

논문 2012-49SP-3-6

신뢰 전파와 디스패리티 맵을 사용한 다관절체 사람 추적

(Articulated Human Body Tracking Using Belief Propagation with Disparity Map)

윤 광 진*, 김 태 용**

(Kwang-Jin Yoon and Tae-Yong Kim)

요 약

본 논문에서는 마르코프 네트워크로 모델링된 다관절체(Articulated body) 사람을 양안 영상(stereo image)을 통해 획득된 디스패리티 맵(disparity map)을 이용해 효과적으로 추적하는 방법을 제안한다. 기존의 색상 정보만을 사용하여 에너지 함수의 우도(likelihood)를 계산하는 방법은 조명 및 그림자의 영향과 배경 색상의 임의성 때문에 강건하지 못 하다. 본 논문에서는 색상 정보에 더불어 디스패리티 정보를 활용하여 우도를 계산하는 방법을 제안한다. 원통형 모양의 사람의 신체 요소(body part)는 2차원 영상으로 사영될 때 직사각형으로 사영되므로 이 직사각형의 디스패리티의 분포가 불연속 하지 않다는 특성을 이용한다. 또한 본 논문에서는 디스패리티 맵을 사용한 조건적 메시지 생성 방법을 제안해 신뢰 전파에서 불필요한 메시지 업데이트 수행을 줄이는 방법을 보여준다. 메시지 업데이트는 신뢰 전파 알고리즘의 전체 수행 시간에 80% 이상을 차지하므로, 조건적 메시지 생성 방법은 기존 대비 9~45%의 속도 향상을 보였다. 또한 사람의 연속적인 움직임 특성을 이용한 다이나믹 모델을 제안해 추적 속도를 향상하였다. 자세한 내용은 4장에 설명되어 있다. 실험 결과 제안하는 디스패리티 정보를 활용한 신뢰 전파를 사용해 다관절체를 추적하는 방법은 기존 대비 강건한 추적 결과와 함께 빠른 속도로 추적할 수 있었다.

Abstract

This paper suggests an efficient method which tracks articulated human body modeled with markov network using disparity map derived from stereo images. The conventional methods which only use color information to calculate likelihood for energy function tend to fail when background has same colors with objects or appearances of object are changed during the movement. In this paper, we present a method evaluating likelihood with both disparity information and color information to find human body parts. Since the human body part are cylinder projected to rectangles in 2D image plane, we use the properties of distribution of disparity of those rectangles that do not have discontinuous distribution. In addition to that we suggest a conditional-messages-update that is able to reduce unnecessary message update of belief propagation. Since the message update has comprised over 80% of the whole computation in belief propagation, the conditional-message-update yields 9~45% of improvements of computational time. Furthermore, we also propose an another speed up method called three dimensional dynamic models assumed the body motion is continuous. The experiment results show that the proposed method reduces the computational time as well as it increases tracking accuracy.

Keywords : Articulated human body tracking, Belief Propagation, Disparity map, likelihood

* 정희원, ** 평생회원-교신저자, 중앙대학교 첨단영상대학원,

(Graduate School of Advanced Image Science, Multimedia, and Film, Chung-Ang University)

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업(NIPA-2010-C1090-1011- 0010)과 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업(2010-0021892) 지원을 받아 수행된 것임.

접수일자: 2011년9월30일, 수정완료일: 2012년1월30일

I. 서 론

영상 속의 사람을 다관절체(Articulated body)로 모델링 하여 움직임을 분석하고 각 신체 요소(Body Parts)를 추적하는 방법에 대한 연구는 실제로 인간-컴퓨터 상호작용 (Human-Computer Interaction) 및 모션 캡처, 비디오 감시 등의 분야에 널리 사용된다. 이와 같은 연구는 컴퓨터 비전 분야에서 꾸준히 연구되어왔으며 도전적인 분야로 분류된다. 그 이유로 첫째, 사람의 움직임은 높은 자유도(보통 20~68)를 가지고 있으며 둘째, 조명 또는 그림자의 영향으로 겉모습(appearance)의 변화가 잦기 때문이다. 셋째는 각 신체 요소끼리의 가려짐(Occlusion)이 생기는 것이며 넷째, 옷차림, 피부색 등의 이유로 일반적인 신체 요소의 색상을 사전에 알기 힘들다는 것이다. 다섯째로는 사람의 팔, 다리의 빠른 움직임 속도를 들 수 있다^[2].

다관절체(articulated body)로써 사람 추적은 칼만 필터^[7], 파티클 필터, 신뢰 전파^[2, 4, 9, 11] 등의 베이즈 추론 알고리즘을 이용한 방법으로 강건한 추적을 제안하는 논문들이 많이 쓰여 졌다. T.X Han^[2]은 2차원 추적기로써 사람의 신체 요소를 마르코프 랜덤 필드(Markov Random Field)로 모델링 하여 신뢰 전파를 통한 메시지의 전파로 각 신체 요소의 위치를 추론하는 방법을 제안 하였다. 그의 방법은 최초 추적 후 인접 위치에서 추적 하도록 해 기존보다는 빠른 추적 속도를 보이고 있지만 여전히 실시간 처리 속도에 못 미친다는 단점과 RGB 공간에서의 템플릿 매칭으로 조명의 변화에 민감하다는 단점이 존재한다. D. Ramanan^[4]은 특정 자세 (pose)에서 사람의 겉모습(색상 정보)을 취득하여 추적 하는 방법에 대하여 연구 하였다. 그의 방법은 색상 정보 이후 강건한 추적이 가능하지만 취득을 위해 필요한 제약 조건이 많으며 다수의 파라미터가 존재 하다는 것이 문제가 된다. P. Felzenswalb^[3, 5]는 그의 논문에서 각 관절들의 효율적인 표현과 함께 동적 계획법 (Dynamic Programming) 및 거리 변환(Distance Transform) 방법으로 계산 복잡도가 작은 알고리즘을 제안 하였다. 하지만 배경차영상을 사용하였기 때문에 변화하는 배경에 적용이 불가능하다는 것과 각 신체 요소가 영상 속에서 존재 할 수 있는 위치를 이산화 시켜 연속적인 움직임을 추적하기에 부적합 하다는 단점이 있다.

본 논문에서는 디스패리티 맵을 신뢰 전파에 사용하

여 영상 내 사람의 신체 요소를 효율적으로 찾는 방법을 제안한다. 더불어 조건적 메시지 생성 및 3차원 다이나믹 모델(Dynamic model)을 통해 속도를 향상하는 방법을 제안한다.

다관절체를 추적 하기위해 영상 속의 객체(Object)를 검출 하는데 있어서 배경차영상은 탐색 공간을 줄이는데 유효한 방법이지만 카메라나 배경이 움직이는 상황 등과 같이 일반적인 상황에 적용하기에는 부족하다^[11]. 제안하는 방법은 전처리 과정으로써 매 프레임 배경의 색상 정보를 업데이트해 영상 속에서 객체가 위치 할 수 있는 후보 영역을 줄이게 된다. 또한 색상 정보만을 사용하게 되면 객체 주위에 비슷한 색상이 존재하게 될 때 잘 못 추적하는 오류를 발생하게 된다^[10]. 이런 단점을 극복하기 위해 3차원에서 2차원 공간으로 사영되어진 신체 요소의 디스패리티 분포의 특성^[11]을 이용하여 색상 정보와 함께 우도를 계산하는 방법에 대해 제안한다. 자세한 방법은 3.2절에 설명 되어있다. 또 사람의 움직임은 영상 속에서 연속적으로 움직이기 때문에 현재 프레임에서의 객체의 위치는 이전 프레임의 추적 결과로부터 근처에 위치하게 된다. 본 논문에서는 3차원 다이나믹 모델을 통해 이전 프레임의 정보를 현재 프레임의 추적에 사용하여 계산 속도와 강건함을 높이는 방법에 대해 제안한다.

본 논문의 구성은 다음과 같다. II장에서는 신뢰 전파를 다관절체 추적에 사용하는 방법을 소개한다. III장에서는 본 논문에서 제안하는 디스패리티 맵을 신뢰 전파에 사용하여 다관절체로써의 사람 관절 추적을 하는 방법 및 3차원 다이나믹 모델과 3차원 연결 제약에 대해서 설명하였으며, 제안하는 조건적 메시지 생성법에 대해서도 알아본다. IV장에서 실험 결과를 보여준다.

II. 신뢰 전파와 다관절체로써의 사람 추적

신뢰 전파는 마르코프 랜덤 필드나 베이지안 네트워크와 같은 그래프 형태의 모델에서 추론을 하는 알고리즘으로써 지역적(local) 메시지 전파를 통해 글로벌(global) 한 추론 결과를 내는 알고리즘이다^[6].

본 논문에서 사용하는 마르코프 랜덤 필드는 그림 1과 같이 순환이 없는 트리 형태의 마르코프 랜덤 필드이다. 간략한 표현을 위해 관측 노드는 생략하였다. 그림 1과 같은 형태의 마르코프 네트워크에서 은닉 노드 X 의 최적의 상태 추론은 식 (1)의 에너지 함수가 최소

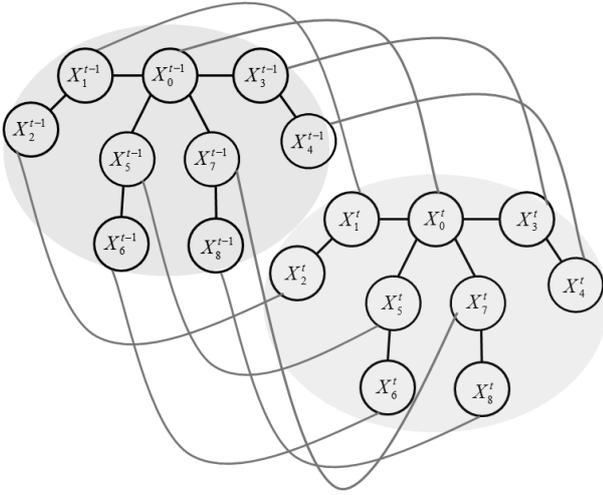


그림 1. 다관절체 사람 추적을 위한 마르코프 네트워크
Fig. 1. Markov network for articulated human body.

가 되는 값을 찾음으로써 구할 수 있다^[8]. [8]에서는 에너지 함수를 결합 확률 모델링하여 에너지 함수가 최대가 되는 값으로 최대 사후확률을 계산하였다. 본 논문에서는 이 에너지 함수의 양변에 음의 로그 함수를 취하였기 때문에 결과적으로 식 (1)이 최소가 되는 값을 찾는 것과 같다^[3].

$$P(X_1, X_2, \dots, X_N, Y_1, Y_2, \dots, Y_N) = \prod_{(i,j)} \Psi_{i,j}(X_i, X_j) + \sum_k (\Psi_k^c(X_k, Y_k) + \Psi_k^t(X_k, X_k^{t-1})) \quad (1)$$

한 노드가 다른 모든 노드로부터 정보를 얻기 위해서는 트리의 깊이만큼의 반복적인 메시지 업데이트를 수행하여야 한다^[8]. 체요소 k 에서 j 로의 메시지 업데이트는 식 (2)와 같이 하게 된다.

$$M_j^k = \min_{X_k} (\Psi_{j,k}(X_j, X_k) + \Psi_k^c(X_k, Y_k) + \Psi_k^t(X_k, X_k^{t-1}) + \sum_{l \neq j} \tilde{M}_k^l) \quad (2)$$

\tilde{M}_k^l 는 M_k^l 의 이전 반복(iteration)에서 계산된 메시지이다. 최초의 메시지 \tilde{M}_k^l 는 0으로 설정 된다.

$\Psi_k^c(X_k, Y_k)$ 는 신체요소 X_k 가 특정 위치에 존재 할 때 t 번째 입력 영상 I^t 와 디스패리티 지도 D^t 와의 우도를 나타낸다. 즉 확률 $P(I^t, D^t | X_k, T_k)$ 과 같다. T_k 는 신체 요소 X_k 의 템플릿이다.

$\Psi_k^t(X_k^t, X_k^{t-1})$ 은 다이나믹 모델로써 확률 분포 $P(X_k^t | X_k^{t-1})$ 과 같다. 본 논문의 다이나믹 모델은 사람

의 연속적인 움직임 특성을 이용한다. 즉 어떤 신체요소 i 가 t 시간에 존재하는 위치 X_i^t 는 $t-1$ 시간의 위치 X_i^{t-1} 의 이웃에 존재한다는 성질을 이용한 모델이다.

$\Psi_{i,j}(X_j, X_i)$ 은 연결 제약 모델로써 확률 $P(X_i | X_j)$ 과 같다. 이웃한 신체 요소끼리의 상관성으로 사람의 신체 요소가 사람의 모양처럼 연결 되도록 해준다.

모든 노드가 메시지 업데이트를 완료하면 식 (3)을 이용하여 신체 요소 j 의 위치를 추론한다.

$$X_{jMAP} = \arg \min_{X_j} (\Psi_j^c(X_j, Y_j) + \Psi_j^t(X_j, X_j^{t-1}) + \sum_k M_j^k) \quad (3)$$

식 (3)을 이용해 추론된 노드의 상태는 최대 사후 확률이 된다^[8].

III. 제안 방법

1. 색상 정보를 이용한 픽셀 분류

본 논문에서는 신체 요소의 색상 정보(템플릿)를 첫 프레임이나 각 신체 요소끼리의 가려짐이 없는 프레임으로부터 획득하여 색상 정보를 가지고 있다고 가정한다. 획득한 신체 요소의 템플릿으로 각각의 히스토그램을 식(4)와 같이 모델링 할 수 있다^[12].

$$P(YCrCb | bodypart_i) = \frac{bodypart_i[YCrCb]}{T_{bodypart_i}} \quad (4)$$

$T_{bodypart_i}$ 는 i 번째 신체 요소의 총 픽셀 수이고 $bodypart_i[YCrCb]$ 는 $YCrCb$ 의 색상을 갖는 i 번째 신체 요소의 픽셀의 개수 이다. 그리고 배경의 색상 모델은 식 (5)로 구해진다.



그림 2. 팔 색상으로 픽셀 분류가 된 그림
Fig. 2. classified pixel by color of arm.

$$P(YCrCb|Background) = \frac{Background[YCrCb]}{T_{Background}} \quad (5)$$

$T_{Background}$ 은 모델에 사용된 배경 픽셀의 총 개수이고, $Background[YCrCb]$ 는 $YCrCb$ 의 색상을 갖는 배경 픽셀의 개수이다. 배경 색상을 얻는 방법은 최초 추적 결과에서 신체 요소를 제외한 주변의 픽셀들을 이용함으로써 배경의 히스토그램을 모델링 할 수 있다. 여기서 사용된 히스토그램은 조명의 변화에 강건하게 하기 위해 YCrCb 색상 공간을 사용하였다. 이렇게 각각 모델링 된 히스토그램을 이용해 식 (6)의 베이저안 추론을 사용하여 픽셀들을 그림 2와 같이 신체 요소(흰색)와 배경(검은색)으로 분류 하였다^[12].

$$P(YCrCb|Bodypart_i) > \lambda P(YCrCb|Background) \quad (6)$$

λ 는 임계치로써 이용되는 $P(YCrCb|Background)$ 의 값을 선형적으로 높여주어 픽셀의 분류를 강건히 하기 위해 사용된 상수이다. 실험을 통해 알맞은 값을 설정 할 수 있었으며, 본 논문에서는 1.9로 설정 하였다.

본 논문에서는 색상 정보를 이용한 우도는 획득한 템플릿과 입력 영상과 템플릿 매칭을 수행하여 계산된다. 템플릿 매칭은 계산 부담이 높아 전체 알고리즘의 수행 속도 저하를 야기한다. 따라서 분류된 픽셀 정보를 이용하여 신체 요소로 분류된 픽셀 위치에서만 템플릿 매칭을 수행하는 것은 연산 수행 시간을 단축시켜 효율적인 방법이 된다. 또한 이러한 방법은 영상 속에서 효과적으로 각 신체 요소를 샘플링 할 수 있게 하는 방식이다^[4]. P.Felzenswalb^{[3][5]}는 각 신체 요소가 존재 할 수 있는 위치를 이산화 시켜 연속적인 신체의 움직임을 표현하는데 어려움이 있었는데 반해 샘플링을 통한 방식으로 사람의 연속적인 움직임을 표현할 수 있다. 또한 분류된 픽셀은 추후 추적에서 우도를 계산하는데도 사용된다. 자세한 설명은 3.2절에서 이어진다. 그리고 사람 모습이 대칭성을 이루는 성질 이용해 팔, 다리의 상, 하박은 하나의 템플릿으로 좌/우 같은 색상 모델로 모델링하였다^[4].

2. 디스패리티 맵을 사용한 다관절체 사람 추적

본 논문에서는 우도(likelihood)를 계산하기 위해 디스패리티 정보와 색상 정보를 사용한다. 디스패리티 정보는 특정 영역의 디스패리티 분포 특성을 활용한다. 따라서 디스패리티 지도를 정확히 구해내는 것이 중요

한 것이 아니므로 빠른 계산 속도를 가지는 블록 매칭을 사용하였다. 블록 매칭은 절대차합(sum of absolute difference) 알고리즘을 사용하였다.

다관절체로써 사람을 추적함에 있어서 우도 함수 $\Psi_k^c(X_k, Y_k)$ 의 계산 결과는 연결성 제약이나 다이나믹 모델의 강건함보다도 추적기의 성능을 좌우하는 중요한 요소이다^[4]. 우도 함수는 확률 분포 $P(Y_i|X_i)$ 과 비례하게 모델링 되도록 해야 한다. 즉 우도는 한 장면의 영상과 디스패리티 정보를 이용해 특정 위치의 신체 요소에 대한 확률을 계산한다. 템플릿과 분류된 픽셀 정보를 색상 정보, 디스패리티 맵을 디스패리티 정보라 할 수 있다. 본 논문에서 제안하는 방법은 템플릿 정보, 분류된 픽셀 정보, 디스패리티 정보를 종합해 우도 함수 $\Psi_k^c(X_k, Y_k)$ 을 계산하는 방법을 제안한다.

본 논문에서는 기존의 마르코프 랜덤 필드의 에너지 함수에서 음의 로그를 취해 최소가 되는 값을 찾고 있다. 따라서 우도 함수 $\Psi_i^c(X_i, Y_i)$ 은 최솟값이 될수록 $P(Y_i|X_i)$ 의 확률이 높아짐을 의미한다. 템플릿은 입력 영상과의 제곱차합(Sum of Squared Difference) 매칭 방법으로 색상 정보를 활용하였다. 템플릿과 입력 영상과의 제곱차합 C_{ssd} 는 식 (7)과 같다.

$$C_{ssd}(X_i) = 1 - \exp\left(-\sum_{x,y} (T(x',y') - I(x+x',y+y'))^2\right) \quad (7)$$

x 와 y 는 X_i 의 성분이고 $X_i = (x,y,\theta)^T$ 이다. T 는 템플릿이고 I 는 입력 영상이다. 제안하는 템플릿을 이용한 우도함수는 1에서 빼도록 해 매칭이 잘 될수록 작은 값을 가지게 한다. 그리고 전처리 과정을 통해 획득된 신체 요소/배경으로 분류된 픽셀은 신체 요소 크기의 직사각형 영역에 잘못 분류된 픽셀의 수에 비례한 값을 더하도록 한다. 왜냐하면 우도 함수는 매칭이 잘 될수록 최소값을 가지게 되기 때문에 잘 못 분류된 픽셀에 비례한 값을 더하여 구한다.

색상 정보만을 이용해 우도를 구하게 될 경우 배경에 비슷한 색상이 존재하거나 조명의 변화로 인해 오류를 발생하게 된다. 이런 경우 배경과 전경의 디스패리티 값의 차이를 이용하면 우도를 강건하게 계산하게 할 수 있다. 사람의 각 신체 요소는 3차원 공간에서 원통형 모양으로 생겼기 때문에 2차원 이미지에 사영될 경우 직사각형으로 사영된다. 즉 하나의 신체 요소는 직사각

형으로 모델링되며 사영된 신체 요소의 디스패리티는 연속적이거나 균일한 기울기를 가지게 된다. 따라서 신체 요소를 가로, 세로로 가로지르며 디스패리티의 불연속 정도를 사용한다. 불연속 정도 $G(X_i)$ 는 신체 요소의 중심을 포함해 가로와 세로로 가로질러가며 기울기를 구하는데 기울기가 불연속적이거나 급격히 변화할수록 큰 값을 가진다.

$$G(X_i) = \frac{N(|d(x_\theta, y_\theta) - d(x'_\theta, y'_\theta)| > t)}{N(p)} \quad (8)$$

(x_θ, y_θ) 는 신체요소 X_i 를 가로지르는 픽셀 위치이고 (x'_θ, y'_θ) 는 가로지르는 방향을 따라 바로 이웃한 픽셀의 위치이다. 첨자 θ 가 붙은 이유는 X_i 의 회전을 반영해야 하기 때문이다. $d(\cdot)$ 는 해당 위치에서의 디스패리티를 의미한다. p 는 신체 요소를 가로, 세로로 가로지르는 경로에 위치한 픽셀들을 의미하고 $N(\cdot)$ 는 조건을 만족하는 픽셀의 개수를 의미한다. 즉 분모의 $N(p)$ 는 정규화를 위한 상수이다. 본 논문의 실험에서 t 는 2로 설정하였다. 직사각형을 가로, 세로로 가로지르며 디스패리티의 불연속성에 대해 검사하는 이유는 가로, 세로 직선이 만나는 중점이 직사각형의 중심점이 되기 때문에 직사각형이 좌우로 움직이거나 회전을 하더라도 반드시 가로와 세로의 직선을 지나게 되기 때문이다. 이러한 성질을 이용해 직사각형 영역의 불연속을 경제적으로 검사할 수 있다. 그러면 색상 정보와 디스패리티 정보를 활용한 우도 함수를 구할 수 있다. 우도는 최종적으로 식(9)에 의하여 색상 정보와 디스패리티 정보를 동시에 활용하여 구해진다.

$$\Psi_i^t(X_i, Y_i) \cong C_{ssd}(X_i) + \frac{\text{misclassified pixels}}{h \cdot w} + G(X_i) \quad (9)$$

다이나믹 모델 $\Psi_i^t(X_i^t, X_i^{t-1})$ 은 확률 $P(X_i^t | X_i^{t-1})$ 에 비례한다. 따라서 다이나믹 모델은 다음의 수식을 이용하여 계산한다.

$$\Psi_i^t(X_i^t, X_i^{t-1}) \cong I(|X_i^t - X_i^{t-1}| < T_v) \quad (10)$$

$I(\cdot)$ 는 조건식을 만족할 경우 0, 만족하지 않을 경우 ∞ 를 가진다. T_v 는 벡터로써 $(S_x, S_y, S_d)^T$ 의 3차원 벡터를 이루며 각각 x 축, y 축 탐색 공간 및 디스패리티 지도에서 깊이 탐색 공간을 의미한다. 다이나믹 모델은 T_v 가 3차원 벡터를 이루므로 3차원 다이나믹 모

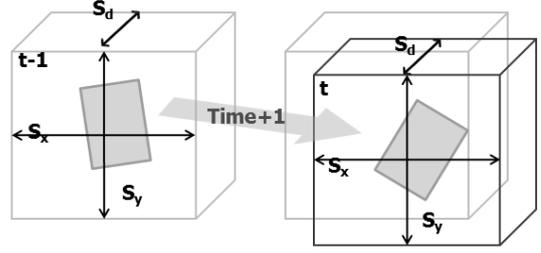


그림 3. 3차원 다이나믹 모델, 노란 직육면체는 $t-1$ 시간의 T_v 를 의미하며 초록 직육면체는 t 시간의 T_v 를 의미한다

Fig. 3. Three dimensional dynamic model, yellow cube represents T_v of $t-1$ time and green cube represents T_v of t time.

델을 형성한다. 본 논문에서 제안하는 에너지 함수는 가장 작은 값에서 최대 사후 확률을 가지게 된다. 따라서 ∞ 를 가진다는 것은 이전 시간의 위치에서 T_v 를 벗어난 곳에서는 메시지를 생성하지 않는다는 것과 동일한 의미이다. 이와 같은 방법으로 탐색 공간을 줄일 수 있어 연산 부담을 덜 수 있게 된다.

따라서 제안하는 3차원 다이나믹 모델은 사람의 연속적인 움직임 특성을 이용하여 어떤 신체요소 i 가 t 시간에 존재하는 위치 X_i^t 는 $t-1$ 시간에 존재 하던 위치 X_i^{t-1} 의 이웃에 존재한다는 성질을 이용한 모델이다. 이에 따라 $t-1$ 시간의 위치를 알게 되면 t 시간에는 이전 위치에서부터 정해진 거리의 3차원 범위만큼만 탐색하게 하는 모델이다(그림 3).

연결 제약 모델은 [3]에서 사전 확률로 사용되었다. 본 논문에서도 사전 확률로써 이웃 한 신체 요소끼리의 연결 제약을 나타내는 함수로 사용된다. 연결 제약 모델 $\Psi_{i,j}(X_j, X_i)$ 확률 $P(X_j | X_i)$ 로 나타내어진다. i 와 j 는 서로 이웃한 신체 요소이다. 식 (11)은 연결 제약 모델을 나타낸다.

$$\Psi_{i,j}(X_j, X_i) \cong 1 - \exp(-|K_{ij} - K_{ji}|) \quad (11)$$

K_{ij} 는 신체 요소 i 에서 j 로 연결되는 관절의 위치 및 i 의 디스패리티이다. $|K_{ij} - K_{ji}|$ 는 다음과 같다.

$$|K_{ij} - K_{ji}| = |x_{ij} - x_{ji}| + |y_{ij} - y_{ji}| + |d_i - d_j| \quad (12)$$

d_i 는 신체 요소 i 의 평균 디스패리티로 정해진다. 따라서 관절의 위치 K_{ij} 는 디스패리티 정보를 삽입하여 3차원 벡터를 형성하게 되며, 두 관절의 거리는 L1 거리

로 정의 된다. 이러한 연결 제약 모델은 3차원 연결 제약 모델로 정의 할 수 있으며, 이와 같은 방법은 2차원 추적기로서 보다 강건한 연결 제약을 나타낼 수 있다. 기존의 방법^[2-5, 7]은 색상 등의 이유로 혼란을 야기시키는 배경에서 위치를 추론할 때 실패할 경우가 있었지만 3차원 연결 제약으로 이웃한 신체 요소간의 디스패리티의 차이로 보다 강건한 추론에 성공하게 된다.

3. 조건적 메시지 업데이트 방법

영상의 화각(Field of View)은 영상 속의 객체의 크기를 결정 하는 요소이다. 화각이 넓은 영상일수록 영상 속의 객체는 작아지며 더 넓은 시야를 가지게 된다. 따라서 영상의 종류를 그림 4와 같이 원거리 영상 및 중간 거리 영상, 근거리 영상 세 가지로 분리 할 수 있다^[1].

첫째는 원거리 영상으로 야구 중계와 같이 먼 거리에서 야구장의 전반을 한 프레임에 담은 영상을 의미한다. 둘째는 중간 거리 영상으로 영상 속에 사람 한 명 정도의 신체가 전체로 잡히는 영상으로 중간 거리의 영상에서는 영상 속의 사람의 몸통과 팔, 다리를 식별 할 수 있다. 셋째로는 근거리 영상인데, 근거리 영상의 경우 특정 부위나 물체가 확대되어진 영상을 의미한다. 사람의 얼굴이나, 손 또는 특정한 물체가 확대되어 한 프레임에 담겨진 영상을 의미한다. 원거리 영상의 사람은 매우 작아 사람의 전신을 하나의 작은 블랍(blob)으로 다루게 된다^[1]. 근거리 영상은 특정 객체를 확대한 영상으로써 관심 있는 객체를 식별하기 위해 사용된다. 이와 같은 영상들은 사람의 관절이 너무 크거나 혹은 영상에 사람의 전신이 다 나타나지 않아 사람의 관절을 추적하기에 부적합하다^[1]. 그러므로 다관절체 사람을 추적하기 위한 영상은 중간 거리 영상을 사용한다. 중간 거리 영상은 사람의 팔과 다리, 몸통의 식별이 가능한

영상이다. 또한 중간거리 영상에서 사람이 차지하는 영역의 크기는 전체 영상 크기의 50% 미만이다. 그리고 추적되는 사람은 객체로서 카메라로부터 가까이 서 있게 된다. 그러므로 영상의 평균 디스패리티보다 디스패리티가 작은 곳은 배경으로 간주 할 수 있다. 즉, 평균 디스패리티보다 작은 곳에서는 메시지를 생성할 필요가 없게 된다. 왜냐하면 사람은 중간거리 영상에서 50%미만의 영역을 차지하고 있으며 카메라와 가까이 위치하고 있어 평균 디스패리티보다 큰 디스패리티를 가지기 때문이다. 따라서 식(13)과 같이 메시지를 조건이 만족할 경우에만 생성하도록 할 수 있다. 이와 같은 메시지 생성 방법을 조건적 메시지 업데이트라 한다.

$$M_j^k = \begin{cases} \min_{X_k} (\Psi_{j,k}(X_j, X_k) + \Psi_k^t(X_k, Y_k) + \Psi_k^t(X_k^t, X_k^{t-1}) + \sum_{l \neq j} \tilde{M}_k^l) & , \text{if } d(X_k) \geq d_m \\ & \text{and } d(X_j) \geq d_m \\ \infty & , \text{if } d(X_k) < d_m \\ & \text{and } d(X_j) < d_m \end{cases} \quad (13)$$

d_m 은 영상의 평균 디스패리티를 의미한다. $d(\cdot)$ 은 (\cdot) 위치에서의 디스패리티이다. 메시지 업데이트는 신뢰 전파 알고리즘 연산에서 80% 이상을 차지한다^[8]. 따라서 식(13)과 같이 X_i 위치의 디스패리티를 이용해 객체가 존재 할지 여부에 대한 타당성을 검토한 후 메시지를 업데이트하도록 하면 계산 부담을 덜 수 있다. IV장의 실험 결과를 참고하면 기존 메시지 업데이트 방법에 비해 조건적 메시지 업데이트 방법이 9.21~45.77%의 속도 향상률을 보이고 있다. 향상률에 차이가 나는 이유는 영상의 특성에 기인하는데, 영상의 평균 디스패리티가 높으면서 사람이 카메라로부터 멀리 떨어져 있고 사람 외의 다른 물체가 카메라로부터 가까이 존재하지 않는 것이 많을수록 속도가 향상된다.

IV. 실험

본 장에서는 3장에서 제안한 알고리즘을 여러 가지 환경에서 테스트하여 그 성능을 측정하고 관찰 하도록 한다. 실험은 Intel(R) Core^(TM)2Duo 2.8Ghz 시스템에서 실험하였다. 실험에 쓰인 영상은 320 x 240의 해상도를 가지는 중간 거리 영상을 사용하였다. 본 실험에 사용된 세 가지 영상의 대표 프레임을 그림 5에 표시하였다.



그림 4. 영상의 세 가지 분류, 좌측부터 원거리 영상, 중간 거리 영상, 근거리 영상

Fig. 4. Three types of image. Far view image, middle view image, and closed range image.



그림 5. 실험에 사용된 세 가지 영상, 좌측 A영상, 중앙 B영상, 우측 C영상
Fig. 5. Three types of videos used in the experiment.

1. 결과 및 기존 방법과의 비교

본 연구에서는 디스패리티 지도를 이용하여 다관절체를 효과적으로 찾는 방법에 대해 제안하였다. 즉, 우도 계산을 색상 정보와 함께 신체 요소 디스패리티 분포의 연속적인 특성을 사용하여 기존에 비해 강건하게 하고, 3차원 연결 제약 및 3차원 다이내믹 모델을 이용하여 각 신체 요소들이 사람같이 연결 될 수 있도록 하였다.

그림 6은 제안한 방법을 사용해 추적한 결과이다. 제안한 방법과 기존의 방법을 비교 실험을 해보았다. 비교한 기존의 방법은 두 가지로써 Tony X. Han^[2]의 방법은 우도를 계산할 때 템플릿과 영상과의 제곱차합 연산으로만 계산하여 신체요소와 비슷한 색상이 존재하는 배경과 같은 상황이나 조명에 민감한 추적 결과를 보여 주고 있다. P. Felzenszwalb^{[3][5]}는 각 객체들이 위치 할 수 있는 공간을 이산화 시켜 연속적인 사람의 움직임을 추적할 때 실패할 확률이 높았다. 또한 배경차영상을 이용해 우도를 계산하기 때문에 움직이는 배경 또는 카메라가 움직이는 환경에 적용할 수 없는 단점을 가지고 있다.

표 1은 A 영상의 총 102 프레임에서의 추적 결과이다. 정확도는 (추적에 성공한 신체요소의 수)/(총 추적을 시도한 신체요소 수)를 퍼센트 단위로 변환하여 구했다. Tony X. Han의 경우 팔보다 다리를 잘 찾았는데, 그 이유는 우도를 RGB공간에서 제곱차합만으로 구하는데 A 영상의 조명의 변화가 팔보다 다리에서 적었기



그림 6. 제안한 방법을 사용해 추적한 결과
Fig. 6. result of proposed method.

표 1. A 영상에서 다른 방법과의 비교
Table 1. Compared result from video A.

	몸통	팔	다리	전체
T X. Han ^[2]	89.3%	58.1%	78.9%	70.8%
Pedro F. ^{[3][5]}	100%	85.9%	42.7%	68.4%
제안 방법	100%	92.4%	90.4%	92.4%

표 2. B 영상에서 다른 방법과의 비교
Table 2. Compared result from video B.

	몸통	팔	다리	전체
T X. Han ^[2]	97.1%	82.9%	89.8%	87.5%
Pedro F. ^{[3][5]}	99.0%	77.2%	75.0%	75.7%
제안 방법	98.1%	86.2%	95.6%	91.6%

때문이다. P. Felzenszwalb의 방법은 전체 68.4%로 가장 낮는데 각 축 별로 이산화된 상태공간을 가지고 있기 때문에 각 신체요소의 움직임에 민감하기 때문이다.

표 2는 B영상의 총 103프레임에서의 추적 결과이다. P. Felzenszwalb가 75.7 %로 가장 낮은 정확도를 보이고 있고 Tony X. Han이 87.54 %로 두 번째로 괜찮은 추적 결과를 보이고 있다.

실험 결과를 통해 제안한 방법의 성능이 가장 우수함을 볼 수 있다. 색상 정보와 함께 디스패리티 정보를 사용한 우도와 3차원 연결 제약 및 3차원 다이내믹 모델로써 각 신체 요소가 3차원 공간상에서 사람같이 연결

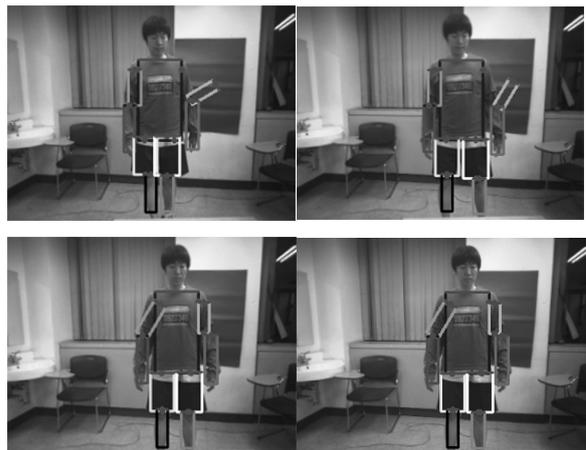


그림 7. 사람과 비슷한 배경을 가진 영상에서의 실험 결과, 윗줄은 T X. Han^[2]의 방법을 사용한 추적한 결과, 아랫줄은 제안한 방법에 의한 추적 결과.

Fig. 7. top row shows the results of T X. Han^[2] which uses color information and bottom row shows the results of the proposed method.

되도록 유도하였기 때문이다.

그림 7은 배경에 신체요소와 비슷한 색상이 있을 때 색상 정보만 사용했을 때의 추적 성능과 디스패리티 정보를 같이 사용했을 때의 추적 성능을 비교하기 위한 예로써, 그림 7을 보면 아랫줄의 제안하는 방법이 윗줄의 기존의 방법^[2]보다 강건한 추적 결과를 보임을 확인할 수 있다.

그림 7의 윗줄의 그림에서 왼쪽 위팔이 잘 못 추적된 것을 볼 수 있다. 색상과 함께 디스패리티 정보를 사용한 우도와 3차원 연결 제약이 기존의 방법^[2]보다 신체요소 끼리 잘 연결 될 수 있도록 우도를 계산하였기 때문이다.

2. 속도 향상에 관한 실험

본 논문에서 사용한 속도 향상 방법은 두 가지 이다. 첫째는 조건적 메시지 업데이트이고 둘째는 다이나믹 모델이다. 메시지 업데이트는 신뢰 전파 연산의 80%를 차지한다^[8]. 식 (13)의 조건적 메시지 업데이트 방법은 신체 요소 위치의 디스패리티 정보가 조건을 만족 할 때에만 메시지를 생성하도록 하여 메시지 업데이트에 따른 연산 부담을 줄이는 방법이다. 표 3은 제안하는 조건적 메시지 업데이트 방법을 통한 속도 향상 정도를 나타내고 있다. 일반 메시지는 식(2)를 사용하여 메시지 업데이트를 하는 것을 말한다. 향상률은 (일반 메시지-

표 3. 조건적 메시지 생성을 통한 속도 향상 정도
Table 3. Improvement of computation speed with conditional message update.

	A영상	B영상	C영상
일반 메시지 (sec/frame)	2.8118	3.9864	5.6584
조건적 메시지 (sec/frame)	2.1798	2.1618	5.1370
향상률	22.47%	45.77%	9.21%

표 4. 다이나믹 모델을 통한 연산 속도 향상 정도
Table 4. Improvement of computation speed with dynamic model.

	A영상	B영상	C영상
최초 추적 (sec/frame)	2.1798	2.1618	5.1370
다이나믹 모델 (sec/frame)	0.8042	0.7130	2.5178
향상률	69.1%	67%	51%

조건적 메시지)/(일반 메시지)×100으로 계산 하였다.

신뢰 전파 연산의 80%를 차지하고 있는 메시지 업데이트를 조건적으로 하였음에도 불구하고 그에 상응하는 향상률을 보이지 못한 이유는 본 논문의 전처리 과정인 픽셀 분류 과정을 통해 이미 많은 상태 공간이 축소되어졌기 때문이다. 또 영상에 따라 향상률이 차이가 있는데 조건적 메시지 생성 방법이 영상의 디스패리티 분포 특성에 기인하기 때문이다. 영상의 평균 디스패리티가 높으면서 사람이 카메라로부터 멀리 떨어져 있고 사람 외의 다른 물체가 카메라로부터 가까이 존재 하지 않는 것이 많을수록 속도 향상률이 높아지게 된다.

다이나믹 모델은 사람의 움직임 특성을 이용한 모델로써 사람의 신체 요소는 움직임에 연속성을 가지고 있기 때문에 t시간의 신체 요소의 위치는 바로 이전 시간인 t-1시간 위치의 이웃에 존재하게 되는 성질을 이용한다. 표 4는 다이나믹 모델을 사용하였을 때 속도의 향상 정도를 나타낸 것이다. 최초 객체의 위치를 파악 한 후 다음 프레임부터 다이나믹 모델을 적용하므로 최초 추적을 할 때에는 다이나믹 모델을 적용하지 않는다. 표 4에서의 향상률은 (최초 추적-다이나믹 모델)/(최초 추적)×100으로 계산 하였다.

V. 결 론

본 논문에서는 신뢰 전파(Belief Propagation)를 디스패리티 지도(disparity map)와 함께 사용하여 다관절체 사람을 강건히 찾는 방법을 제안 하였다. 실세계의 3차원의 신체 요소가 2차원 영상에 투영될 때 그 디스패리티 분포는 연속적이거나 일정한 기울기를 가지게 된다. 신체 요소의 디스패리티의 기울기 값을 우도에 반영하여 불연속이 존재하는 위치나 혼란을 야기하는 배경이 있는 위치 등에서 효과적으로 추적할 수 있었다. 또한 조건적 메시지 생성과 3차원 다이나믹 모델을 사용하여 기존 보다 빠르게 추적 하는 방법과 3차원 연결 제약 및 3차원 다이나믹 모델(Dynamic model)을 이용하여 각 신체 요소가 강건히 연결되어 추적할 수 있음을 나타내었다.

디스패리티 정보를 이용한 우도는 기존의 방법보다 전체적으로 4.1%~24% 정확도 향상을 보였으며, 조건적 메시지 업데이트는 기존 보다 9~45% 정도의 향상된 연산 속도를 보였으며 3차원 다이나믹 모델은 기존 보다 50~70%의 연산 속도 향상을 보였다.

추후에는 디스패리티 정보를 활용해 사람의 3차원 위치를 추적하는 방법에 대한 연구를 진행할 계획이다.

참 고 문 헌

[1] D. A. Forsyth, O. Arikan, L. Ikemoto, J. O'Brien, D. Ramanan. "Computational Studies of Human Motion: Part 1, Tracking and Motion Synthesis" Foundations and Trends in Computer Graphics and Vision, Hanover, Massachusetts., 2006.

[2] Tony X. Han and Thomas S. Huang, "Articulated Body Tracking Using Dynamic Belief Propagation," IEEE International Workshop on Human-Computer Interaction, in conjunction with ICCV (ICCV-HCI 2005), pp. 26-35., 2005

[3] P. Felzenszwalb, D. Huttenlocher, "Pictorial Structures for Object Recognition", International Journal of Computer Vision, Vol. 61, No. 1, Jan 2005.

[4] Ramanan, D., Forsyth, D. A., Zisserman, A., "Tracking People by Learning their Appearance," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 29, pp. 65-81, Jan 2007.

[5] P. Felzenszwalb, D. Huttenlocher "Efficient Matching of Pictorial Structures" IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2000.

[6] J.S. Yedidia, W.T. Freeman, and Y. Weiss, "Understanding Belief Propagation and Its Generalizations," Proc. Int'l Joints Conf. Artificial Intelligence, Aug. 2001.

[7] Nebojsa Jovic, Matthew Turk, Thomas S. Huang, "Tracking Self-Occluding Articulated Objects in Dense Disparity Maps," International Conference on Computer Vision (ICCV), pp. 123-130., Korfu, Greece, Sep 1999,

[8] W. Freeman, E. Pasztor, and O. Carmichael, "Learning low-level vision," IJCV, vol. 40, pp. 25-47, 2000.

[9] O. Bernier, P. Cheung-Mon-Chan, and A. Bouguet "Fast Nonparametric Belief Propagation for Real-Time Stereo Articulated Body Tracking", Computer Vision and Image Understanding, Vol. 113, no. 1, pp. 29-47, Jan 2009.

[10] YoungJoon Chai, JinYong Park, Kwang-Jin Yoon, and TaeYong Kim, "Multi Target Tracking Using Multiple Independent Particle Filters For Video Surveillance", IEEE

Transaction on Consumer Electronics, Las Vegas, US., 2010.

[11] Kwang-Jin Yoon and Tae-Yong Kim, "Articulated Body Tracking Using Belief Propagation with Disparity Map", ISCE, Singapore, 2011

[12] Jones, M.J, Rehg, J.M, "Statistical color models with application to skin detection", Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.

————— 저 자 소 개 —————



윤 광 진(정회원)
 2009년 삼육대학교 컴퓨터과학과 학사 졸업.
 2011년 중앙대학교 첨단영상대학원 영상학과 석사 졸업.
 <주관심분야 : 컴퓨터 비전, 트래킹, 객체 인식>



김 태 용(평생회원)
 1986년 한양대학교 전기공학과 학사 졸업.
 1988년 한양대학교 전자통신공학과 석사 졸업.
 1998년 포항공과대학교 전자계산학과 박사 졸업.

2003년~현재 중앙대학교 첨단영상대학원 영상학과 교수
 <주관심분야 : 컴퓨터 비전, 영상처리, 컴퓨터 게임>