

로봇-휴먼 인터랙션을 위한 인간 몸의 실루엣 및 액티브 스킴레톤 추출

소제운, 김진규, 주영훈, 박진배*
 군산대학교 전자정보공학부, * 연세대학교 전기전자공학과

Silhouette and Active Skeleton Extraction of Human Body for Robot-Human Interaction

Jeon Yun So, Jin Gyu Kim, Young Hoon Joo, Jin Bae Park*
 School of Electronic & information Engineering, Kunsan National University
 * Electrical and Electronic Engineering, Yonsei University

Abstract - 본 논문에서는 로봇과 인간의 인터랙션을 위해 인간 몸의 실루엣 및 액티브 스킴레톤 추출 기법을 제안한다. 연속된 이미지 정보로부터 얻어진 옷영역등의 정보에서 background subtraction를 이용한 adaptive fusion을 통해 추출된 인간 몸의 실루엣을 바탕으로 active contour와 가상 신체 모델인 skeleton model을 응용하여 작은 움직임에 보다 강한 active skeleton model을 이용하여 인간 몸의 특징 점 위치를 추출하는 방법을 한다.

1. 서론

오늘날 현대사회에서는 인간-컴퓨터간, 인간-로봇간, 컴퓨터-로봇간 등 피드백이 발생하는 모든 요인을 고려한 인간 공학적이고 형태학적인 분석이 필요하다. 그 중에서 인간-로봇 상호 작용[4]은 인간의 시각 기능을 기반으로 하여 로봇이 인간의 행위 및 의사 정보를 알 수 있고, 또한 인간이 로봇의 의사를 알 수 있도록 전달하는 인간-로봇 상호 작용기술로서 제스처 인식이 연구 되었으며, 사람과 복잡한 로봇, 그리고 컴퓨터 작업환경에서 효과적인 작업을 위해서, 인간과 로봇, 컴퓨터간의 인터페이스 분석이 이루어져왔다. 인간의 동작 추출은 연속된 이미지로부터 움직임 영역의 정확한 특징점을 찾는 데 있다. 본 논문에서는 연속된 이미지로부터 피부색, 옷영역, 시·공간차 정보등을 혼합하여 인간 몸의 실루엣을 추출하고, OpenCV에서 제공된 얼굴 탐색기법[10]과 peer의 고정 영역 피부색 추출 기법[5], CHB 기법[7]을 이용하여 연속된 영상 정보로부터 얻어진 신체정보를 기반으로 스킴레톤 특징점 추출을 한다. 기존의 인간의 움직임을 중요한 관례의 위치만으로 표현한 skeleton 특징점 표현 형태와 active contour 기법을 응용한 active skeleton 모델 기법을 제안한다. 또한 원형 탐색 알고리즘을 이용하여 팔꿈치 영역 특징점을 추출하여 특징점 추출을 위한 최소한의 탐색 공간을 탐색함으로써 탐색시간을 줄이고 특징점 추출 정확도를 높인다.

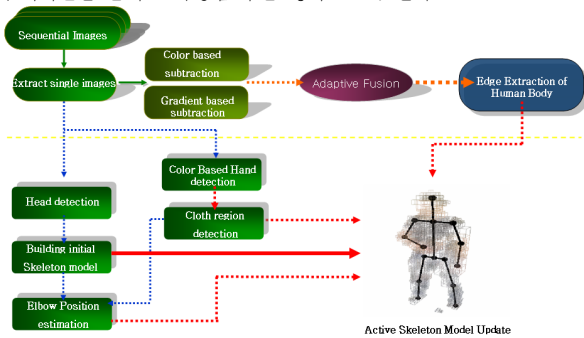


그림 1. 인간의 실루엣과 active skeleton 추출 흐름도

2. 실루엣 및 액티브 스킴레톤 추출

2.1 Silhouette extraction using Background subtraction

본 논문에서 제안한 실루엣 특징점 추출 기법은 Stauffer와 Grimson [1][2][3]가 사용한 background subtraction 기법을 통해 특징점 추출에 필요한 다양한 영상 정보를 추출한다. 연속된 이미지로부터 color, edge, gradient 정보를 이용하여 인간 몸의 실루엣을 추출한다. 칼라 기반의 움직임 추출과 인간의 움직임 추출에서 가장 기본적인 두 가지 정보인 영상의 spatial gradient(SG) 정보와 영상간의 temporal gradient(TG) 정보[7][8]를 이용하여 인간 몸의 실루엣을 추출한다. 칼라 정보는 조명이 민감하고 SG는 영상에 나타는 모든 edge정보를 가지지만 인간 몸만의 edge 정보 구분이 어려운 반면, TG는 영상에서 인간이 움직임을 있을 경우 인간의 고유의 edge 정보를 가지지만 움직이지 않는 영역에 대해 부분적인 손실이 있어 전체적인 인간 몸의 edge를 가지지 못한다. 따라서 세 가지 정보를 바탕으로 adaptive fusion을 통한 장점 최대화 기법으로 인간 몸의 실루엣을 추출 한다.

2.1.1 Color based subtraction

칼라 영역에서 배경제거를 위해 Stauffer와 Grimson의 칼라 영역의 mixture of K Gaussians 기법을 사용한다. 시간 t에 대해 픽셀의 위치가 (i,j)일 때 K Gaussians 확률밀도함수는 다음 식 (1)과 같다.

$$n(I_{i,j}^t, \mu_{i,j}^t, \Sigma_{i,j}^t) = \frac{1}{(2\pi)^{\frac{n}{2}} \left| \Sigma_{i,j}^t \right|^{\frac{1}{2}}} \exp - \frac{1}{2} (I_{i,j}^t - \mu_{i,j}^t)^T \Sigma_{i,j}^{-1} (I_{i,j}^t - \mu_{i,j}^t) \quad (1)$$

2.1.2 Gradient based subtraction

본 논문에서 사용된 연속적인 영상 정보 시간 t에 대해서 다음 식(2)와 같이 정의 된다.

$$I(x, y, ; t) = [I_r(x, y, t), I_g(x, y, t), I_b(x, y, t)]^T \quad (2)$$

TG는 연속적인 두 이미지 사이의 temporal difference 이다. Temporal gradient와 Spatial gradient edge 정보는 다음 식(3)과 같이 정의된다.

$$I_t = \frac{\partial I}{\partial t} \quad I_s = \left[\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right] \quad (3)$$

위와 같이 윤곽선 정보와 Gradient 정보는 형태가 다르기 때문에 다음 식(4)과 같은 Mapping 방법을 사용하여 변경한다.

$$\overline{I}_t = \| I_t \|, \quad \overline{I}_s = \| I_s \| \quad (4)$$

여기서 배경 제거는 영상의 모든 edge정보를 갖는 SG 만을 적용하여 얻어진 인간 몸의 실루엣 정보에 TG 정보를 다음 식(5)의 convex summation을 이용하여 gradient 기반의 실루엣을 추출한다. 최종적으로 color 기반 정보와 gradient 기반 정보를 이용하여 adaptive fusion을 통해 인간 몸의 실루엣을 추출한다.

$$n \overline{I}_s(x, y) + (1 - n) \overline{I}_t(x, y) \quad (5)$$

2.2 Adaptive active skeleton model method

인간의 움직임 영역을 추출하기 위해서는 움직임 영역 탐색 시 탐색 공간을 줄여줄 수 있는 인간의 신체구조적 정보를 이용하는 것이 효율적이다. 인간의 신체구조적 정보로 사용되는 Skeleton model 형태는 아래 Figure 2과 같다. skeleton model은 움직임을 표현하기 가장 간단한 형태의 모델로 인간 몸의 움직임을 표현하는데 매우 용이하다. skeleton model은 총 15개의 특징점을 가지며, 14개 뼈대를 가진다. OpenCV에서 제공하는 얼굴탐색기법을 이용하여 얼굴의 위치와 가로 길이를 얻고 미리 계산된 신체 구조비를 이용하여 skeleton model의 생성하는 방법을 사용하였다. 하지만 사람마다 신체비가 다르기 때문에 얼굴을 기준으로 한 특징 벡터가 잘못 생성될 수가 있어 active contour 기법을 응용한 adaptive active skeleton model을 제안한다.

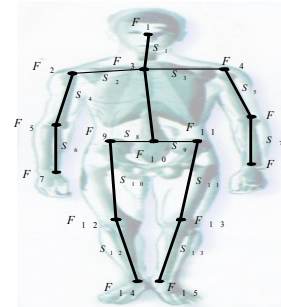


그림 2. skeleton model

각각의 신체구조비를 통해 얻어진 초기 점을 기준으로 에너지 함수의 값을 옥십쟁이 기법을 사용하여 최소로 하는 최적의 특징점을 찾아 active skeleton model[8][9]를 생성한다. active skeleton model에서 각 특징점의 거리는 S_1, K, S_{13} 로 표현된다.

Active skeleton model을 이용하여 동정되는 특징점은 몸통 안에 있는 특징점 ($F_1, F_2, F_3, F_4, F_9, F_{10}, F_{11}$)이다. 몸통 밖의 특징점 손, 팔꿈치는 움직임이 크기 때문에 동정하기가 힘들다. 따라서 peer의 색상 모델을 이용하여 피부색을 탐지하고 mean shift 탐색 기법을 사용하여 손영역을 추출 하였으며, 원형 탐색 기법을 이용하여 팔꿈치 영역의 특징점을 추출한다. 특징점 동정에 사용된 에너지 함수는 내부, 외부 에너지 함수로 구성된다. 이 알고리즘은 snake 알고리즘과 유사하며 에너지 함수를 최소화 시키는 점을 찾는 방향으로 알고리즘을 수행한다. 내부에너지 함수는 뼈대 정보와 특징점의 기하학적 관계를 바탕으로 계산되며 외부에너지 함수는 실루엣 이미지 픽셀 정보를 이용하여 정의된다. 내부에너지 함수와 외부에너지 함수는 다음 식(6)과 같다

$$E_i^{int} = |S_i - \|f_j - f_k\||$$

$$E_i^{ext} = - \frac{f_{ix-k+m} f_{iy-k+m}}{\sum_{x=f_{ix-k}} \sum_{y=f_{iy-k}} T_b(x, y)} \quad (6)$$

최종적으로 에너지 함수 E_i 는 내·외부에너지 함수 합으로 표현된다.

$$E_i = \alpha E_i^{ext} + (1 - \alpha) E_i^{int} \quad (7)$$

초기 픽셀이 주어졌을 때 인접한 모든 픽셀의 에너지 값을 계산하고 에너지 값이 작은 쪽을 새로운 특징점으로 잡는 방법을 통해 특징점을 찾는다. k 는 특징점과 실루엣 사이의 거리를 나타내며 m 은 탐색공간을 α 는 불록합 파라미터 값이다.

2.2.1 Color based hand extraction method

손과 같이 움직임 정보가 많은 곳은 실루엣 정보가 정확하지 않기 때문에 peer의 색상 모델 기법[5]을 이용하여 피부색을 탐지 한다. 두 손이 겹치지 않는 경우와 얼굴 위치에 겹치지 않는다는 가정 하에 몸 중앙을 중심으로 가로, 세로 방향의 히스토그램을 만들고 mean shift 탐색 기법을 사용하여 피부색 정보가 많은 부분을 손이라 판단한다 고정영역 조건식은 식(8)과 같다.

$$(I_r > 95) \cap (I_g > 40) \cap (I_b > 20) \cap (\max(I_r, I_g, I_b) - \min(I_r, I_g, I_b) > 15) \cap (I_r > I_g) \cap (I_r > I_b) \cap (I_r > I_b) \quad (8)$$

2.2.2 Robust skeleton extraction method

색상 기반 손 탐색은 색상과 움직임 정보만을 사용하기 때문에 근처에 다른 사람의 손이 오는 경우는 구분하지 못한다. 따라서 본 논문에서는 어깨선 추출과 옷영역 추출을 통해 보다 정확한 손 위치를 파악한다.

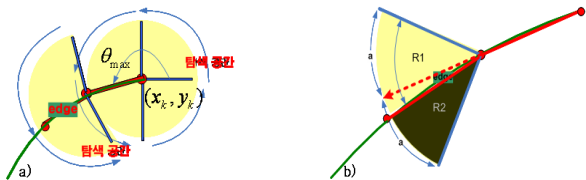


그림 3. (a) Shoulder line search space, (b) Energy region
Figure 3과 같이 원형 탐색 기법을 사용하여 원형의 탐색 공간을 각 원형으로 돌면서 탐색을 하는 방법으로 어깨선을 추출한다. 탐색 영역안의 각 각들은 R1과 R2 영역을 가지며 식(9)의 에너지함수에 의해 에너지 값이 가장 큰 각으로 탐색 하는 방법으로 어깨선을 추출한다. 옷 영역은 color histogram backprojection 기법을 이용하여 추출하였다.

$$S = (\sum_{\forall x \in R2} f(x) - \sum_{\forall x \in R1} f(x)) / r + \sum_{\forall x \in L} f(x) \quad (9)$$

3. 실험 및 결과 고찰

본 논문에서 사용된 입력 영상은 320 × 240 24bit RGB이며 프레임 속도는 10 frame/sec 이상에서 실험하였고, 사용된 실험영상은 1인 영상 11개, 2인 이상 영상 4개를 사용하여 실험하였다. 그림 4는 Background subtraction 기법을 이용한 인간 몸의 실루엣 추출 생성 실험 결과를 나타낸다. 그림 5는 그림 4에서의 영상정보로부터 얻어진 특징점을 이용하여 adaptive active skeleton 생성 실험 결과를 보여준다.

표 1 어깨 추출 실험 결과

구분	프레임 수	얼굴추출 프레임 수	손 추출 프레임 수	어깨추출 프레임 수	추출 성공률
1인 영상	2310	1820	1580	1210	76.6%
2인 영상	840	650	520	380	73.1%
총 어깨선 추출 성공률					74.9%

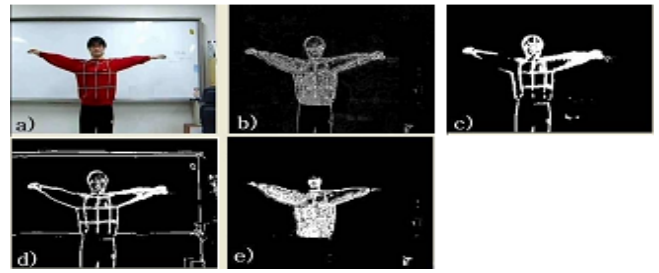


그림 4. (a) Original image, (b) Result image, (c) Temporal gradient, (d) Spatial gradient, (e) Color based edge image

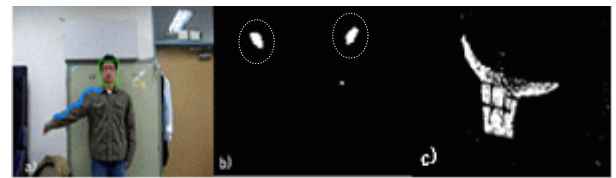


그림 5.추출실험결과 (a) Shoulder line and head, (b) Hand detection, (c) Cloth region

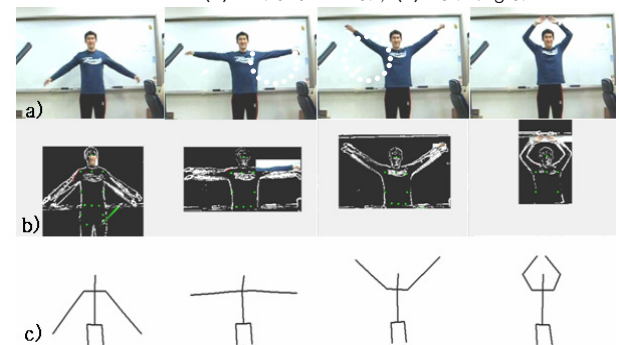


그림 6. 실험결과 (a)original image, (b)feature of skeleton, (c) active skeleton model

4. 결론

본 논문에서는 로봇과 인간의 인터랙션을 위해 인간 몸의 실루엣 및 액티브 스키텔론 추출 기법을 제안하였다. 또한, 연속된 이미지 정보로부터 얻어진 피부색, 옷영역등의 정보에서 background subtraction를 이용한 adaptive fusion을 통해 추출된 인간 몸의 실루엣을 바탕으로 active contour와 가상 신체 모델인 skeleton model을 응용하여 작은 움직임에 보다 강한 active skeleton model을 이용하여 인간 몸의 특징 점 위치를 추출하는 방법을 제안하였다. 제안한 방법을 실험을 통해 그 응용 가능성을 보였다.

참고문헌

- [1] C.Stauffer and W. E. L. Grimson: Adaptive background mixture models for real-time tracking in: Computer Vision and Pattern Recognition Fort Collins, Colorado(Jun. 1999)
- [2] I. Haritaoglu, R. Cutler, D. Hawood and L. Davis, "Backpack:detection of people carrying objects using silhouette," Computer Vision and Image Understanding, 2001.
- [3] Y. Ivanov, C. Stauffer, A. Bobick and W. E. L. Grimson: Video surveillance of interactions in: Second IEEE Workshop on Visual Surveillance Fort Collins, Colorado(Jun. 1999).
- [4] J. K. Aggarwal and Q. Cai., "Human motion analysis: a review" Computer Vision and Image Understanding, 1999.
- [5] P. Peer, J. Kovac, and F. Solia, "Human skin colour clustering for face detection," EUROCON, 2003.
- [6] G. Healey and D. Slater, "Using illumination invariant color histogram descriptors for recognition", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1994.
- [7] Moon Hwan Kim, "Extraction of Features in Human Body Using Adaptive Skeleton Model", Korea ICASE. 2005.
- [8] Moon Hwan Kim, "Extraction of Human Body Using Hybrid silhouette extraction method in intelligent robot system", Korea Fuzzy Logic and Intelligent System Society, 2005
- [9] G. Halevy and D. Weinshall: Motion of disturbances:detection and tracking of multi-body non-rigid motion Machine Vision and Applications(1999)
- [10] <http://opencvlibrary.sourceforge.net>