

상반신 포즈 추적을 위한 키포즈 기반 예측분포

오 치 민[†] · 이 칠 우^{††}

요 약

Pictorial Structures(PS)는 동적 프로그래밍을 이용하여 인체의 포즈 추적 및 인식 하는 것에 매우 효과적인 방법으로 알려져 있다. 본 논문에서 상반신 포즈는 PS와 Particle filter(PF)에 의한 동적 프로그래밍 기법으로 추적된다. PF와 같은 동적프로그래밍에서 마코프 연쇄 (Markov Chain) 기반 동적 움직임 모델은 높은 자유도를 갖는 상반신 포즈를 예측하기 어려운 단점이 있다. 본 논문에서 제안하는 방법은 키포즈 기반 예측분포이며, 이것은 상반신 실루엣과 키포즈(Key Pose)들 사이의 유사도를 참고하여 파티클(Particle)을 적절히 예측하는 것이다. 실험 결과를 통해 제안된 방법은 기존 방법 성능을 70.51% 향상시킨 것을 확인하였다.

키워드 : 상반신 포즈, 그림 구조체, 파티클필터, 예측분포, 키포즈

Key Pose-based Proposal Distribution for Upper Body Pose Tracking

Chi-Min Oh[†] · Chil-Woo Lee^{††}

ABSTRACT

Pictorial Structures is known as an effective method that recognizes and tracks human poses. In this paper, the upper body pose is also tracked by PS and a particle filter(PF). PF is one of dynamic programming methods. But Markov chain-based dynamic motion model which is used in dynamic programming methods such as PF, couldn't predict effectively the highly articulated upper body motions. Therefore PF often fails to track upper body pose. In this paper we propose the key pose-based proposal distribution for proper particle prediction based on the similarities between key poses and an upper body silhouette. In the experimental results we confirmed our 70.51% improved performance comparing with a conventional method.

Keywords : Upper Body Pose, Pictorial Structures, Particle Filter, Proposal Distribution, Key Poses.

1. 서 론

상반신 포즈 인식은 오랫동안 활발한 연구 중 하나이다. 이것은 보안과 감시분야에 쓰일 수 있으며 특히 최근 각광을 받고 있는 인간-컴퓨터 상호작용과 인간-로봇 상호작용 분야에서 사용될 수 있다. 상호작용에 응용하기 위해서는 상반신 포즈에 대한 적절한 정의를 통해 문제를 해결할 수 있어야 하며, 높은 자유도와 동적인 움직임을 추적할 수 있는 시스템을 필요로 한다.

기존 연구들[1-5]은 Pictorial Structures(PS)[1]처럼 사람의 형상을 간략히 나타낸 포즈 모델을 포즈추적에 사용하였다. PS는 인체를 관절별로 구분하고 각 관절은 파트(Part)로 정의하고 관절 사이의 연결성은 트리 그래프 구조로 표현한다. PS는 단순한 형태이지만 인간의 포즈공간을 충분히 표현하므로 인식된 포즈를 이해하는 데에 충분한 정보를 제공한다. 또한 PS모델은 템플릿 매칭처럼 영상과 직접적으로 비교가 가능하므로 포즈 추정 문제를 쉽게 한다.

PS를 이용하더라도 각 관절의 움직임은 매우 동적이고 비선형이기 때문에 포즈를 추적하는 것은 여전히 어려운 일이다. 기존 연구[1-5]에서는 동적이고 비선형적인 변화를 추적하기 위해서 동적프로그래밍 기반 추적기법[1]을 적용하였다. 대표적인 방법으로는 PF[6], MCMC[7]과 Unscented Kalman filter[13]가 있으며, 현재 영상에 나타난 포즈에서 다음 영상의 포즈를 예측하기 위해 동적 움직임 모델을 이

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 지능로봇을 위한 3D 센싱 및 비전 기반 사람/물체 인식 기술 개발 지원사업의 연구결과로 수행되었음 (NIPA-2010-C7000-1001-0007).

† 준 회 원 : 전남대학교 전자컴퓨터공학부 박사과정

†† 정 회 원 : 전남대학교 전자컴퓨터공학부 교수

논문접수 : 2010년 8월 2일

수정일 : 1차 2010년 9월 29일

심사완료 : 2010년 10월 14일

용한다. 대표적 동적 움직임 모델은 마코프 연쇄(Markov Chain)에 따라 상태를 예측하는 예측 모델이다.

기존 마코프 연쇄 기반 동적 움직임 모델은 후보 포즈 샘플, 즉 파티클을 효과적으로 예측하지 못하는 단점이 있으며 이유는 상반신 포즈의 자유도가 높으며 관절 움직임이 매우 동적이기 때문이다. 예측분포¹⁾(Proposal distribution)[8]는 마코프 과정 기반 움직임 모델 보다 더 효과적으로 파티클들을 예측한다. 마코프 연쇄 기반 동적 움직임 모델은 현재 상태를 예측하기 위해 이전 상태 값을 참조한다. 예측분포는 이전 상태와 현재 관측 값을 참고로 한다. 즉 예측된 현재 상태 값은 현재 관측 값을 참고하여 예측될 수 있다.

본 논문의 예측분포는 기존 마코프 연쇄 기반 동적 움직임 모델에서 일부 파티클을 예측하면서도, 오류가 큰 파티클들은 입력 실루엣과 유사한 키포즈를 변형한 파티클을 예측한다. 키포즈 기반 예측분포는 최초 파티클 초기화를 자동화 할 수 있으며 동적 움직임 모델 예측이 실패한 파티클들을 변형된 키포즈 PS모델로 재초기화 할 수 있는 장점을 가지고 있다.

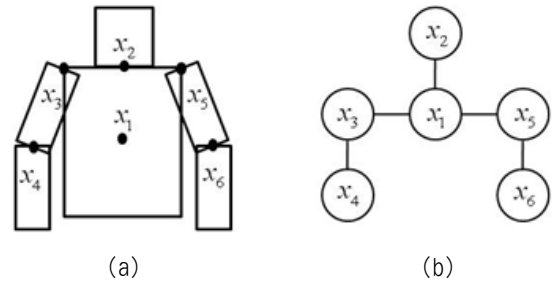
본 논문에서는 키포즈 기반 예측분포를 기반으로 하는 PF를 이용하여 PS로 모델링된 상반신 포즈를 연속적인 영상에서 추적하는 방법을 서술한다. 본 논문의 내용은 다음과 같이 구성된다. 2장에서는 PS에 의해 상반신 포즈를 설계하는 방법을 서술한다. 3장에서는 상반신 포즈 추적을 위한 PF 동적 프로그래밍 기법과 예측분포를 설명한다. 4장에서는 키포즈 기반 예측분포를 설명한다. 5장에서는 우도 계산방법이 설명되며, 6장에서는 제안한 방법의 효율성 평가 결과를 나타낸다. 7장에서는 현 연구에 대한 결론과 향후 연구방향에 대해 언급하며 본 논문을 맺는다.

2. PS기반 상반신 포즈

PS[1,10]는 변형 가능한(Deformable) 객체를 다수의 파트(Part)들로 분리하고 파트 사이의 연결성을 모델링하는 방법이다. (그림 1)은 본 논문의 상반신 PS모델과 트리 구조의 관절 연결 관계를 나타낸다. 상반신 포즈는 (그림 1(a))의 PS 모델을 이용하여 머리, 몸통, 왼손 위, 왼손 아래, 오른손 위, 오른손 아래 총 6개 파트로 구성된 객체로 정의된다. 상반신 PS 모델은 (그림 1(b))처럼 그래프 구조로 모든 파트가 연결되어 있으며, 각 파트의 위치 및 회전량은 부모의 위치에 의해 결정되도록 서로 유기적으로 연결되어 있다. 이러한 상반신 포즈를 구성하는 파라미터는 식 (1)의 구성 벡터 Θ 와 같다.

$$\Theta = [s, x_1, x_2, x_3, x_4, x_5, x_6]^T \quad (1)$$

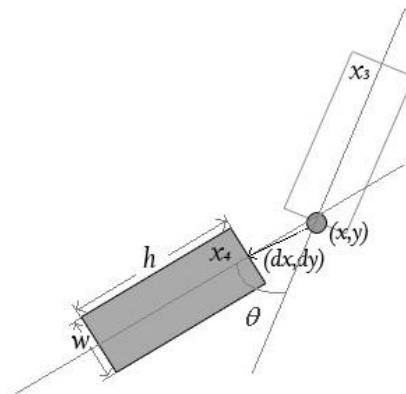
1) 예측분포는 제안분포라는 용어로도 알려져 있다.



(그림 1) 본 논문에서 사용되는 상반신 PS 모델
(a) 상반신 모델의 외형 (b) 파트 연결 그래프 구조

구성 벡터의 s 는 상반신 포즈의 크기변화에 대한 변수이다. x_i 는 식 (2)의 관절 파트의 파라미터 집합이다. x_i 파라미터들은 부모의 위치에 의해 결정되는 파트의 위치 (x, y) , 그 위치로부터 부모 파트로부터 스프링처럼 움직일 수 있는 위치 이동량 (dx, dy) , 회전량 θ 와 파트 크기 (w, h) 이다. (그림 2)는 자식 x_4 파트가 부모 x_3 파트에 의존하여 위치와 회전이 결정되는 예를 나타낸다.

$$x_i = [(x, y), (dx, dy), \theta, (w, h)]^T \quad (2)$$



(그림 2) 부모 파트 x_3 에 상대적으로 이동 (dx, dy) 및 회전 (θ) 하는 자식 파트 x_4

3. 상반신 포즈 추적

3.1 에러 최소화

입력 영상의 상반신 포즈를 결정하는 것은 상반신 입력 영상에 가장 유사한 상반신 PS 모델의 구성벡터를 찾는 것이다. 가장 적절한 구성벡터 결정은 에러 최소화 문제로 표현할 수 있다. 즉 주어진 입력영상에서의 포즈 결정은 목적함수 E 를 최소화 하는 구성벡터를 찾아내는 것이다. 포즈 측정은 식 (3)의 목적함수 E 를 최소화 하는 시간 t 의 PS 구성벡터 Θ_t 를 찾는 것과 같다.

$$E(\Theta_t) = \sum_{x_i \in \Theta_t} \gamma(I_S, I_E, x_i) + \sum_{(x_i, x_j) \in G} \delta(x_i, x_j) + \sum_{i=t-n}^t \Delta(\Theta_i) \quad (3)$$

식 (3)의 Θ_t 의 에러가 최소화되기 위한 조건은 3가지이다. 첫 번째 조건은 주어진 PS가 실루엣 영상과 유사해야 하는 것이다. 두 번째 조건은 주어진 PS 내부 파트들 사이의 연결 관계가 (그림 1(b))의 그래프 구조로부터 만족해야 하는 것이다. 마지막 조건은 주어진 PS의 포즈가 최근 포즈들과 차이가 너무 크지 않아야 하는 것이다.

식 (3)의 에러 함수는 3가지 조건에 대한 에러를 계산한다. 첫 번째 에러함수 γ 는 Θ_t 의 PS 모델이 두 영상의 상반신 형상과 차이점을 에러로 계산하는 함수이다. γ 는 포즈를 나타내는 두 가지 영상인 실루엣 영상 I_S 와 에지영상 I_E 를 이용하여 PS모델 Θ_t 의 오차를 측정한다. 두 번째 에러함수 δ 는 (그림 1(b))에 나타난 그래프 형태 G 를 참고하여 주어진 Θ_t 의 PS모델을 구성하는 파트들의 위치가 적절 한지에 대한 에러 값을 계산하는 함수이다. 세 번째 에러함수 Δ 는 시간에 따른 상반신 포즈의 변화가 마코프 과정을 따른다고 가정하고, 주어진 Θ_t 를 포함한 n 개의 연속적인 최근 포즈 구성벡터들의 유사도가 낮을수록 높은 에러로 계산하는 함수이다.

3.2 최대사후확률에 의한 포즈인식 문제

에러 최소화 문제는 동적 프로그래밍 적용이 쉬운 최대사 후확률 (Maximum a Posteriori, MAP) 문제 형태로 표현 될 수 있다. MAP문제 해결을 위해서는 Bayes' Rule에 따라 사후확률을 나타내는 우도와 사전확률 정의가 필요하다. 기존 에러함수들은 우도와 사전확률과 의미가 유사하다. 에 러함수의 값을 확률로 변환하는 것은 Gibbs 분포[15]를 통 해 해결할 수 있다. 입력영상에서 주어진 Θ_t 의 에러를 계 산하는 함수 γ 는 우도 $P(I|\Theta_t) = P(I_S, I_E|\Theta_t)$ 와 유사하 며 δ 함수와 Δ 함수는 $p(\Theta_t|\Theta_{t-1}, \dots, \Theta_{t-n})$ n 차 마코프 과정 기반 사전확률로 표현된다. 만일 1차 마코프 과정을 따르는 경우 사전확률은 $p(\Theta_t|\Theta_{t-1})$ 이며, 최적 포즈를 결 정하는 에러 최소화 문제는 식 (4)와 같이 표현된다.

$$\Theta_t^* = \arg \max_{\Theta_t} p(I|\Theta_t)p(\Theta_t|\Theta_{t-1}) \quad (4)$$

식 (4)에서 Θ_t^* 는 사후확률분포를 구성하는 샘플 중에서 가장 높은 사후확률을 갖는 포즈 구성 벡터이다.

3.3 Particle Filter (PF) 기반 사후분포추정

식(4) MAP 문제 해결은 사후확률분포 추정을 필요로 한 다. 사후분포는 매시간 영상이 발생할 때마다 일부 분포의

확률이 변경되는 특징이 있다. 사후분포 추정을 위해서는 시간에 연속적으로 변화하는 사후확률분포를 효과적으로 측 정할 수 있는 방법이 필요하다.

PF는 실시간으로 변경되는 사후확률분포를 이전 시간의 사후분포로부터 측정한다. PF는 사후분포 예측(prediction) 과 갱신(update) 두 가지 단계로 구성된다. 첫 번째 단계인 PF 사후분포 예측은 이전 시간의 사후확률분포를 구성하는 파티클들을 현재 사후확률분포에서 높은 확률을 갖도록 이 전 사전확률분포를 구성하는 파티클에 잡음을 가하는 것이 다. 두 번째 단계인 PF 사후분포 갱신은 잡음을 가한 파티 클들이 관측변수에서의 우도를 계산함으로써 현재 관측변수 에 대한 사후확률분포로 변경하는 것이다.

PF는 시간에 연속적으로 식(5)의 현재 사후분포를 구성 할 파티클들이 높은 확률을 갖도록 이전 사후분포를 구성하 는 파티클들을 사전정보를 기반으로 현재 사후분포에 맞도 록 예측하고 현재 관측변수에 대한 가중치 계산을 통해 갱 신하는 과정을 재귀적으로 실행한다.

$$p(\Theta_t|I_t) = \alpha p(I_t|\Theta_t) \int_{\Theta_{t-1}} p(\Theta_t|\Theta_{t-1})p(\Theta_{t-1}|I_{t-1}) \quad (5)$$

식 (5)는 PF에 의해 이전 사후분포로부터 예측과 갱신이 된 현재 사후분포이다. 식 (6)은 식 (5)에 서 기존 사후분포 를 현재 상태에 대한 분포로 예측한 확률분포 식이다. 사후 분포 예측은 기존 사후분포 $p(\Theta_{t-1}|I_{t-1})$ 을 현재 상태에 대한 분포로 갱신하기 위해 동적 움직임 모델 $p(\Theta_t|\Theta_{t-1})$ 의 확률 내에서 상태 Θ_{t-1} 에서 Θ_t 로 변경된 $p(\Theta_t|I_{t-1})$ 를 계산하는 것이다.

$$p(\Theta_t|I_{t-1}) = \int_{\Theta_{t-1}} p(\Theta_t|\Theta_{t-1})p(\Theta_{t-1}|I_{t-1}) \quad (6)$$

예측 다음 단계인 갱신에서는 $p(I_t|\Theta_t)$ 우도를 통해 예측 된 사후분포 $p(\Theta_t|I_{t-1})$ 를 현재 관측변수 I_t 에 대한 분포 로 변경시키므로 $p(\Theta_t|I_t)$ 사후분포를 측정한다. 식 (7)은 우도를 통해 갱신된 사후분포 확률의 합을 1로 정규화하기 위한 상수 α 를 포함한 $p(\Theta_t|I_t)$ 를 나타낸다.

$$p(\Theta_t|I_t) = \alpha p(I_t|\Theta_t)p(\Theta_t|I_{t-1}) = \frac{p(I_t|\Theta_t)p(\Theta_t|I_{t-1})}{p(\Theta_t|I_{t-1})} \quad (7)$$

PF는 연속적인 사후확률 분포를 효율적으로 갱신할 수 있다. 그렇지만 최초 파티클 집합은 초기화가 필요하며 또 한 예측이 실패했을 경우 오류로부터 복원되지 않는다. 이 러한 문제들은 PF 예측단계에서 $p(\Theta_t|\Theta_{t-1})$ 대신 식(8)

의 예측분포 $q(\Theta_t|\Theta_{t-1}, I_t)$ 를 파티클들을 예측하는데 이용하면 해결할 수 있다.

$$\Theta_t \sim q(\Theta_t|\Theta_{t-1}, I_t) \quad (8)$$

예측분포를 파티클 예측에 이용하는 것은 두 가지의 장점이 있다[9]. 첫 번째 장점은 오류 복원이 가능하다는 것이다. 사후확률이 낮은 파티클들은 관측 영상 I_t 를 분석하여 영상에 나타난 포즈와 유사한 구성벡터 값을 갖도록 할 수 있기 때문이다. 그리고 두 번째 장점은 초기 파티클 집합의 값을 초기화할 수 있다는 것이다. 초기 파티클집합 $\Theta_0^{(i)} (1 \leq i \leq r)$ 은 정규분포이므로 모두 I_t 를 참고하여 $\Theta_1^{(i)}$ 으로 예측되기 때문이다.

4. 키포즈 기반 예측분포

4.1 예측분포 구성 방법

예측분포[16]는 이전 사후분포의 파티클들을 기반으로 현재 사후분포의 파티클들을 생성하는 확률밀도함수이다. 기존 예측분포는 현재 상태를 나타내는 확률 분포를 이전 상태 조건으로 하는 조건부 확률밀도함수이다. 기존 PF에서의 예측분포인 동적 움직임 모델 $p(\Theta_t|\Theta_{t-1})$ 은 이전 파티클 집합에 오류가 있을 경우 새롭게 생성되는 파티클은 여전히 오류가 있던 이전 파티클에 영향을 받는 문제점이 있다.

근래의 예측분포는 관측변수를 추가로 참고하는 $q(\Theta_t|\Theta_{t-1}, I_t)$ 모델이 사용되는 추세이다. 새로운 예측분포는 기존 파티클 집합을 참고하여 새로운 파티클들을 예측하는 $p(\Theta_t|\Theta_{t-1})$ 과 함께 관측영상을 참고하는 식(9)와 같은 $q(\Theta_t|\Theta_{t-1}, I_t)$ 으로 정의 될 수 있다.

$$q(\Theta_t|\Theta_{t-1}, I_t) = \alpha p(\Theta_t|\Theta_{t-1}) + (1-\alpha)p(\Theta_t|I_t) \quad (9)$$

$p(\Theta_t|\Theta_{t-1})$ 은 기존 상태를 평균으로 현재 상태를 나타내는 정규분포이다. $p(\Theta_t|I_t)$ 는 관측영상과 PS구성벡터들의 포즈 유사도 분포이다. 상반신 포즈 공간간의 크기는 사실상 무한하므로 모든 PS구성벡터에 대한 확률 분포 계산은 불가능하다. 현실적인 $p(\Theta_t|I_t)$ 분포 계산은 모든 공간의 포즈 대신에 일부 포즈만을 모은 다수의 (포즈영상-PS 모델) 예제 포즈들을 이용하는 것이다. 포즈영상은 관측영상과의 유사도를 계산하는데 이용된다. 예제 포즈를 이용함으로써 $p(\Theta_t|I_t)$ 분포는 효과적으로 예제포즈와 관측영상과의 정규화된 유사도 분포로 추정될 수 있다.

예제포즈들을 관측영상과 비교한 유사도순서에 따라 예제 포즈의 PS모델 구성벡터들을 정렬할 경우, 가중치가 낮은

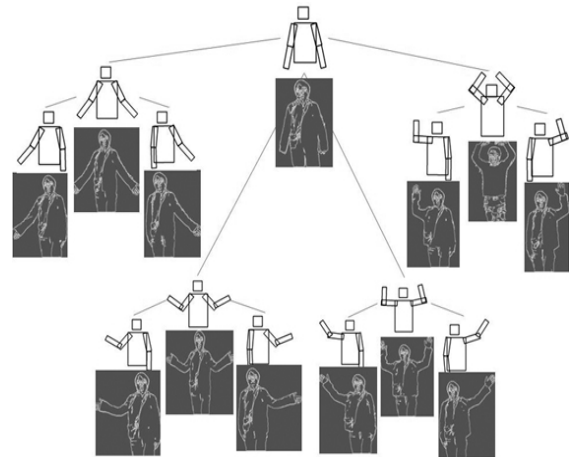
각 파티클은 유사도가 가장 큰 PS 모델 구성벡터 외에 정렬된 순서에 따른 PS모델 구성벡터로 설정될 수 있다. 정렬된 순서에 따라 파티클을 예측하는 방법은 예측분포의 다양성을 보장할 수 있다. 또한 각 파티클은 예제포즈의 PS모델 구성벡터에 정규분포를 따르는 잡음을 가한 값으로 설정됨으로써 예제 포즈들 외의 포즈들도 예측될 수 있다.

4.2 키포즈 라이브러리(Key Pose Library) 구축

본 논문에서는 예제 포즈 데이터베이스를 키포즈 라이브러리라고 정의한다. 키포즈가 의미하는 것은 상반신 정면 포즈 중에서 주요한 포즈를 가리킨다. 키포즈는 포즈 공간에서 차지하는 비율을 통해 선택될 수 있으며, 사용자가 경험적으로 특정 포즈들 중에서 선택할 수도 있다.

관측영상에서 예측분포 계산을 위해서 키포즈 라이브러리를 구축해야 한다. 각 키포즈는 식(10)처럼 k_i 으로 정의되며 예지영상 I_E^i 와 특징벡터 f_i , PS 구성벡터 m_i 를 갖는다. K 는 키포즈를 n 개를 가진 키포즈 라이브러리이다.

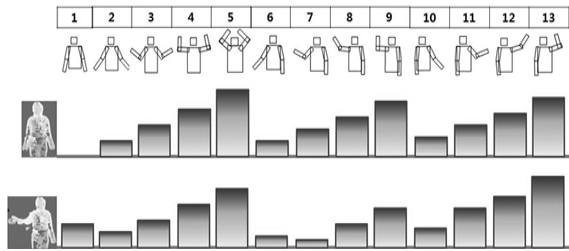
$$k_i = (I_E^i, f_i, m_i) \quad K = (k_1, \dots, k_n) \quad (10)$$



(그림 3) 키포즈 라이브러리(키포즈 k_i 는 PS모델 구성벡터 m_i 와 예지영상 I_E^i , 특징벡터 f_i 로 구성됨)

(그림 3)과 같이 모든 키포즈들은 예지영상과 PS 모델을 가진다. PS 구성벡터는 예지영상에서 자체 GUI툴을 이용하여 예지의 포즈와 유사한 형태의 PS포즈의 구성벡터로 획득된다. 특징벡터 f_i 은 해당 키포즈의 예지 영상과 모든 키포즈의 PS 구성벡터의 거리 값들이다. 본 논문에서 사용하는 키포즈의 수는 13개이므로 특징벡터 f_i 의 차원의 수는 키포즈의 수와 같다. 키포즈 특징벡터 f_i 는 식(11)과 같다.

$$f_i = [v(I_E^i, m_1), v(I_E^i, m_2), \dots, v(I_E^i, m_n)]^T \quad (11)$$



(그림 4) 두 키포즈 특징벡터 f_i 의 예, 모든 키포즈와의 거리 히스토그램

식(11)에서 f_i 는 키포즈 k_i 의 에지영상 I_E^i 와 모든 키포즈와의 거리 히스토그램이다. 거리 값들은 에지영상 I_E^i 이 나타내는 상반신 포즈와 모든 키포즈의 PS모델 형상의 차이 값이며, i 번째 키포즈가 전체 키포즈 공간에서의 위치를 가리킬 수 있다. 함수 v 는 챔퍼매칭을 이용하여 에지영상과 PS 모델 사이의 거리 값인 챔퍼거리(Chamfer Distance)[14]를 계산하는 함수이다. (그림 4)는 계산된 특징벡터의 예를 보여주고 있으며, 키포즈 공간에서 해당 키포즈의 위치를 나타낸다. 이러한 축소된 포즈 공간과 각 포즈와의 관계로 특징을 추출하는 것을 임베딩 기법[17]이라 한다. 임베딩 특징을 이용한 두 포즈 특징의 비교는 단순한 두 특징의 비교가 아니라 전체 키포즈 공간을 고려한 비교가 가능한 장점이 있다.

키포즈 라이브러리를 이용한 본 논문에서 사용하는 예측분포는 식(12)과 같이 표현된다.

$$q(\theta_t | \theta_t, I_t, K) = \alpha p(\theta_t | \theta_{t-1}) + (1 - \alpha) \sum_{j=1}^n p(\theta_t | m_j) p(I_t | m_j) \quad (12)$$

식 (12)의 예측분포는 동적 움직임 모델 $p(\theta_t | \theta_{t-1})$ 과 키포즈 라이브러리와 파티클 θ_t 사이의 유사도 분포

$$\frac{1}{n} \sum_{j=1}^n p(\theta_t | m_j) p(I_t | m_j)$$

를 혼합한 확률분포이다. 파티클들은 식 (12)의 예측분포에서 확률이 높은 것으로 평가되는 임의의 값으로 예측된다. 기존 사후분포에서 가중치 값이 높았던 파티클들은 $p(\theta_t | \theta_{t-1})$ 확률 안에 존재하는 새로운 상태 값들로 설정된다. 즉 파티클의 새로운 상태 θ_t 는 가중치가 높았던 이전 상태 θ_{t-1} 를 평균으로 하는 정규 분포에 존재하는 임의의 값으로 설정된다. 기존 사후분포에서 가중치가 낮았던 나머지 파티클들은 $p(I_t | m_j)$ 분포에 존재하는 임의의 값으로 예측된다. 즉 새로운 상태 $\theta_t^{(i)}$ 는 이전 상태 $\theta_{t-1}^{(i)}$ 를 참고하지 않고 관측영상 I_t 와 유사한 키포즈의 PS 구성벡터 m_j 에 잡음을 더한 값으로 예측된다.

본 논문에서 사용하는 예측분포는 관측영상을 적절하게 분석한 정보와 기존 동적 움직임 모델을 그대로 이용함으로

써 파티클들의 가중치에 따라 적절하게 예측하는 모델로 사용될 수 있다. 파티클 필터링 과정 중에서 가중치가 낮은 파티클이 다량 발생할 수 있는 오류를 보정할 수 있으며 최초 파티클들의 초기화도 파티클 필터링 내에서 모두 해결되는 장점이 있다. 5장에서 다루게 될 관측모델이 정확하다면 실시간 시스템에서 포즈를 나타내는 사후분포 추정은 제안된 방법의 예측분포로 인해 강건하게 이루어 질 수 있다.

5. 관측모델 (Observation Model)

5.1 관측모델 정의

관측모델이란 예측된 파티클의 가중치 즉 우도를 결정하기 위한 방법과 일련의 절차를 가리킨다. 관측모델의 계산 결과는 우도이지만 우도 계산을 위한 절차들은 다양한 형태로 정의될 수 있다. 본 논문에서 사용하는 관측모델은 실루엣과 에지 관측모델을 결합함으로써 파티클의 우도를 식 (13) 결합 우도로 계산한다.

$$p(I_t | \theta_t) = p(I_S | \theta_t) p(I_E | \theta_t) \quad (13)$$

식(13)의 결합우도는 실루엣 우도 $p(I_S | \theta_t)$ 와 에지 우도 $p(I_E | \theta_t)$ 를 결합한 것이다. 실루엣 우도는 예측된 θ_t 의 PS 모델의 실루엣이 실루엣 관측영상 I_S 와 겹침 정도를 의미한다. 실루엣 우도는 양손이 몸통과 겹쳤을 경우 우도의 변화가 없기 때문에 에지우도를 함께 사용한다. 에지영상에서 우도를 계산하는 것은 실루엣 우도에 비해 오류가 있지만 속도가 빠르고 신체 파트간 중첩현상에서도 우도를 계산할 수 있기 때문이다.

5.2 실루엣 관측 모델

본 논문에서의 실루엣 우도 계산을 위한 실루엣 관측 모델은 Makoto Kato[11]가 제안한 방법이며, 모델 실루엣영상과 포즈 실루엣영상의 차영상에서 나타난 실루엣 차이영역은 포즈의 오차로 볼수 있으며 이것을 통해 우도를 계산할 수 있다. 실루엣 관측 모델은 두 영상사이의 모든 픽셀에서 뺄셈연산이 수행되므로 속도가 느리지만, 실루엣 차이가 포즈의 차이를 직접적으로 나타내므로 실루엣 우도는 에지우도에 비해 정확하다.

(그림 5)는 실루엣 관측모델에서 추출된 영상들이다. (그림 5(a))는 비교 대상인 관측영상과 예측된 한 파티클의 PS 모델이다. (그림 5(b))는 비교대상인 관측영상의 실루엣 I_S 과, PS모델의 실루엣은 I_M 을 보여주고 있으며, I_S 는 GMM[18] 기반 배경제거를 통해 얻은 전경 영상이다. 모델의 실루엣 영상은 예측된 파티클의 PS 모델을 검은 배경 영상에 하얀색으로 출력한 영상이다.



관측영상 I_t PS모델(파티클)

(a) 관측영상 I_t 와 본 논문의 예측분포를 통해 예측된 한 파티클의 PS 상반신모델



실루엣 I_S 모델실루엣 I_M 합영상 I_O 차영상 I_{S-M} 차영상 I_{O-M} 차영상 I_{O-S}

(b) 실루엣 우도는 두 실루엣의 차영상 I_{S-M} 와 두 실루엣 합영상에서 관측영상 실루엣을 뺀 I_{O-M} 와 모델 실루엣을 뺀 I_{O-S} 차영상들로부터 계산됨

(그림 5) 실루엣 우도 계산을 위한 관측영상과 모델의 실루엣 차영상 계산 과정

실루엣 우도는 (그림 5(b))의 I_S 와 I_M 의 차분한 영상의 모든 픽셀 값의 합으로부터 계산되며 3가지 종류의 차영상을 이용한다. 첫 번째 차영상 I_{S-M} 은 I_S 와 I_M 의 서로 차분한 영상이며 $I_{S-M}=|I_S-I_M|$ 로 계산된다. I_{S-M} 은 두 실루엣영상의 교차하는 영역만 제거되고 교차하지 않는 영역이 차영상에 남는다. 합영상 $I_O=|I_S+I_M|$ 은 나머지 두 개의 차영상을 얻는데 사용된다. 두 번째 차영상 I_{O-M} 은 관측영상 실루엣에서 PS모델을 차분한 것과 같으며 $I_{O-M}=|I_O-I_M|$ 으로 계산된다. 세 번째 차영상 I_{O-S} 는 PS모델에서 관측영상의 실루엣을 차분한 것과 같으며 $I_{O-S}=|I_O-I_S|$ 으로 계산된다. I_{S-M} 은 서로 겹치지 않는 부분을, I_{O-M} 은 관측영상의 실루엣의 남은부분을, I_{O-S} 는 PS모델 실루엣의 남은 부분을 나타낸다.

실루엣 우도 $p(I_S|\theta_t)$ 는 (그림 5(b)) 3가지 차분영상을 이용하여 계산된 식(14)이다.

$$p(I_S|\theta_t) = p(I_{S-O}|\theta_t)p(I_{M-O}|\theta_t)p(I_{S-M}|\theta_t) \quad (14)$$

$$p(I_{S-O}|\theta_t) = \exp\left(-\frac{1}{wh} \sum I_{S-O}(x,y)\right) \quad (15)$$

$$p(I_{M-O}|\theta_t) = \exp\left(-\frac{1}{wh} \sum I_{M-O}(x,y)\right) \quad (16)$$

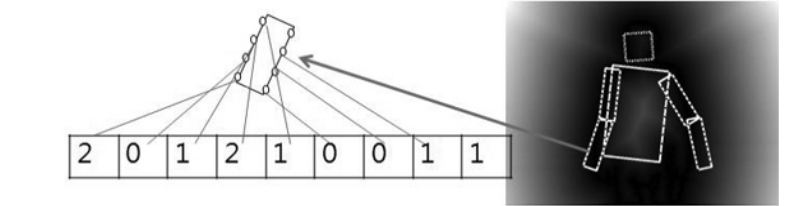
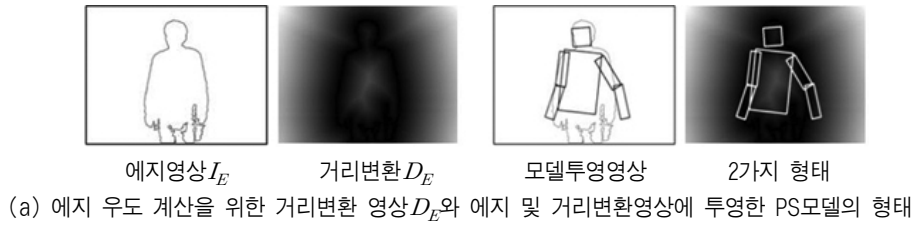
$$p(I_{S-M}|\theta_t) = \exp\left(-\frac{1}{wh} \sum I_{S-M}(x,y)\right) \quad (17)$$

식(14)에서 $p(I_S|\theta_t)$ 는 3가지 식(15)-(17)의 결합우도이다. 식(15)-(17)은 거리 값을 확률로 변환하는 Gibbs분포 $\exp(-|거리|)$ 를 이용한 3가지 차영상을 이용한 실루엣 우

도이다. 거리는 차영상의 픽셀 값을 모두 더한 $(w \times h)^{-1} \sum I_{차영상}(x,y)$ 값이며 이 합산된 값은 검은색 영역은 값이 0이고 하얀색영역은 값이 1이므로 차영상에서 나타난 하얀색 픽셀 수이다. $(w \times h)^{-1}$ 는 정규화 상수이며 w 는 영상 가로크기이고 h 는 세로 크기이다. 거리 값을 두 실루엣 영상이 틀릴수록 값이 증가하므로 Gibbs분포를 이용한 실루엣 우도는 두 실루엣의 유사도를 나타낼 수 있다.

5.3 에지 관측 모델

에지 우도 $p(I_E|\theta_t)$ 는 관측 영상 I_E 와 PS모델의 외곽선에 사이의 유사도를 통해 계산된다. 에지 우도 계산을 위한 두 에지영상과의 거리 계산은 챔퍼거리(Chamfer Distance) [12,14]를 기반으로 한다. 챔퍼거리는 에지 영상과 PS모델 사이의 거리를 효과적으로 계산할 수 있다. 챔퍼거리는 (그림 6(b))에 그 계산법이 개념적으로 도시되어 있다. 챔퍼거리는 (그림 6(b))의 PS모델 상의 특징점들의 위치에서 에지 거리 값들을 읽어오는 챔퍼매칭 과정을 통해 얻은 합산된 거리 값이다. PS모델 파트의 각 특징점 위치에서 얻은 픽셀 값들의 합이다. 이 픽셀 값들은 에지 거리 값이며, (그림 6(a))처럼 입력된 에지영상 I_E 를 거리변환(Distance Transform)을 통해 계산한 영상 D_E 의 픽셀 값들이다. 각 픽셀 값의 밝기는 상반신 실루엣 에지로부터의 거리의 정도를 비례적으로 나타내므로, 챔퍼거리가 클수록 관측영상과 PS모델간의 유사도는 낮아진다. 식 (17) 챔퍼거리는 PS모델 특징점 위치의 D_E 영상 픽셀 값들을 합산하여 특징점 개수 k 로 정규화 한 것이다.



(그림 6) 에지 우도 계산을 위한 PS모델 투영 후 특징점에서의 챔퍼거리 계산과정

$$v(I_E, \theta) = \frac{1}{3k} \sum_{j=1}^k D_E(x_p, x_j) \quad (17)$$

식 (17)의 v 함수는 에지 관측영상과 PS모델의 챔퍼매칭 함수이며, 결과는 식 우편의 챔퍼거리 값이다. v 함수는 에지 관측영상 I_E 를 거리변환 영상 D_E 를 계산한 뒤 θ 의 PS모델의 k 개의 특징점 위치에서 D_E 의 픽셀 값들을 합산한 값을 정규화한 챔퍼거리 값을 계산한다.

식 (18)은 에지 우도 $p(I_E|\theta_t)$ 를 의미하며 식 (17)에서 계산된 챔퍼거리를 Gibbs분포를 이용하여 유사도 값으로 바꾼 값을 나타낸다.

$$p(I_E|\theta_t) = \exp(-v(I_E, \theta_t)) \quad (18)$$

6. 추적 결과

실험 환경은 Intel i3 CPU와 2G RAM, 1394B 단안 카메라로 구성된다. 추적결과를 평가하기 위해서 자체 제작한 지상검증(Ground-Truth) 자료는 자체 제작한 GUI 프로그램을 통해 테스트 영상의 PS모델을 획득한 데이터베이스이다. 지상검증 자료와 추적결과와의 차이 값을 식 (19) RMS(Root Mean Square)를 통해 계산하였다.

$$RMS(\theta_t, \theta_G) = \sum_{x_i \in \theta_t, x_j \in \theta_G} \sum_{p_k \in x_i, p_l \in x_j} \sqrt{(p_k - p_l)^2} \quad (19)$$

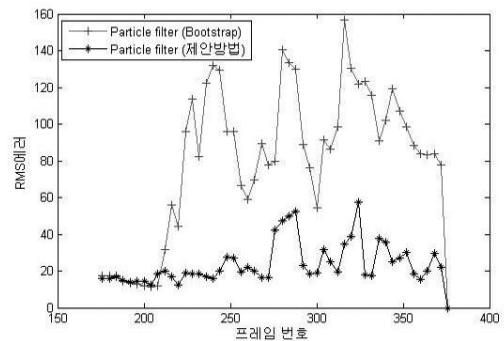
식 (19) RMS에러는 추적결과 θ_t 와 지상검증자료 θ_G 와의 거리 값을 계산하는 식이다. (그림 7)에서는 지상검증자

료의 PS가 굵은색으로 추적결과와 PS가 보다 얇은 색으로 나타나 있다. 정확한 계산을 위해서 서로 관련이 있는 두 모델의 파트 x_i 와 x_j 들의 거리를 모두 합산하였다. 거리 계산을 위해서 각 파트의 사각형 4 꼭짓점을 기준으로 했으며 두 모델의 서로 연관이 있는 꼭짓점 p_k 와 p_l 사이의 RMS거리를 기준으로 하였다.



(그림 7) RMS에러에 따른 결과 포즈 예

- (a) $0 \leq RMS \leq 20$ (b) $20 < RMS \leq 60$
- (c) $60 < RMS \leq 100$



(그림 8) 동적 움직임 모델을 사용한 경우와 본 논문에서 제시한 예측분포를 사용한 경우의 추적 RMS 에러 결과 그래프



(a) 제안 방법의 추적결과 (사용한 파티클 개수 : 200)



(b) 기존 방법의 추적결과 (사용한 파티클 개수 : 400)

(그림 9) 다양한 사용자에게 적용한 상반신 추적 비교결과

RMS 값에 따른 PS추적 결과를 나타낸 (그림 7)은 3개의 비교 영상이 있으며 비교 결과를 수치적으로 표현하였다. (그림 7)에서 RMS에러가 20이하일 경우 지상검증 자료와 추적결과가 거의 흡사한 것을 볼 수 있다. RMS에러가 20 이상 60 이하일 경우 PS 파트 한 개 정도가 지상검증 포즈로부터 벗어난 것을 볼 수 있으며 60이상 100 이하일 경우 두 개 파트가 크게 벗어난 것을 확인할 수 있다.

추적 결과 평가는 비디오영상 각 프레임에서 RMS에러를 계산하여 수행되었다. (그림 8)은 비디오 영상에서 테스트한 RMS에러를 그래프로 표현한다. (그림 8)은 두 개 방법의 RMS 결과를 나타낸다. 첫 번째 방법은 예측분포를 단순히 마코프 모델을 이용한 방법인 Bootstrap이라 알려진 파티클 필터이며, 두 번째 방법은 본 논문에서 제안한 키포즈 기반 예측분포를 이용한 파티클 필터이다. 동적 움직임 모델만 사용할 경우에 비해 성능이 향상되었는지 점검하였으며 (그림 8)의 그래프에 의하면 키포즈 기반 예측분포는 동적인 모션 모델에 대해 더 나은 성능을 보이는 것을 확인할 수 있다. Bootstrap기반 PF는 평균 RMS가 81.44였으며 제안된 방법은 평균 RMS가 29.4838이었다. 제안된 방법은 기존 방법의 성능을 70.51% 향상 시켰다.

(그림 9)는 상반신 포즈 추적 비교 영상이다. 기존 방법에 비해 나은 추적 결과를 갖는다. 또한 다양한 사용자에 적응적으로 추적이 가능하다. 실험자의 옷 색상과 무늬, 신체 크기 및 성별에 구애받지 않는 것을 확인할 수 있다. 기존 Bootstrap 기반 PF의 경우에는 갑작스런 비선형적인 포즈변화에 잘 대처하지 못하는 문제점을 드러냈었지만 제안된 방법은 포즈 변화가 크더라도 실패 없이 상반신 포즈를 추적할 수 있었으며 간혹 큰 에러가 생기더라도 복원될 수 있는 장점을 나타내었다.

포즈 추적 대상은 반드시 카메라를 정면으로 주시해야 한다는 단점이 있지만 (그림 9)에서 사용자가 정면을 주시하지 않아도 추적이 가능한 상황이 발생하는 것을 볼 수 있기 때문에 앞으로 정면 외 포즈 추적 문제도 해결할 수 있을 것으로 파악된다. 단점이 발생한 이유는 얼굴 정면 검출기 (Viola&Jones Adaboost)를 이용하여 전경에서 얼굴이 검출될 경우 최초 추적이 이루어지도록 시스템이 설계되었고 그때 얼굴 크기를 기준으로 전체적인 신체 사이즈를 조절했기 때문이다. 전경을 사람인 것을 판단할 수 있는 방법이 얼굴 검출기에 의존적이지 않고, PS 신체 사이즈를 실루엣 크기에 맞추는 방법이 얼굴 크기에 의존되지 않도록 하는 방법이 있다면 얼굴 주시방향에 따른 상반신 포즈 추적이 실패하는 단점은 개선될 수 있다. 또 다른 문제점으로는 현재 사용자의 팔이 몸통 안으로 들어올 경우 추적이 어려운 단점이다. 이는 관측 모델의 경우 실루엣 영상에 크게 의존하고 있고, 예지 관측 모델이 있지만 실루엣 관측모델에 비해 충분한 단서를 제공하지 못하고 있기 때문이다. 이는 옷의 텍스처로부터 예지영상이 발생하고 있기 때문으로 보인다.

신체 파트간의 중첩현상을 해결하기 위한 문제점 해결을 위해서는 앞으로 많은 연구가 필요할 것으로 보인다.

7. 결 론

본 논문에서는 상반신 포즈 추적을 위해 시간에 따른 포즈변화를 예측하기 위한 키포즈 기반 예측분포에 대한 방법을 서술하였다. 제안된 방법은 키포즈 라이브러리와 입력영상과의 유사도를 계산함으로써 파티클을 예측할 때 사후확률이 낮파티클들은 키포즈와 유사한 포즈 구성벡터를 갖도록 유도하는 방법이다. PF 예측분포에 키포즈 라이브러리 검색 과정을 결합하여 구현하였다. 실험결과를 통해 제안된 방법 기존의 방법에 비해 70.51%의 성능향상을 보여주었다. 향후 연구에서는 알고리즘 최적화 및 보다 외관이 자연스러운 PS를 설계할 예정이며, 제스처 인식을 통해 로봇-인간 상호작용에 적용하는 것이 최종 목표이다.

참 고 문 헌

- [1] M. Andriluka, S. Roth and B. Schiele, "Pictorial Structures Revisited: People Detection and Articulated Pose Estimation", Int. Conf. Computer Vision & Pattern Recognition, pp.1014-1021, 2009.
- [2] H. Jiang, "Human Pose Estimation Using Consistent Max-Covering", Int. Conf. Computer Vision, pp.1357-1364, 2009.
- [3] X. Zhang, C. Li, X. Tong, W. Hu, S. Maybank, and Y. Zhang, "Efficient human pose estimation via parsing a tree structure based human model", Int. Conf. Computer Vision, pp. 1349-1356, 2009.
- [4] A. Agarwal and B. Triggs, "A Local Basis Representation for Estimating Human Pose from Cluttered Images", Asian Conf. Computer Vision, pp. 50-59, 2006.
- [5] O. Freifeld, A. Weiss, S. Zuffi and M. J. Black, "Contour People: A Parameterized Model of 2D Articulated Human Shape", In Int. Conf. Computer Vision and Pattern Recognition, pp. 639-646, 2010.
- [6] S. Maskell, "A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking", IEEE Trans. Signal Processing, Vol. 50, No. 2, pp. 174-188, 2002.
- [7] K. M. Manson, "Markov Chain Monte Carlo Posterior Sampling with the Hamiltonian method", Proc. Sensitivity Analysis of Model Output, pp. 456-467, 2001.
- [8] M. W. Lee and I. Cohen, "Proposal Maps driven MCMC for Estimating Human Body Pose in Static Images", In Int. Conf. Computer Vision and Pattern Recognition, vol. 2, pp.334-341, 2004.

[9] K. Okuma A. Taleghani N. de Freitas J. Little and D. Lowe, "A Boosted Particle Filter: Multitarget Detection and Tracking", European Conf. Computer Vision, pp. 28-39, 2004.

[10] S. X. Ju, M. J. Black and Y. Yacoob, "Cardboard People: A Parameterized Model of Articulated Image Motion", Int. Conf. Automatic Face and Gesture Recognition, pp. 38-44, 1996.

[11] Makoto Kato, Yen-Wei Chen and Gang Xu, "Articulated hand tracking by pca-ica approach", Int. Conf. Automatic Face and Gesture Recognition, pp. 329-334, 2006

[12] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles and H. C. Wolf, "Parametric Correspondence and Chamfer Matching: Two New Technique for Image Matching", Int. Conf. Artificial Intelligence, pp. 1175-1177, 1997.

[13] E. A. Wan and R. van der Merwe, "The Unscented Kalman Filter for Nonlinear Estimation", Adaptive Systems for Signal Processing, Communications, and Control Symposium, pp. 153-158, 2000.

[14] Borgefors and Gunilla "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 6, pp. 849-865, Nov 1988.

[15] R. Kindermann and J. L. Snell, "Markov Random Fields and Their Applications", American Mathematical Society, Providence, RI, 1980.

[16] K. Nishiyama, "Fast and Effective Generation of the Proposal distribution for Particle filters", Vol 85(12), pp. 2412-2417, 2005.

[17] J. Bourgain, On Lipschitz, "Embeddings of Finite Metric Spaces in Hilbert Space", Israel Journal of Mathematics, Vol. 52, pp. 46-52, 1985.

[18] N. A. Setiawan, S. J. Hong, J. W. Kim and C. W. Lee, "Gaussian Mixture Model in Improved HLS Color Space for Human Silhouette Extraction", Int. Conf. Artificial Reality and Telexistence, pp. 732-741, 2006.



오 치 민

e-mail : sapeyes@gmail.com
 2007년 전남대학교 전자컴퓨터공학부(학사)
 2009년 전남대학교 대학원 전자컴퓨터
 공학부(석사)
 2009년~현 재 전남대학교 전자컴퓨터
 공학부 박사과정

관심분야: 컴퓨터 공학, 객체 추적, 제스처인식, 인간-컴퓨터 상호작용



이 철 우

e-mail : leecw@chonnam.ac.kr
 1986년 중앙대학교 전자공학과(학사)
 1988년 중앙대학교 대학원 전자공학과
 (공학석사)
 1992년 동경대학교 대학원 전자공학과
 (공학박사)

1992년~1995년 이미지 정보과학 연구소 수석 연구원 오사카 대학 기초공학부 협력연구원 및 리츠메이칸 대학 특별 초빙강사

1996년 1월~현 재 전남대학교 전자컴퓨터공학부 교수

2002년 1월~2003년 2월 미국 NC A&T State University 방문교수

2006년 3월~2008년 2월 정보통신부 자체평가위원

2008년 10월~2009년 8월 KAIST 및 오사카대학 초빙교수

2007년 1월~2008년 8월 전남대학교 산학연구부처장 겸 산학협력부단장

2006년 3월~현 재 전남대학교 문화콘텐츠기술연구소 소장

2009년 3월~현 재 전남대학교 차세대휴대폰인터페이스연구 센터(ITRC) 센터장

관심분야: 컴퓨터비전 및 영상처리 응용기술, 지능형 휴먼 인터페이스, 실감형 디지털콘텐츠 표현 및 제작