

## 능동형태모델 기반 다시점 영상 추적

임재현\*, 김대희\*, 최종호\*\*, 백준기\*

### Multiview Tracking using Active Shape Model

Jae-hyun Im\*, Dae-hee Kim\*, Jong-ho Choi\*\*, Joon-ki Paik\*

#### 요 약

다시점에서의 다중 객체 추적은 여러 분야에서 연구되고 있다. 다시점 영상 추적은 두 객체가 서로 근접하면 하나로 인식하는 문제점을 가지고 있다. 이러한 문제를 해결하기 위한 하나의 방법으로 능동형태모델(active shape mode: ASM)을 들 수 있다. ASM은 훈련집합을 이용하여 다른 객체에 가려진 목표 객체를 추적할 수 있다. 본 논문에서는 겹쳐진 객체를 추적하기 위해 ASM 기반의 다시점 추적 알고리즘(Multi-view tracking using ASM: MVTA)에 대해서 제안한다. 제안된 추적 방법은 (i) 영상 획득, (ii) 객체 추출, (iii) 객체 추적, 그리고 (iv) 현재 형태의 업데이트, 4가지 단계로 나눌 수 있다. 첫 번째 단계에서는 여러 대의 카메라를 사용해서 다시점 영상을 획득하며, 두 번째 단계에서는 객체를 배경으로부터 분리하며, 겹쳐진 객체로부터 목표 객체를 분리해낸다. 세 번째 단계에서는 추적을 위해 ASM을 사용하며, 마지막 단계인 네 번째 단계는 현재 입력 영상의 업데이트이다. 실험결과 제안한 MVTA는 겹쳐진 객체를 추적 시에 생기는 문제에 대해서 향상 된 결과를 보여준다.

▶ Keyword : 능동형태모델, 다시점 영상 추적

#### 1. 서 론

능동형태모델(active shape mode: ASM) 기반 영상 분리는 영역적 최적의 특징 점을 이용함으로써 표현된다. 반면 현존하는 방법은 정규화 된 윤곽선(profile)을 이용한다[1]. 통계적인 형태 모델은 라벨링(labeling) 된 영상의 기준점(land mark point)로 구성된 훈련 집합(training set)에 의해서 생성하며, 정렬 후 벡터  $\bar{x}$ 를 연산한다.

벡터  $\bar{x}$  연산 후 중요하지 않은 성분의 차원 수를 줄

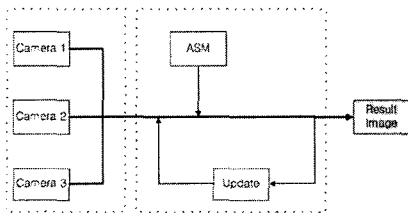
이기 위해 주성분 분석(principal component analysis: PCA)를 적용한다. 이러한 과정을 통해 얻어진 형태는 새로운 형태를 생성하는데 사용할 수 있다.

다시점 추적 시스템은 영상 추적에서 굉장히 중요하며, 선택이 가능할 정도의 겹침 현상이 적고, 조도의 차이가 추적에 영향을 미치지 않는 곳이어야 하는 조건이 따른다. 사람이 화면에 들어오면, 얼굴의 특징들, 질감, 옷 색상 등 비슷한 데이터들이 추출되며, 정합, 구성되어 배경으로부터 분리된다.

• 제1저자 : 임재현

\* 중앙대학교 첨단영상대학원 영상공학과, \*\*강남대학교 전자공학전공

가시제 다시점 추적 시스템에서는 추적의 성능을 저하시키는 겹침 문제를 피할 수 없다. 많은 연구자들이 추적에서의 겹침 문제를 개선하기 위해 칼만 필터(kalman filter)[2] 혹은 파티클 필터(particle filter)[3]를 이용했다. 본 논문에서는 가림현상(occlusion)을 해결하기 위해 다시점 추적 시스템에 ASM을 접목시켰으며, MVTA (multi-view tracking with ASM)라고 부른다. 제한한 시스템은 두 가지 처리 과정으로 나눌 수 있다. 하나는 다시점 영상을 생성하는 부분이며, 나머지 하나는 ASM을 이용하여 추적을 하며 새로운 형태를 업데이트하기 위한 업데이트 부분이다. 시스템의 구성은 <그림 1>에 나타나있다. 본 논문은 2장에서 다시점 영상을 생성하는 시스템에 대해 설명하며, 3장에서 MVTA 시스템에 대해 설명한다. 4장에서는 실험 결과를 소개하며, 마지막으로 5장에서 결론을 내리면서 논문을 마친다.



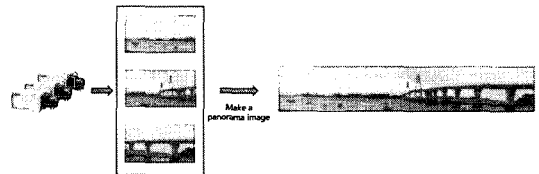
<그림 1> MVTA의 블록 다이어그램

## 2. 다시점 영상 시스템

최근 많은 연구자들은 다수의 카메라를 이용한 추적에 관심을 가지고 있다. 그 이유는 다수의 카메라를 이용함으로써 감시 영역을 확장할 수 있기 때문이다. 다시점 추적은 외형(contour) 정합과 템플릿(template) 정합으로 간략하게 나눌 수 있다[4]. 외형(contour) 기반의 추적은 템플릿 정합과 다르게 목표 객체를 확인하기에 유용한 방법[5]이다. 반면, 템플릿(template) 정합은 영상 안에서 목표 객체를 찾기 위한 가장 직접적인 방법이다[6]. 그러나 템플릿(template) 매칭만으로는 겹침 문제 때문에 다수의 객체를 추적하기에 충분치 않다. 그러므로 본 논문에서는 가림현상(occlusion)을 해결하기 위해 다수의 카메라와 외형(contour) 정합을 적용한다.

본 논문에서의 실험은 세 대의 카메라를 이용하며, 그것들은 다른 시점의 영상을 동시에 획득한다. 카메라로부터 얻은 입력 영상은 여러 좌표를 하나로 정합함으로써 하

나의 영상이 된다. <그림 2>는 다시점 영상의 시스템을 보여준다. 정합된 좌표는 하나의 파노라마 영상으로부터 특정 점을 선택하는데 사용된다.



<그림 2> 다시점 영상 시스템

## 3. MVTA 시스템

앞서 말했듯이, 카메라를 이용하여 객체를 추적할 시 가림현상(occlusion)이 발생하게 되면 템플릿(template) 기반의 추적 시스템에서는 객체를 하나로 간주해 버리게 된다. 이러한 문제점을 극복하고자 본 논문에서는 형태(contour)기반의 추적을 한다고 소개했었다. 다시 말해서, 본 논문에서는 다시점 영상 시스템에 ASM을 접목시켜 가림현상(occlusion)이 발생하더라도, 객체를 형태로 분리하여 추적한다.

### 3.1 객체 추출

객체를 추적하기 위해서는 우선 객체를 배경으로부터 추출해야하며, 겹쳐진 객체의 경우 목표 객체를 추출 해야 한다. 추출 알고리즘은 다양하지만 일반적으로 변화 검출(change detection) 기반 추출[7], 움직임(motion) 기반 추출[8], 형태학적(morphological) 추출[9][10], 그리고 공간-시간적 추출[11]등이 있다.

변화 검출(change detection)기반 추출[7]은 전역적인 문턱치 처리를 이용하며, 문턱치 값은 경험에 의해 정해진다. 그러나 이 방법은 객체로부터 영상의 가림현상(occlusion)의 여부를 구별하기에 적합하지 못하다. 이러한 문제점을 해결하기 위해 생겨난 것이 움직임(motion) 기반 추출이다. 이 방법은 배경의 가림 여부의 문제를 추출될 영역을 동일한 움직임과 위치 매개변수로 정의함으로써 해결한다. 이러한 정합 매개변수들은 화소의 밝기 값(pixel intensity values)의 공간-시간적 차이로부터 추정된다. 하지만 이러한 접근 방법은 전역적인 문턱치 처리 때문에 여전히 몇 가지 문제점을 가지고 있다. 앞에서 설

명한 방법들은 두 장의 연속적인 영상 프레임 사이의 차분에 의해서 공간적인 정보를 이용하는 반면, 형태학적(morphological) 추출은 수학적인 형태학(morphology) 처리를 이용한다. 이것은 추출에 굉장히 유용하다. 수학적인 형태학(morphology)은 크기, 연결성(connectivity), 대비(contrast), 그리고 외형(shape)과 같이 객체의 특징이라 생각할 수 있는 기하학적인 특징 점을 효과적으로 다룰 수 있다. 그러므로 형태학적(morphological) 추출 알고리즘은 객체의 외형을 보존 할 수 있다. 그러나 객체 자신이 들로 나뉘져 있다면, 효과적으로 표현하기 힘들다. 앞에서 언급한 시간적, 공간적 정보를 개발함으로써 추출 문제점을 해결하고자 하는 시도가 있었다[11]. 하지만 그것들은 객체의 정의가 주관적이고 영상 프레임의 밝기만을 이용해서 자동적으로 추출하는 방법이어서 추출의 문제점을 완벽하게 해결하진 못했다. 따라서 본 논문에서는 반자동적인 추출 방법을 사용한다.

### 3.2 ASM

연속적인 비디오 영상에서 비정형 객체를 추적하기 위해서는 기본적으로 객체의 형태와 위치를 추출하는 작업이 필요하다. 대부분의 2차원의 가변성 모델은 객체를 표현하기 위해 경계 정보를 이용한다. 가변성 모델을 추적할 시, 목표 객체의 사진 정보가 적용 가능하다는 조건을 만족하면, ASM은 정확성과 효율성면에서 최적의 방법 중 하나이다.

ASM은 목표 객체의 외곽 형태의 선형적 정보(a priori knowledge)를 이용한다. 따라서 불완전 잡음이 많이 들어간 영상에서도 강건하게 객체의 외곽을 정합하여 추출할 수 있다. ASM은 훈련 집합(training set)이 필요한데, 사용자가 직접 자동 혹은 수동으로 훈련 영상으로부터 객체의 외곽에 기준점들을 표시하여 여러 가지 형태 모델을 구축해야 한다. 기준점은 객체의 외곽 특징을 잘 표현할 수 있어야 한다. ASM의 알고리즘은 크게 (i)기준점(landmark point), (ii)주요 성분 분석(principal component analysis: PCA), (iii)지역구조(local structure) 모델링, 그리고 (iv) 모델 정합과 같이 크게 네 가지 단계로 나눌 수 있다.

#### 3.2.1 기준점(landmark point) 할당

주어진 영상으로부터 적당한 기준점을 객체의 윤곽선에 수동으로 할당한다. 최적의 기준점은 현재 영상에서 정해진 특징 점이 다음 영상에서도 특징 점으로 연결이 되는 것이다. 2차원 영상에서 우리는 n개의 기준점을 2n차원의 벡터를 사용해서 다음과 같이 표현할 수 있다.

$$x = [x_1, \dots, x_n, y_1, \dots, y_n]^T \dots\dots\dots (1)$$

본 논문에서 적용될 실험에 사용할 기준점은 n=40로 하였다. 기준점의 역할은 모델의 윤곽 형태를 조절하는 것이다. 즉 초기 값으로 설정된 기준점은 계속 유지되는 것이 아니라, 각각의 기준점에서 경계의 수직방향에 따라 추적하는 객체의 실제 윤곽과 편차를 최소화 하는 방향으로 갱신된다.

#### 3.2.2 주성분 분석(PCA)

n개의 기준점의 집합은 객체의 형태를 표현한다. 학습 집합(training set)에서 각각의 형태는 2차원 공간에서 존재하지만, 우리는 PCA 기술을 기반으로 하여 좀 더 적은 수의 매개변수를 가지고 객체의 형태를 효과적으로 모델링할 수 있다. PCA 알고리즘은 다음과 같다.

1. 훈련 집합에서 m개의 표본 형태의 평균을 구한다.

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i \dots\dots\dots (2)$$

여기서 m은 각각의 형태에 따른 훈련 집합을 의미한다. (i = 1, ..., m)

2. 훈련 집합으로부터 공분산 행렬을 구한다.

$$S = \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})^T \dots\dots\dots (3)$$

3. 다음과 같은 행렬을 만든다.

$$\Phi = [\phi_1 | \phi_2 | \dots | \phi_q] \dots\dots\dots (4)$$

여기서  $\phi_j$ (단,  $j = 1, 2, \dots, q$ )는 S에서 q개의 가장 큰 고유 값(eigen value)에 해당하는 고유 벡터(eigen vector)를 의미한다. 이 때 q는 데이터의 전체적인 분산 값에의 98%에 해당하는 모델형태의 개수를 사용한다.

4. 주어진  $\Phi$ 와  $\bar{x}$ 로부터 각각의 형태는 다음과 같이 간략화 시킬 수 있다.

$$x_i \approx \bar{x} + \Phi b_i \dots\dots\dots (5)$$

여기서

$$b_i = \Phi^T(x_i - \bar{x}) \dots\dots\dots (6)$$

벡터  $b$ 는 변형 가능한 모델 파라미터의 집합을 정의한다.  $b$ 의 요소를 변하시키면서 형태를 변화시킬 수 있다.

### 3.2.3 모델 정합(model fitting)

모델 좌표프레임  $x$ 와 영상 좌표프레임  $y$ 를 정합하기 위해서 다음과 같은 여러 함수를 최소화함으로써 이루어질 수 있다.

$$E = (y - Mx)^T W(y - Mx) \dots\dots\dots (7)$$

여기서  $W$ 는 각각 기준점에 대한 가중치를 나타낸다. 그리고  $M$ 은 회전성분  $\theta$ , 전이성분  $t$ , 스케일링 성분  $s$ 로 이루어진 기하학적인 변환을 나타내는 행렬이다. 하나의 점을  $[x_0, y_0]^T$ 으로 표현한다면, 기하학적인 변환행렬은 다음과 같이 정의 될 수 있다.

$$M \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = s \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \dots\dots\dots (8)$$

기하학적인 파라미터가 결정되면, 모델좌표 프레임으로 투영된 값은 다음과 같이 주어진다.

$$x_p = M^{-1}y \dots\dots\dots (9)$$

결국 모델의 파라미터는 다음과 같이 갱신된다.

$$b = \Phi^T(x_p - \bar{x}) \dots\dots\dots (10)$$

최적으로 갱신된 기준점의 조합은 영상좌표 프레임  $y$ 에 새로운 형태를 만들어낸다. 새로운 형태는 식 (7)을 사용하여 가장 근접한 형태를 찾는 데 이용될 수 있다.  $M$ 을 통해서 최적의 위치를 도출한 뒤, 이 새로운 형태는 주어진 학습 집합에 주요 요소에 포함된  $\Phi$ 로 투영될 수 있다. 이 과정은 모델 파라미터  $b$ 를 갱신한다. 그 결과로 주요 요소에 해당되는 비슷한 편차만이 모델 파라미터에 영향을 줄 수 있다. 모델 파라미터를 계산한 후에  $x$ 에 의해 주어지는 새로운 형태는 식 (5)에 의해 구할 수 있고, 적당한 반복을 거쳐, 최종적인 형태가  $x$ 로 얻어진다.

### 3.2.4 지역구조 모델링

ASM을 사용하여 입력영상의 형태를 해석하기 위해서는 영상과 모델을 정합시키는 기준점의 집합을 찾아야 한다.

우선  $j$ 번째 기준점을 가로지르는  $K$ 만큼의 길이를 갖는 지역적인 윤곽선(profile)의 정규화 된 미분 값을  $g_j(j = 0, \dots, n)$ . 해당하는 평균과 분산을 각각  $\bar{g}_j$ 와  $S_j$ 로 둔다. 가장 근접한 윤곽을 구하기 위해서 모델의 표본과 평균사이의 Mahalanobis 거리를 최소화함으로써 얻을 수 있다. Mahalanobis 거리는 다음과 같이 나타낼 수 있다.

$$f(g_{j,m}) = (g_{j,m} - \bar{g}_j)^T S_j^{-1} (g_{j,m} - \bar{g}_j) \dots\dots\dots (11)$$

여기서  $g_{j,m}$ 은 해당되는 경계의 수직방향에 존재하는 한  $m$ 개의 표본만큼 이동된  $g_j$ 를 나타낸다.

## 4. 실험 결과

본 절에서는 제안된 알고리즘의 성능을 실험을 통하여 확인한다. <그림 3>은 카메라 3대를 이용하여 얻은 다시점 영상이다. 이 영상에서 목표 객체의 추적을 위해 ASM을 이용하였다. <그림 4>는 목표 객체를 ASM을 이용하여 외형을 추적하는 것을 보여준다. 실험 결과에서 볼 수 있듯이 ASM을 이용하여 움직이는 객체를 강건하게 추적을 하였다.



<그림 3> 다시점 영상



<그림 4> 객체의 외형 추적

## 5. 결론

본 논문에서는 다중 카메라 환경에서 전 방향 추적 시스템을 보였다. 다수의 카메라를 사용하여 보다 넓은 영역에서 객체를 감시 추적할 수 있고, 능동 형태 모델을 사용하여 부분 가림현상에 강건하게 대처할 수 있다. 앞으로 완전히 가려진 객체에 관해 강건한 추적을 위해 메타데이

터를 이용하는 추적, 그리고 좀 더 정확한 추적을 위해 세밀한 객체 분할과 인식을 할 수 있는 감시 시스템을 발전해 나아갈 것이다.

### 참고문헌

- [1] T. Cootes and C. Taylor, "Active shape models," Proc. 3rd British Machine Vision Conference, vol. 3, pp. 266-275, September 1992.
- [2] Y. Yao and R. Chellappa, "Tracking a dynamic set of feature points," IEEE Trans. Image Processing, vol. 4, pp. 1382-1395, 1995.
- [3] K. Otsuka and N. Mukawa, "A particle filter for tracking densely objects based on explicit multiview occlusion analysis," Proc. IEEE, Conf, Pattern Recognition, vol. 4, pp. 745-750, 2004.
- [4] Y. Cheng and C. Huang, "Multiple people visual tracking in a multi-camera system for cluttered environments," Proc. IEEE, Conf, Intelligent Robots and Systems, pp. 675-680, 2006.
- [5] S. Birchfield, "An elliptical head tracker," Conf, 31th Asilomar, Signal, System & Computers, vol. 2, pp. 1710-1714, 1997.
- [6] M. Lew, N. Sube, and T. Huang, "Improving visual matching," Proc IEEE, Conf, Computer Vision and Pattern Recognition, vol. 2, pp. 58-65, 2000.
- [7] T. Aach and A. Kaup, "Statistical model-based change detection in moving video," Signal Processing, vol. 31, pp. 165-180, 1993.
- [8] J. Choi, S. Lee, and S. Kim, "Spatio-temporal video segmentation using a joint similarity measure," IEEE trans. Circuit and System for Video Technology, vol. 7, pp. 297-286, April 1997.
- [9] J. Choi, "A user-assisted segmentation method for video object plane generation," ITC-CSCC 1998, pp. 7-10, July 1998.
- [10] J. Choi, M. Kim, M. Lee, and C. Ahn, "Automatic segmentation based on spatio-temporal information," ISO/IEC JTC1/SC29/WG11, MPEG4/m209, April 1997.
- [11] M. Hotter and R. Thoma, "Image segmentation based on object oriented mapping parameter estimation," Signal Processing, vol. 15, pp. 315-334, 1998.