

# 인간 로봇 상호작용을 위한 강인한 스켈레톤 특징점 추출기 설계

## Design of Robust Skeleton Feature Extractor for Human-Robot-Interaction

\*김문환, \*\*주영훈, \*\*\*박진배

\*연세대학교 전기전자공학과, \*\*군산대학교 전자정보공학부

M. H. Kim, Y. H. Joo, J. B. Park

\*Department of Electrical Engineering, ROK Naval Academy

\*\*School of Electronic and Information Engineering, Kunsan National University

\*\*\*Department of Electrical and Electronic Engineering, Yonsei University

### 요 약

본 논문에서는 인간 로봇 상호 작용을 위해 정확한 스켈레톤 특징점을 추출하는 강인한 추출기를 설계한다. 제안된 특징점 추출기는 인간의 움직임 정보로부터 얻어진 색상, 윤곽선, 시간차 정보 및 가상 신체 모델을 이용하여 정확한 특징점 위치를 찾아낸다. 또한 특징점 추출에 소요되는 탐색 시간을 줄이기 위해 격자박스를 이용한 원형 탐색 기법을 도입하였다. 최종적으로 기법의 우수성을 확인하기 위해 다양한 동작의 특징점 추출 실험을 수행하였다.

### 1. 서론

최근 10년간 휴먼노이드 로봇 기술은 눈부신 발전을 이루고 있다. 특히 단순 보행 기능뿐만 아니라 인간과 상호 작용을 하는 기술 개발이 활성화 되고 있다 [1]-[8]. 로봇 인간 상호작용 기법은 인간의 의도를 로봇이 알 수 있고 또한 로봇의 의사를 인간이 알 수 있도록 하는 기술로서 크게 접촉식과 비접촉식으로 나눌 수 있다. 접촉식 상호작용 기법은 기존의 여타 상호작용 방법과 같이 버튼 혹은 별도의 입력 장치의 조작을 통해 로봇에게 의사를 전달하는 방법이다. 이는 정확하며 복잡한 형태의 의사 전달이 가능한 장점이 있는 반면 입력 장치의 조작이 필요하다는 단점을 가지고 있다. 비접촉식 상호작용 기법은 사용자의 영상이나 음성을 통해 얻어진 정보를 바탕으로 간접적으로 인간의 의사를 판단하고 이를 이용하여 상호작용에 이용하는 방법이다. 이는 사용자의 별도의 조작 없이 자연스러운 의사 표현으로 로봇과 상호작용을 할 수 있다는 장점을 가진 반면 높은 성능을 가지는 상호작용 기법을 개발하기 힘들다는 단점을 가진다.

인간의 동작을 인식에 앞서 가장 중요한 것은 동작을 가장 잘 표현하는 정확한 특징점을 찾는

데 있다. 이러한 특징점은 원통형, 막대형, 타원형 등등 여러 형태로 표현이 가능하다. 본 논문에서는 스켈레톤 형태의 막대형 특징점 추출 방식을 사용한다. 스켈레톤 형태는 인간의 움직임을 중요한 관절의 위치만으로 표현한 특징점 표현 방식이다.

기존의 스켈레톤 특징점 추출은 배경 제거 기법 혹은 색상 추출 기법 등을 이용하여 수행되어 왔다. 하지만 배경 제거 기법과 색상 추출 기법이 그 성능이 우수함에도 불구하고 극히 제한된 환경에서만 동작이 가능하다는 단점을 가진다. 가정용 로봇과 같은 움직이면서 다양한 조도와 배경을 가진 환경에서는 그 성능이 매우 낮아 실제적으로 사용하기가 힘들다.

본 논문에서는 이러한 기존의 기법의 한계점을 극복하고 가정용 로봇에 실적용 가능한 강인한 스켈레톤 추출 기법을 제안한다. 제안된 기법은 스켈레톤 특징점 추출을 위해 움직임 정보, 피부색 정보, 옷 색 정보 그리고 윤곽선 정보를 혼합하여 이용한다. 또한 정밀하고 빠른 특징점 추출을 위해 신체정보를 기반으로 한 원형 탐색 기법을 제안한다. 원형 탐색 기법은 특징점이 존재할 수 있는 최소한의 탐색 공간을 탐색함으로써 탐색 시간을 크게 줄이고 성능을 높였다.

## 2. 스켈레톤 특징점

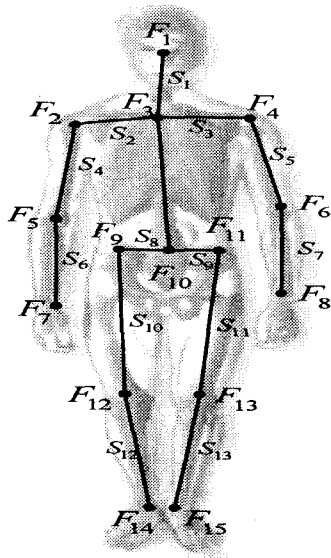


그림 1 스켈레톤 특징점

스켈레톤 특징점은 정확한 동작은 작은 정보를 통해서 표현하는 효과적인 특징점 표현 기법이다. 본 논문에서 사용된 특징점은 그림 1 과 같다. 전체 스켈레톤의 특징점은 15개이며 14개의 막대로 이루어져 있다. 그림 1에서  $F_1, \dots, F_{15}$ 은 특징점을  $S_1, \dots, S_{14}$ 는 막대를 나타낸다. 본 논문에서는 사용자의 상호작용에 많은 사용되는 상반신 특징점 8개만을 사용한다. 사용된 특징점은  $F_1, \dots, F_8$  이다. 특징점 중 머리에 해당하는  $F_1$ 은 OpenCV에서 제공된 얼굴 탐색기법에 의해서 탐색이 되며 나머지 특징점은 제안된 기법에 의해서 추출이 된다 [10].

## 3. 강인한 특징점 추출 기법

본 논문에서 제안한 강인한 특징점 추출 기법은 크게 세단계로 나누어진다. 가장 먼저 영상 전처리 과정을 통해 특징점 추출에 필요한 다양한 영상 정보를 추출한다. 그리고 OpenCV에서 제공하는 얼굴 탐색기법을 이용하여 얼굴의 위치를 탐색한다. 최종적으로 얻어진 얼굴의 위치를 기준으로 가상의 인간 모델에서 얻어진 비례 정보를 이용하여 탐색 공간을 설정하고 팔꿈치, 손순서로 특징점을 추출한다. 그림 2는 특징점 추출 기법의 순서를 나타낸다.

### 3.1 특징점 탐색을 위한 영상 정보 추출

본 논문에서 사용되는 영상정보는 움직임 영상 정보, 윤곽선 정보, 옷색 정보, 피부색 정보 총 4

가지이다. 각각의 영상 정보는 영상 전처리 과정

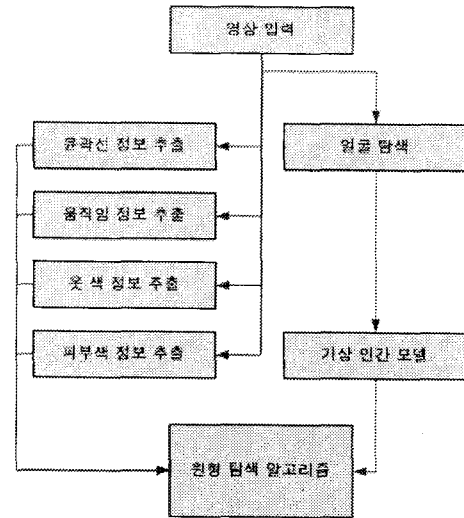


그림 2 특징점 추출 기법 순서도

에서 추출이 되며 매 특징점 추출시 다시 계산이 된다. 시간  $t$ 에 주어진 영상 정보는  $I(x, y, t)$  다음과 같이 정의 된다.

$$I(x, y, t) = [I_r(x, y, t), I_g(x, y, t), I_b(x, y, t)]^T \quad (1)$$

이때  $I_r, I_g, I_b$ 는 각각 RGB 영상 정보를 의미한다. 동작 정보  $I_t$  및 윤곽선 정보  $I_s$ 는 다음과 같이 계산된다.

$$I_t = \frac{\partial I}{\partial t} \quad (2)$$

$$I_s = \left[ \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right]$$

피부색 정보  $I_p$ 는 peer의 고정 영역 피부색 추출 기법을 사용하였다 [9]. 사용된 고정 영역의 조건은 다음과 같다.

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

옷색 정보  $I_c$ 는 color histogram backprojection (CHB) 기법을 이용하여 추출을 하였다 [11]. CHB를 사용하기 위해 가슴 영역에서 얻어진 옷색을 이용하였다.

### 3.2 가상 인간 모델을 이용한 탐색 공간 설정

OpenCV의 얼굴 탐색 기법에 의해 얼굴 영역이 탐색 되면 가상 인간 모델을 바탕으로 인간의 움직임 가능 영역이 도출된다. 가상 인간 모델은 인간의 얼굴 영역을 기준의 각 스켈레톤의 특징점의 위치를 가상으로 설정해 놓은 모델이다. 이를 이용하면 얼굴 영역이 탐색 되었을 때 양 어깨의 특징점 정보  $F_2$ 와  $F_1$ 를 쉽게 얻을 수 있다.

인간의 움직임은 관절을 바탕으로 움직이기 때문에 주요 관절의 위치를 알면 전체적인 움직임 영역을 유도할 수 있다. 그림 3 은 얼굴의 위치와 가상 인간 모델을 통해 얻어진 초기 스켈레톤 모델과 그 탐색 공간을 나타낸다. 여기서  $R1$  영역은 인간의 손이 움직일 수 있는 최대 영역이며  $R2$ 는 인간의 손의 움직임이 없는 영역이다. 본 논문에서는 영상 정보를 이용하여 탐색을 할 시 탐색 시간을 줄이기 위해  $R2$  영역은 탐색 공간에 넣지 않는다. 하지만 실제적으로  $R1$ 을 원형 영역으로 설정하기 힘들기 때문에  $R3$ 과 같은 사각형 영역을 이용하여 특징점을 탐색한다.

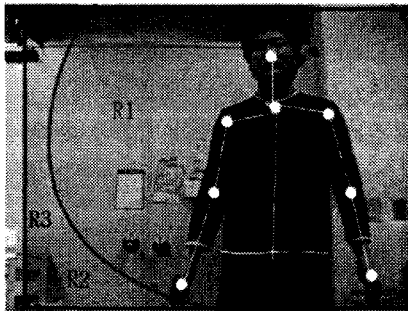


그림 3 특징점 추출 영역

### 3.3 원형 탐색 기법을 이용한 특징점 추출

본 논문에서는 특징점 추출을 위한 최소한의 탐색을 위해 원형 탐색 기법을 사용한다. 원형 탐색 기법은 그림 4와 같이 어깨의 특징점인  $F2$ 를 중심으로 팔꿈치까지 반원을 그리는 탐색 영역  $L1$ 을 설정하여 그 안에서만 특징점의 위치를 추적한다. 또한 손의 위치는  $L2$  영역에서만 탐색을 하여 탐색 시간을 단축 시키고 정확도를 높인다.

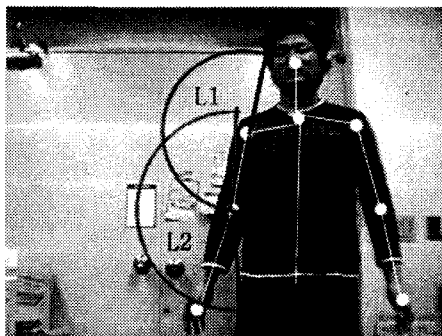


그림 4 원형 탐색 기법의 탐색공간

그림 5는 원본 영상과 앞서 계산된 4가지 색상, 움직임, 윤곽선 정보를 융합한 형태의 영상지도이다. 그림에서 밝은 영역이 인간의 몸으로 추정되는 부분이 된다. 최종 영상지도는 다음과 같다.  

$$I_f(x,y) = \alpha_1 I_r(x,y) + \alpha_2 I_s(x,y) + \alpha_3 I_l(x,y) + \alpha_4 I_p(x,y)$$

여기서  $\alpha_1, \dots, \alpha_4$  는 정도 융합을 위한 파라미터 값이다.



(a) 원본 영상 (b) 최종 영상 지도

그림 5 원형 탐색 기법을 위한 영상지도

탐색 공간내의 원형 탐색은 각 특징점을 중심으로 각 회전각에 대한 특징점이 위치할 확률을 계산한다. 이때 각 회전각의 확률 계산은 그림  $S_1$ 과 같이 넓이가  $S_w$  이고 높이가  $S_h$  인 사각형내의 영상지도 값의 합을 바탕으로 계산된다. 최종 회전각  $\theta$ 에 대한 확률  $p_\theta$ 는 다음과 같이 계산된다.

$$p_\theta = \frac{w_\theta}{\sum_{\forall \theta} w_\theta}$$

$$w_\theta = \sum_{\forall (x,y) \in S_1} I_f(x,y)$$

최종적으로 회전각 값 중 가장 큰 확률을 가지는 회전각 방향의 탐색 공간의 끝 부분이 특징점의 위치가 된다. 팔의 탐색은 위와 같은 동일한 방법으로 팔꿈치를 기준으로 탐색을 수행하면 된다.

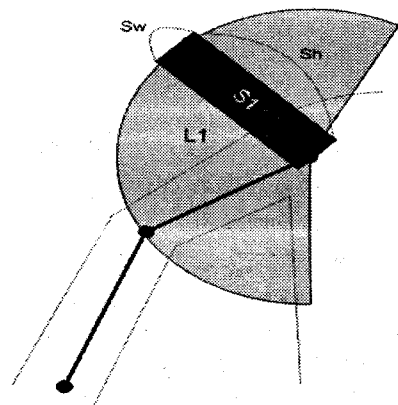


그림 6 회전 탐색 기법

## 4. 모의실험

는 최종적으로 제안된 특징 추출기의 성능을 평가하기 위해 다양한 동작에서 특징점을 추출하여 보았다. 배경을 일반적인 사무 환경으로 하였으

며 영상 입력은 320×240 24bit RGB 이며 수행 속도는 10 frame/sec 이상 동작한다.

그림 7은 제안된 강인한 특징점 추출기의 실험 결과는 나타낸다. 다양한 동작에서 정확한 동작을 추출하고 있음을 알수 있다.

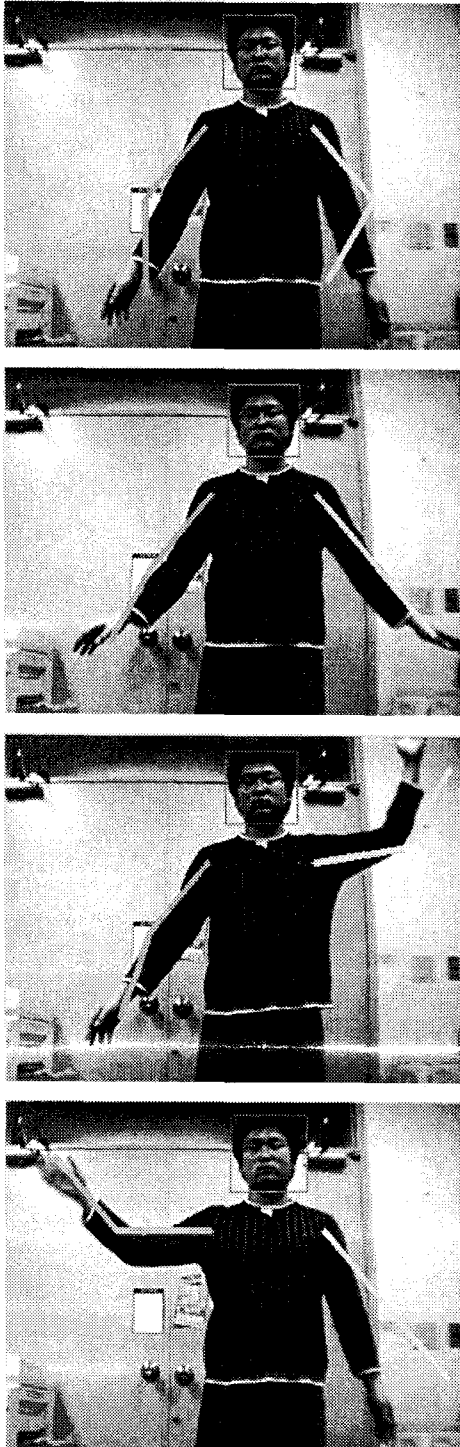


그림 7 특징점 추출 결과

## 5. 결론

로봇과 인간의 상호작용은 매우 중요한 이슈로

많은 연구가 진행되고 있다. 하지만 아직 높은 성능을 가진 상호작용 기법이 개발되고 있는 못한 실태이다. 본 연구에서는 보다 높은 성능의 특징점 추출기 개발을 위해 다양한 정보를 사용하고 가상 인간 모델로부터 얻어지는 정보를 바탕으로 지능적으로 특징점을 찾는 기법을 개발한다. 실험을 통해 제안된 기법의 성능을 확인하였다.

감사의 글 : 본 논문은 2005년도 정부재원 (교육 인적 자원부 학술연구 조성사업비)으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2005- 042-D00268)

## 참고문헌

- [1] J. K. Aggarwal and Q. Cai., "Human Motion Analysis: A Review," *Computer Vision and Image Understanding*, pp. 428-440, 1999
- [2] B. Fan; Z.-F. Wang, "Pose estimation of human body based on silhouette images," *International Conference on Information Acquisition Proceedings.*, pp. 296-300, June, 2004.
- [3] I. Haritaoglu, R. Cutler, D. Hawood and L. Davis, "Backpack: Detection of people carrying objects using silhouettes," *Computer Vision and Image Understanding*, pp. 385-397, No. 3, 2001.
- [4] I. Haritaoglu, D. Harwood, and L. Davis, "A real time system for detection and tracking people" *Journal of Image and Vision Computing*, 1999.
- [5] I. Haritaoglu, D. Hawood and L. Davis, "Who? When? Where? What? A Real Time System for Detecting and Tracking People", *Automatic Face and Gesture Recognition*, pp. 222-227, 1998.
- [6] A. Blake, M. Isard, and D. Reynard, "Learning to track curves in motion of contours," *Artificial Intelligence*, pp.101-133, 1995.
- [7] C. Anderson, P. Burt, and G. van der Wal., "Change detection and tracking using pyramid transformation techniques," *In Proceedings of SPIE Intelligent Robots and Computer Vision*, Vol 573, pp. 72-78, 1985.
- [8] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, pp. 42-77, 1994.
- [9] P. Peer, J. Kovac, and F. Solina, " Human skin colour clustering for face detection," *EUROCON*, pp. 144-148, 2003.
- [10] <http://opencvlibrary.sourceforge.net>
- [11] G. Healey and D. Slater, "Using illumination invariant color histogram descriptors for recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 355 - 360, 1994