 함수를 정의하고자 한다. 그러나 이전 프레임에서 추적된 타원의 크기가 실제 머리의 윤곽선보다 크게 예측되는 경우, 타원 경계의 배경색이 포함되어 부정확한 추적결과를 가져오며, 이러한 오류는 추적이 진행되면서 누적 된다. 따라서 이러한 오류 누적의 문제를 해소하기 위하여, 타원의 크기를 충분히 축소시키는 것이 중요하며 다음 식에 따라 타원의 크기를 조정한다.

$$\eta' = r \times \eta \quad (6)$$

이 식에서  $\eta$  와  $\eta'$  는 각각 이전 프레임에서 추적된 타원 단축 길이와 축소된 타원 단축 길이를 나타내며  $r$  은 0에서 1 사이의 실수이다.  $r$  을 결정하기 위해, 모델 색에 포함되지 않는 색이 배경색임을 가정하여, 이전 프레임에서의 최종 추적된 타원의 색 히스토그램과 모델의 색 히스토그램의 Bhattacharyya 계수를 구하여 이 값이 작을수록  $r$  을 작게 조정한다. 그림 2 는 이전 프레임에서의 타원 크기가 실제보다 크게 예측된 경우 적

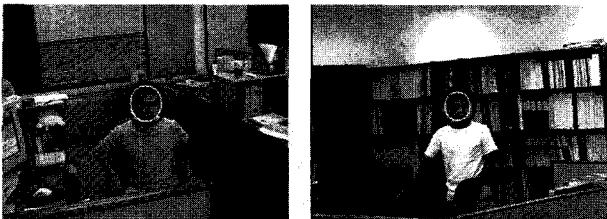


그림 2. 오류가 제거된 색 히스토그램을 추출하기 위해 타원 크기를 조정한 두가지 예. 검정색의 타원을 수축하여 흰색의 타원으로 결정한다.

Fig. 2. Two examples of ellipse size adjustment for proper color histogram extraction. The white ellipse is determined by shrinking the black ellipse.

절히 수축된 예를 보여준다. 결과적으로, 현재 프레임에서 위치  $y$  의 타원 후보에 대한 색 유사도 함수를 다음과 같이 정의한다.

$$\rho(y) = \alpha \rho_{Model}(y) + (1 - \alpha) \rho_{Temporal}(y) \quad (7)$$

이 식에서,  $\rho_{Temporal}(y)$  는 축소된 타원과 후보 타원의 색 히스토그램들의 Bhattacharyya 계수를 의미하며,  $\alpha$  는 두 계수들의 가중치 상수로서 실험적으로 결정된 값(0.2)을 사용하였다. 이 유사도 함수는 mean-shift에 적용 가능한 다음의 식으로 표현할 수 있다.

$$CHS \equiv \sum_{i=1}^N \omega_i^T k \left( \left\| \frac{y - x_i}{h} \right\|^2 \right) \quad (8)$$

여기에서,

$$\omega_i^T = \sum_{u=1}^M \left[ \alpha \sqrt{\frac{\hat{q}_u}{\hat{p}_u(y_0)}} + (1 - \alpha) \sqrt{\frac{\hat{v}_u}{\hat{p}_u(y_0)}} \right] \delta[b(x_i) - u] \quad (9)$$

이 식에서  $\hat{v}_u$  는 이전 히스토그램에서의  $u$  번째 bin의 확률 값이며  $\alpha$  는 가중치 상수이다.

따라서 다음 식에 의해 타원의 위치가 결정된다.

$$y_{j+1} = \frac{\sum_{i=1}^N \omega_i^T x_i}{\sum_{i=1}^N \omega_i^T} \quad j = 0, 1, 2, \dots, K \quad (10)$$

(나) 수렴성 향상을 위한 고속 전역움직임 보상 기법  
Mean-shift를 이용한 최적화 방법이 해의 수렴성을 보장한다는 것은 이미 증명 되었다<sup>[5]</sup>. 하지만 이는 주변의 가장 가까운 국부 최대값으로 수렴함을 의미하기 때문에 초기 위치에 따라 추적 대상에 수렴하지 못하는 경우가 발생한다. 따라서 mean-shift 방법에서 전역 최대값에 수렴하기 위해서는, 그림 3에서 보여진 수렴영역을 나타내는 타원의 내부에 타원의 초기점이 위치할 때만 올바른 해를 보장할 수 있다. 즉, 초기위치가 A로 결정되면 mean-shift과정에서 최초의 후보 타원이 표적과 겹치는 부분이 있기 때문에 표적에 수렴할 수 있지만, B의 경우에는 수렴이 보장 되지 않는다.

한편, 팬-틸트 능동 카메라는 표적의 움직임에 따라 회전하도록 제어가 된다. 이때 표적의 이동 방향이 급격히 바뀌는 경우, 카메라는 제어 속도의 한계로 인해 이전 추적 방향으로 계속 회전하게 되므로 영상 내의 전역 움직임이 표적의 이동 방향에 비하여 반대가 되어 mean-shift의 시작점이 수렴 영역을 벗어날 수 있다.

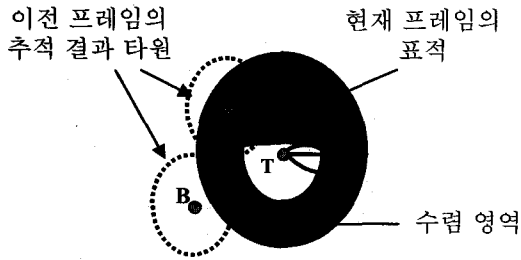


그림 3. Mean-shift의 수렴 영역  
Fig. 3. Convergence area in the mean-shift.

따라서 영상의 전역 움직임을 추정하여 이를 보상함으로써 mean-shift의 시작점의 예측 오류를 보정하여 초기 위치가 수렴 영역에 포함되도록 즉, 올바른 해로 수렴하도록 유도한다. 그러나 전역 움직임 예측 시, 넓은 영역을 탐색해야 하므로 실시간 추적을 위해서는 계산량이 적은 움직임 추정 기법이 필요하다. 고정된 축을 갖는 팬-틸트-줌 능동 카메라의 운동에 의한 영상에서의 전역 움직임은 근사적으로 다음과 같이 모델링 된다<sup>[9]</sup>.

$$u = -f\Omega_Y + \beta x + \frac{\Omega_X}{f}xy - \frac{\Omega_Y}{f}x^2 \tag{11}$$

$$v = f\Omega_X + \beta y + \frac{\Omega_X}{f}y^2 - \frac{\Omega_Y}{f}xy \tag{12}$$

여기에서  $f, \Omega_X, \Omega_Y, \beta$  는 각각 초점 거리, 틸트와 패닝 각도 차이, 배율을 의미한다. 여기에서  $\Omega_X, \Omega_Y \ll f$  라고 가정하면, 위의 식은 다음과 같이 다시 표현된다.

$$u = \alpha_1 + \beta_x x \tag{13}$$

$$v = \alpha_2 + \beta_y y \tag{14}$$

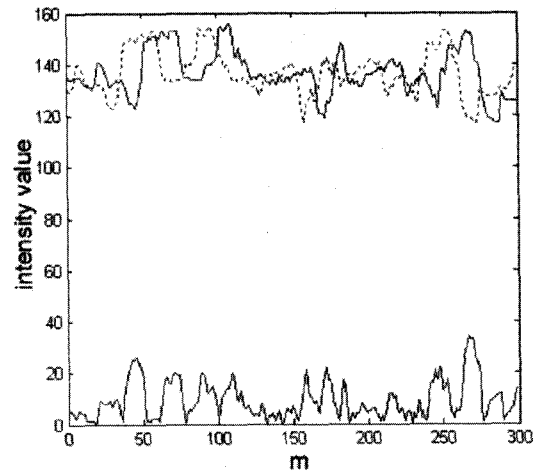
여기에서  $\alpha_1$ 과  $\alpha_2$ 는 각각  $-f\Omega_Y$  와  $-f\Omega_X$  이다. 위 파라미터를 예측하기 위해 먼저 다음과 같이  $x, y$  방향으로 투사(projection)된 데이터를 구한다.

$$P_i^X(m) = \sum_{n=1}^{N_h} I_i(m, n) \tag{15}$$

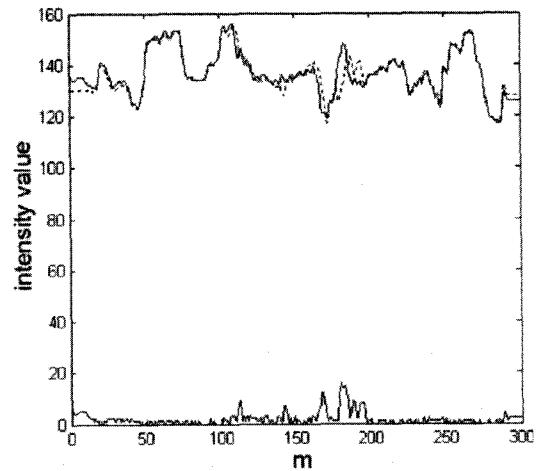
$$P_i^Y(n) = \sum_{m=1}^{N_w} I_i(m, n) \tag{16}$$

여기에서  $I_i(m, n)$  은  $N_w \times N_h$  크기의  $i$  번째 프레임이다. 다음으로 이전 프레임에서의 각 투사 데이터로부터 두 개씩의 데이터 세그먼트를 선택한다. 이 때 움직이는 객체의 독립적인 국부 움직임은 전역 움직임 예측의 성능을 저하 시킬 수 있으므로 이전 추적 결과

타원의 투사된 위치로부터 멀리 떨어진 세그먼트를 선택함으로써 예측 오류를 최소화하였다.  $X_1$ 과  $X_2$ 를  $P_i^X$ 로부터 얻어진 세그먼트의 중심이라 하고 대응되는 움직임 벡터를 각각  $MV_{X_1}$ 과  $MV_{X_2}$ 라 하자. 마찬가지로,  $Y_1$ 과  $Y_2$ 를 각각  $P_i^Y$ 로부터 얻어진 두 세그먼트의 중심이라 하고 대응되는 움직임 벡터를 각각  $MV_{Y_1}$ 와



(a)



(b)

그림 4. 수평 방향을 따라 얻어진 투사데이터 비교. 연속 프레임간의 전역 움직임을 보상하기 (a) 전과 (b) 후. 각 그림에서 위쪽의 점선과 실선은 각각 이전 및 현재 프레임에서의 투사 데이터를 나타내며, 아래쪽의 실선은 그 차이를 보여준다.

Fig. 4. Two projection data of the consecutive frames obtained along the horizontal direction (a) before and (b) after global motion compensation. Dotted and solid lines on the top of the graphs denote the projection data of the previous and current frames, respectively, and their absolute difference is given at the bottom.

$MV_{Y_2}$ 라 하면, 이들 움직임 벡터는 간단한 1-D 정합으로 구할 수 있다. 따라서, 데이터와 탐색 공간이 모두 1차원에 수행되므로 탐색범위가 넓어도 매우 적은 계산량으로 처리가 가능하다. 결국 중심점과 대응되는 움직임 벡터를 식 (13)과 (14)에 대입하여 풀면 다음과 같은 결과를 얻는다.

$$\alpha_1 = \frac{X_1 MV_{x_2} - X_2 MV_{x_1}}{X_1 - X_2} \quad (17)$$

$$\alpha_2 = \frac{Y_1 MV_{y_2} - Y_2 MV_{y_1}}{Y_1 - Y_2} \quad (18)$$

$$\beta_x = 1 + \frac{MV_{x_1} - MV_{x_2}}{X_1 - X_2} \quad (19)$$

$$\beta_y = 1 + \frac{MV_{y_1} - MV_{y_2}}{Y_1 - Y_2} \quad (20)$$

이 결과를 이용하여 보정된 초기 위치는 다음과 같이 표현된다.

$$i_x = \alpha_1 + \beta_x t_x \quad (21)$$

$$i_y = \alpha_2 + \beta_y t_y \quad (22)$$

여기에서  $(i_x, i_y)$ 는 보정된 후의 후보타원의 초기 위치이고  $(t_x, t_y)$ 는 이전 프레임에서의 추적 타원의 중심이다. 그림 4는 어떤 연속 프레임에 대하여 전역 움직임 추정 전과 후의 투사 데이터를 비교해서 보여준다. 움직임 추정 후 데이터의 중간 부분에서 어느 정도의 차이를 보이는 것은 이동 물체의 고유 움직임에 기인한 것이다.

#### 나. 모양 기반 재조정 단계

모양 기반 재조정 단계에서는 색 기반 수렴 단계에서 수렴된 위치를 중심으로 제한된 영역 내에서 타원의 크기와 위치를 보다 정밀하게 찾아낸다. 정확한 예측을 위해 객체의 윤곽선 정보를 이용하며, 이미 수렴된 위치 주변의 국부적 영역만을 탐색하므로 계산량의 부담이 줄어들 뿐 아니라 배경이 복잡한 경우에도 복잡한 경사도로부터 영향이 적게 된다. 이 단계에서는 영상 내 머리의 윤곽선과 후보 타원의 윤곽선들 간의 유사도를 정의한다. 이를 위해 우리는 후보 타원 경계부근의 경사도 성분들이 얼마나 타원과 가까운지를 측정하는 척도를 제안한다. 제안한 모양 유사도 함수의 주요 특징은 다음과 같다.

첫째, 영상의 경사도 값을 이진화하여 사용하였다. 사

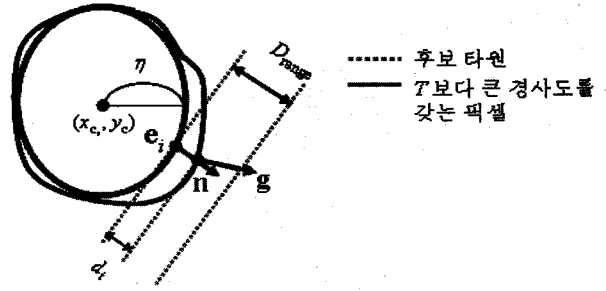


그림 5. 경사도를 이용한 모양 유사도 관련 파라미터들  
Fig. 5. Parameters regarding the shape similarity measure using image gradients.

람 머리의 경계에 해당하는 경사도가 배경에 존재하는 그것보다 항상 크다고 보장할 수 없다. 현재 머리의 위치가 어느 배경 위에 있느냐에 따라 배경의 경사도가 더욱 큰 값을 가질 수 있고, 머리 경계의 일부분에만 경사도가 큰 경우도 있다. 따라서 경사도의 크기를 그대로 사용하면 주변의 큰 경사도와 타원의 일부분에 존재하는 큰 경사도에 귀착하는 문제가 발생한다. 따라서 경사도의 값을 이진화하여 크기보다는 방향성을 고려하였다.

둘째, 타원 가장자리의 픽셀만 고려하지 않고 타원으로부터 법선 벡터 방향으로 일정한 검색영역을 설정하여 가장 가까운 경사도를 고려하였다. 머리의 모양이 정확히 타원에 일치하지 않기 때문에 타원의 위의 픽셀만 고려하면 잘못된 후보가 선정될 가능성이 크기 때문이다.

셋째, 타원의 법선 벡터와 경사도 벡터가 이루는 각이 일정한 문턱 값(threshold)보다 작은 픽셀의 수,  $N_v$ 를 반영하여 타원 경계부근의 경사도가 전체적으로 타원의 모양을 이루는 후보를 선택 한다. 이는 일부분에만 경사도가 있는 것 보다는 전체적으로 타원의 모양을 이루는 후보를 선정하기 위한 가중치이다.

다음 식은 제안된 모양 유사도 함수를 표현한다.

$$\rho_s(s) = \left(\frac{N_v(s)}{N_\sigma}\right)^2 \cdot \frac{1}{N_\sigma} \sum_{i=1}^{N_\sigma} (|g(p_i(s)) \cdot n(e_i(s))|) \times \left(1 - \frac{d_i(s)}{D_{range} + 1}\right) \quad (23)$$

여기에서  $s$ 는 임의의 타원의 위치  $(x_c, y_c, r)$  및 크기  $r$ 를 나타내며,  $N_\sigma$ 은 타원 경계의 샘플 수를,  $N_v$ 는 각 경계 샘플에서 타원의 법선 벡터와 영상의 경사도 벡터가 이루는 각이 일정한 문턱값  $T$  보다 작은 픽셀의 수를,  $e_i(s)$ 는 타원 경계의  $i$ 번째 샘플 위치를,  $p_i$ 는  $e_i$ 로부터 법선 방향으로 가장 가까우며 동시에 경

사도 벡터의 크기가 문턱 값  $T$ 보다 큰 픽셀 위치를,  $d_i$  는  $P_i$ 와  $e_i$  사이의 거리를 의미한다.  $n(e_i(s))$  는  $e_i$  에서 타원 윤곽선의 단위 법선 벡터를,  $g(p_i(s))$  는  $P_i$ 에서 단위 경사도 벡터를 의미한다.  $D_{range}$  안에 해당하는  $P_i$ 가 없는 경우에는  $g(p_i(s))=0$  이다.

### III. 실험 결과

객체의 움직임이 빠른 경우, 객체의 크기가 변화가 심한 경우, 배경 영역이 복잡한 경우 등 다양한 경우를

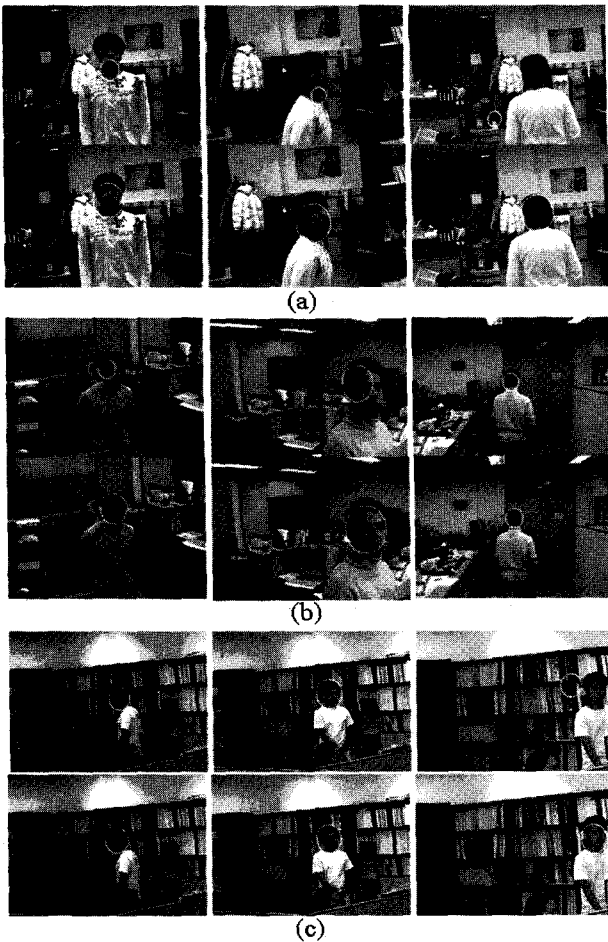


그림 6. 다양한 환경에서의 추적 결과. (a) 움직임이 매우 빠른 경우, (b) 영상 내 머리의 크기가 변하는 경우, (c) 배경이 복잡한 경우. 각 경우, 위와 아래 행은 Comanicu의 방법과 제안한 방법의 추적 결과를 나타낸다.

Fig. 6. Tracking results in various environments, namely, for the cases when (a) the head moves fast, (b) head size significantly varies, and (c) background region is complex. In each case, the upper and lower rows represent the results from the Comanicu's and proposed methods, respectively.

표 1. 추적 알고리즘의 정확도 비교, 크기 오차는 각 장에서 수동으로 구한 타원의 단축의 길이와 예측된 타원의 단축의 길이의 차이의 평균값으로 구함.

Fig. 1. Accuracy comparison among tracking algorithms. Scale error is obtained by averaging the absolute difference values between the estimated and manually-selected scales in each frame.

	Birchfield's	Comanicu's	Proposed
위치 오차 (pixels)	10.72	5.32	3.12
크기 오차 (pixels)	4.10	2.17	1.46

반영하는 320x240 크기의 세 영상을 얻어 실험을 수행하였다. 또한, 동일한 실험 영상에 대해 mean-shift를 기반으로 한 대표적인 방법인 Comanicu의 방법과 제안한 방법을 비교하여 그림 6에 나타내었다. 그림 6으로부터 제안하는 알고리즘이 다른 알고리즘의 경우보다 정확한 결과를 보여줌을 볼 수 있다. 또한 정량적인 비교를 위하여 상대적으로 영상 내의 머리의 이동과 크기의 변화가 심한 그림 6(b)의 실험 영상을 이용하여 mean-shift 기반의 Comanicu의 방법과, 색과 모양을 동시에 고려한 Birchfield의 방법을 제안한 방법과 비교하여 표 1에 나타내었다. 표의 수치는 가장 적합한 타원의 위치와 크기  $\eta$ 를 수동으로 직접 구한 다음 이 결과와 Birchfield 및 Comanicu의 방법에서 얻어진 타원의 위치와 크기의 차이를 전체 360장의 프레임에 대해 평균값으로 나타낸 것이다. 아울러 Pentium IV 2.8 GHz CPU가 장착된 일반 PC에서 약 30 fps로 처리 가능함을 확인하였으며 실시간 추적 시스템에 적용하여 그 성능을 확인하였다.

### IV. 결론

제안한 방법은 실시간 연산을 위해 색 수렴성이 보장되는 저계산량의 mean-shift 방법을 적용한다. 수렴성을 보장하기 위하여 시간적으로 변하는 색을 고려해 주고 카메라의 움직임에 의한 전역 움직임을 보상하기 위한 고속 알고리즘을 제안하였다. 수렴된 결과를 이용하여 보다 정확한 타원을 예측하기 위해 영상의 경사도 값을 사용하는 모양 유사도 함수를 정의하였다.

색과 모양 정보만을 이용한 대부분의 추적 알고리즘은 서로 다른 머리의 겹침(occlusion)에 대해 취약한 성능을 보인다. 추후 연구에서는 색과 모양 정보 외에 표

적 내의 패턴의 유사도를 점검하여 겹침에 대해서도 강한 추적을 수행할 수 있는 알고리즘으로 발전시킬 수 있을 것이다.

### 참고 문헌

- [1] S. M. Smith and J. M. Brady, "ASSET-2: Real-time motion segmentation and shape tracking," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, no. 8, pp. 814-820, Aug. 1995.
- [2] A. Lipton, "Local application of optic flow to analyze rigid versus non-rigid motion," Technical report CMU-RI-TR-99-13, Robotics Institute, Carnegie Mellon University, Dec. 1999.
- [3] A. Baumberg and D. Hogg, "An efficient method for contour tracking using active shape models," Proc. Workshop on Motion of Nonrigid and Articulated Objects., pp. 194-199, Los Alamitos, Calif.: IEEE CS Press, Nov. 1994.
- [4] Comaniciu, "Robust detection and tracking of human faces with an active camera," IEEE Int'l Workshop on Visual Surveillance, pp. 11-18, July 2000.
- [5] D. Comaniciu, "Kernel-based object tracking," IEEE Trans. Pattern Analysis and Machine Intelligence. vol. 25, no. 5, pp. 564-577, May 2003.
- [6] D. Comaniciu and P. Meer, "Mean Shift: A robust approach toward feature space analysis," IEEE Trans. Pattern Analysis and Machine Intelligence. vol. 24, no. 5 May 2002.
- [7] S. Birchfield, "Elliptical head tracking using intensity gradient and color histograms," IEEE Conf. Computer Vision and Pattern Recognition, Santa Barbara, California, June 1998.
- [8] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," IEEE Trans. Communication Technology. vol. 15, no. 1, pp. 52-60, Feb. 1996.
- [9] I. Grinias and G. Tziritas, "Robust pan, tilt and zoom estimation," IEEE Int'l Conf. Digital Signal Processing, vol. 2, pp. 679-682, July 2002.

### 저 자 소 개



**정 동 길**(정회원)  
2003년 부산대학교 전자전기통신 공학부 학사 졸업.  
2005년 한국과학기술원 전자전산학과 전기및전자 공학전공 석사 졸업.  
2005년 현재 국방과학연구소 연구원.

<주관심분야 : 패턴인식, 객체추적, 특징점 추출>



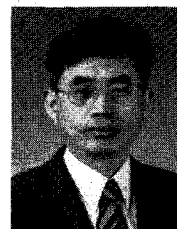
**양 유 경**(정회원)  
2002년 전북대학교 정보통신 공학과 학사 졸업.  
2004년 한국과학기술원 전자전산학과 전기 및 전자공학전공 석사 졸업.  
2005년 현재 KTF Technologies 전임연구원.

<주관심분야 : 신호처리, 통신>



**강 동 구**(학생회원)  
2000년 서강대학교 전자공학과 학사 졸업.  
2002년 한국과학기술원 전자전산학과 전기 및 전자공학전공 석사 졸업.  
2005년 현재 한국과학기술원 전자전산학과 전기 및 전자공학전공 박사과정.

<주관심분야 : 디지털 영상처리, 영상정합, 객체 추적>



**나 종 범**(평생회원)  
1975년 서울대학교 전자공학과 학사 졸업.  
1977년 한국과학기술원 전기및 전자공학과 석사 졸업.  
1983년 한국과학기술원 전기및 전자공학과 박사 졸업.

1983년~1987년 미국 Columbia 대학교 연구 조교수

1987년 7월~현재 한국과학기술원 전자전산학과 교수.

<주관심분야 : 디지털 영상처리, 동영상 신호처리, 3차원 시각화, 의료영상 시스템>