
구성요소 기반 확률 전파를 이용한 2D 사람 자세 추정

2D Human Pose Estimation Using Component-Based Density Propagation

차은미, Eunmi Cha, 이경미, Kyoungmi Lee

덕성여자대학교 지능형 멀티미디어 연구실, 아동게임연구센터

요약 본 논문에서는 인체 추정에 필요한 인체의 각 부위들을 구성요소로 각각 검출하여 연결하는 인체 모델을 통해 각 구성요소를 개별적으로 추정하게 된다. 여기서 인체의 구성요소 중 동작 추정에 가장 필요한 6개 부위로 구성된 구성요소인 머리, 몸통, 왼팔, 오른팔, 왼발, 오른발 등을 검출하여 추적한 후, 각 구성요소의 중심값과 색상정보를 이용하여 이전 프레임과 현재 프레임 간에 연결성을 두어 각 구성요소를 개별적으로 확률 전파를 통해 추적되어지고, 각 구성요소의 추적 결과는 구성요소들의 추정 결과를 구성요소 기반 확률 전파를 이용하여 인체의 동작을 추정하는 방법을 제안한다. 입력 영상에서 피부색 등의 색상 정보를 이용하여 인체 부위 또는 인체 모델의 구성 요소들 각각의 중심값과 색상정보를 가지고 확률전파를 통해 이것이 어떤 동작인지 동작 추정이 가능하다. 본 논문에서 제안하는 인체 동작 추적 시스템은 유아의 동작교육에 이용되는 7가지 동작인 걷기, 뛰기, 양감질, 구부리기, 뺨기, 균형 잡기, 회전하기 등에 적용하였다. 본 논문에서 제안한 인체 모델의 각 구성요소 부위들을 독립적으로 검출하여 평균 96%의 높은 인식률을 나타냈고, 앞서 적용한 7가지 동작에 대해서 실험한 결과 평균 88.5% 성공률을 획득함으로써 본 논문에서 제안한 방법의 타당성을 보였다.

핵심어: *human tracking, human body component model, motion estimation, density propagation*

1. 서론

최근 사람의 동작을 추정하는 것이 컴퓨터 시각 분야에서 중요하게 인식되고 있을 뿐만 아니라 활용할 응용 분야가 급격히 증가되고 있다. 사람의 걷기를 분석하는 의학 분야, 아바타 제어를 필요로 하는 가상현실 분야, 사람-기계간 상호작용이 필요한 자동화 분야, 보안용 감시 시스템 개발 분야, 발레나 체조 동작을 분석하는 체육학 분야 등이 여기에 속한다[1,2].

인체 모델에 기반을 두지 않고 인체 전신으로 동작을 추정한다거나 인체 모델에 기반을 두어 인체를 추적하여 전신이 아닌 주로 상반신만으로 동작을 추정하거나[3], 걷는 동작으로 동작을 한정지어서 추정하거나[4], 마커를 사용하여

아바타를 제어하는 연구들이 대부분이다[5]. 그러나 마커프리 동작 캡처 장비의 최대 문제점은 광학식이나 자기식 장비들에 비하여 그 정확도가 떨어지며, 캡처 장치에서 받을 수 있는 동작에 제한이 있다는 것이다[6]. 그리고 사람의 동작들은 동일한 사람, 같은 동작이라 하더라도 조명, 겹침, 변형이나 잡음, 관측계의 오차, 특정 추출의 불완전성 등으로 인하여 예기치 않은 변화가 많이 포함될 수 있다. 이러한 성질을 가장 잘 반영해 줄 수 있는 것이 구성요소에 기반한 확률 전파를 이용한 검출 방법이다[7-9].

구성요소에 기반을 둔 방법은 겹침 등으로 인한 부분적 왜곡들이 명확하게 검출되지 않아도 적당한 인체 동작을 추정할 수 있다. 또한 인체를 검출하거나 동작을 추적하기 위해서, 인체 전신보다 인체 부위들이 자세나 조명 변화에 덜

민감하고 각 부위들이 독립적으로 검출될 수 있기 때문에, 구성요소에 기반한 인체 검출 방법의 성능이 더 좋다고 알려져 있다. 따라서 여러 개의 추적 대상에 대해 추정된 결과를 구성요소 기반 확률전파를 통한 인체 자세를 추정하는 방법에 대한 연구가 필요하다.

본 논문에서는 동작을 하는 유아에게 어떠한 마커도 부착하지 않고, 영상 처리를 통하여 인체 모델을 6개의 부위로 구성요소를 구성하고, 확률 전파를 이용하여 동작을 추정하는 방법을 소개하고 있다.

2. 구성요소기반 인체 모델

본 장에서는 프레임에서 인체 부위들을 구성요소로 검출하여 연결하는 구성요소에 기반한 인체 모델에 대해 소개한다.

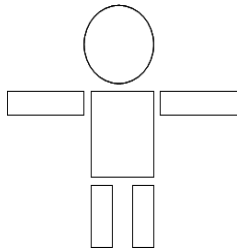


그림 1. 구성요소에 기반한 인체 모델

구성요소에 기반한 인체 모델은 사람의 인체를 6개의 부위로 구성되어진다: 머리, 몸통, 왼팔, 오른팔, 왼발, 오른발. 이렇게 6개의 부위로 구성되어진 인체 모델은 동작 추정에 가장 필요한 구성요소이다. 그림 1은 구성요소에 기반한 인체 모델을 보여준다.

6개의 부위로 이루어진 인체 모델인 *Human*은 식(1)과 같이 표현할 수 있다. 여기서 C_i 는 각 부위의 기하 정보와 외형 정보를 나타내고, R_i 는 각 부위의 연결 정보를 나타내며 i 는 6이다.

$$Human = (C_i, R_i), \quad i = 1, \dots, 6 \quad (1)$$

3. 구성요소 기반 인체 자세 추정

본 장에서는 이전 프레임의 정보들을 통해 현재 프레임에서 사람 자세를 결정하는 과정이다. 본 논문에서는 동작을 추정하기 위해 확률 전파를 이용하여 예측된 각 구성요소의 위치정보와 추적된 구성요소간의 연관 관계를 찾는 방법에 대해 기술한다.

3.1 초기화

확률전파를 통해 동작을 추정하기 전에 모델을 초기화하는 과정이 필요하다. 추적되는 사람이 비디오에 처음 동작을 시작할 때 인체 모델을 초기화하는 것으로 시작될 수 있다.

본 논문에서는 7가지 동작에 대한 영상을 10개의 프레임으로 정규화 시켰고, 그 중 구부리기 동작은 동작 특성상 영상 전체를 하나의 주기로 정의하였다. 배경영상과 전경영상을 분리하는 차영상 계산을 통해 구해진 전경영상으로부터 피부색 등과 같은 색상 정보를 통해 인체 모델을 구성할 영역을 각각 검출하고, 라벨링 과정을 통해 불필요한 잡음을 제거하고, 앞서 구성한 인체 모델에 대한 각 구성요소 영역을 분할하고 인체 모델을 설계하여 추적에 이용할 것이다. 그림 2는 피부색 등과 같은 색상정보를 통해 구해진 영역에 대한 검출을 보여준다.



그림 2. 7가지 동작에 대한 구성요소별 영역 검출

3.2 구성요소 기반 확률 전파

기하 정보와 외형 정보를 가진 인체 모델의 각 구성요소인 C_i 와 각 부위의 연결 정보 R_i 를 이용해 인체 부위의 위치를 추정하게 된다. 이 단계에서 추정된 인체의 움직임의 정보를 가지고 이전 프레임과 다음 프레임에서의 인체의 움

직입에 연결 관계를 반영하여 추적된 정보를 통해 동작을 추정하게 된다.

확률전파 $P(X_k|Z_k)$ 는 파티클 필터와 같이 k-1번째 프레임까지 관찰된 내용을 바탕으로 k번째 프레임에서 어떤 구성요소일 것인지에 대한 확률을 추정하는 것이다. 만약 $P(X_{t-1}|Z_{t-1})$ 이 주어진다면, k번째 프레임에서의 확률은 k-1번째 프레임에서의 표본을 전파함으로써 구할 수 있다. k-1번째 프레임에서 얻은 $P(X_{k-1}|Z_{k-1})$ 으로부터 전파된 표본을 적용하여 $P(X_k|Z_{k-1})$ 을 얻는다. 이를 베이지 정리로 적용하면 다음과 같다.

$$P(X_k|Z_{k-1}) = \frac{P(Z_{k-1}|X_k)P(X_k)}{P(Z_{k-1})} \quad (2)$$

본 논문에서는 한 프레임에 6개의 구성요소 C의 확률인 $P(X_k^C|Z_{k-1}^{C_i})$ 을 이용하여 동작 추정이 가능하다. 임의의 구성요소 C와 가장 유사한 구성요소를 찾기 위해 구성요소별로 구해진 확률들을 비교하여 확률이 가장 높은 것으로 해당 구성요소를 결정한다.

$$\max P(X_k^C|Z_{k-1}^{C_i}) \quad i=1, \dots, 6$$

이렇게 구해진 확률은 가중치의 합으로 융합됨으로써 동작을 추정할 수 있다. 가중치 w_i 는 0과 1사이의 값이며, 그 합은 1이다. 여기서 I는 머리, 몸통, 왼팔, 오른팔, 왼발, 오른발로써 6이다. 식으로 표현하면 다음과 같다.

$$\frac{1}{I} \sum_{i=1}^I w_i P(X_k^C|Z_{k-1}^{C_i}) \text{ 그리고 } \sum_{i=1}^I w_i = 1 \quad (3)$$

4. 실험 및 결과

본 논문에서 실험에 사용된 동영상은 캠코더 Sony DCR-PC330에서 320x240 크기의 영상을 초당 15 프레임으로 획득하여 Pentium-IV 3.0GHz인 CPU와 1GB 메모리를 가진 PC에서 수행하였으며, 프로그램은 Microsoft Visual C++ 6.0을 사용하여 구현되었다.

본 논문에서는 앞서 3.1절에서 설명한 초기화 과정에 대한 결과를 보여준다. 전처리 과정을 통해 배경과 전경을 분리한 후, 배경이 분리된 영상에서 불필요한 잡음을 제거하고 라벨링 과정을 통해 구성요소별 영역을 분할하게 된다. 그림 3은 연속된 원영상과 그에 따른 라벨링 과정을 거친 균형잡기, 양감질, 뛰기, 걷기, 회전하기, 구부리기, 뺨기 등 7가

지 동작들에 대한 라벨링 영상과 각 구성요소별 영역을 검출한 영상을 보여주고 있다.

앞서 3.2절에서 설명한 각 구성요소의 중심값과 색상 정보를 이용하여 확률 전파를 이용하여 구해진 각 구성요소의 궤적을 보여준다. 그림 4는 각 구성요소에 대한 7가지 동작의 궤적을 모두 보여주고 있다.

	구성요소					
	1	2	3	4	5	6
	머리	몸통	왼팔	오른팔	왼발	오른발
구성요소 추적 성공률	93	93	98	99	100	97

표 1. 구성요소별 추적 성공률(%)

표 1은 840개의 테스트 프레임에 대해 각 구성요소별 추적 성공률을 나타낸다. 평균 96%의 인식률을 얻었으며, 머리는 93%, 몸통은 93%, 왼팔은 98%, 오른팔은 99%, 왼발은 100%, 오른발은 97%의 높은 성공률을 나타내었다.

	동작						
	1	2	3	4	5	6	7
	균형잡기	양감질	구부리기	뛰기	뺨기	회전하기	걷기
동작 인식 성공률	95	89	58	90	99	91	98

표 2. 동작 인식 성공률(%)

표 2는 210개의 테스트 프레임에 대한 7가지 동작 인식률을 나타낸다. 평균 88.5%의 인식률을 얻었으며, 균형잡기는 95%, 양감질은 89%, 구부리기는 58%, 뛰기는 90%, 뺨기는 99%, 회전하기는 91%, 걷기는 98%의 인식률을 나타내었다.

구부리기 동작은 영상 전체를 하나의 주기로 정의하여 일정 주기 동안 6개 구성요소 중에 머리 부위가 보여 지지 않아 다른 동작들에 비해 상대적으로 낮은 인식 성공률을 보였다.

5. 결론 및 향후 연구계획

본 논문에서는 구성요소에 기반한 확률 전파를 사용하여 7가지 동작이 이루어지는 동작 추정에 적합한 인체 모델을 제시하였다. 기존의 동작 추정 연구 방법을 개선하기 위하여 이전 프레임의 위치 정보를 활용한 확률 전파의 특징을 반영하여 인체 모델을 제안하였고, 각 구성요소 부위들을 독립적으로 검출하였기 때문에 표 1에서 보는 것과 같이 평균

96%의 높은 인식률을 보였다. 제안된 방법은 조명에 덜 민감할 뿐만 아니라, 마커 없이 동작을 추정함으로써 상대적으로 인체 동작 추적이 용이하며 다양한 동작들이 포함된 영상에서 사람이 취하고 있는 동작을 추정할 수 있다. 또한 겹침 등으로 인해 부분적으로 왜곡되어 모든 구성요소가 명확하게 검출되지 못하여도 확률 전파를 이용하여 동작을 추정하기 때문에 적당한 인체 동작을 추정할 수 있다. 7가지 동작에 대해 실험한 결과 표 2에서 보는 것과 같이 평균 88.5%의 인식률을 획득함으로써 본 논문에서 제안한 방법의 타당성을 보였다.

향후 연구계획으로는 본 논문에서 앞서 제안했던 인체 동작 추적 시스템은 2D가 아닌 3D로 확장하여 구성요소에 기반한 동작을 추정함으로써 좀 더 정확한 추정 결과를 보완해야 한다. 또한 추정방법에 있어서도 단순한 확률전파가 아닌 Kalman filter 등 기타 다른 방법을 모색하여 연구를 계속 진행할 것이다.

Acknowledgement

본 연구는 2005년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행되었음.
(KRF-2005-042-D00285)

참고문헌

- [1] L. Campbel and A. Bobick, "Recognition of Human Body Motion Using Phase Space Constraints", MIT Media Lab Perceptual Computing Group TR 309, 1995.
- [2] E. Clergue, et al. "Automatic Face and Gestual Recognition for Video Indexing", Int' l Workshop on Automatic Face and Gesture Recognition, pp.110-115, 1995.
- [3] O. Bernier, P. Cheung-Mon-Chan, "Real-time 3D Articulate Pose Tracking using Particle Filtering and Belief Propagation on Factor Graphs", bmvc, 2006.
- [4] X. Zhiwei, V. Chi, V. Nikolaos, Boulgouris, "Human gait recognirion based on matching of body components", IEEE, 2006.
- [5] B. Rosenhahn, U.G. Kersting, A.W. Smith, J.K.Gurney, T. Brox, R. Klette, "A System for Marker-less Human Motion Estimation", LNCS, pp.230-237, 2005.
- [6] 이재호, 박창준, 이인호, "동작 인식기를 이용한 3D 게임제어", HCI 2006 학술대회 발표논문집, 2006.
- [7] S. Frey et al. "Das Berner System zur Unter-shchung nonverbaler Interaktion", Methoden der Analyse von Face-to-Face-Situationen, Stuttgart, Metzler, pp. 203-236, 1981.
- [8] K.S. Fu "Syntactic Methods in Pattern Recognition", Academic Press, pp. 124-229, 1974.
- [9] R.C. Gonzalez and M.G. Thomason, "Syntactic Pattern Recognition", Addison-Wesley Publicshing Company, pp. 117-270, 2005.
- [10] C.Cheng, A.Rashid, K.Ashfaq, "Density Propagation for Tracking Initialization with Multiple Cues", icassp, 2004.
- [11] 이경미, 김혜정, 이윤미, "사람 자세 추정을 위한 모델 기반 추적", HCI 2006 학술대회 발표논문집, 2006.
- [12] 조경은, 조형제, "확률적 정규 문법 추론법에 의한 사람 몸동작 인식", 정보과학회논문지 : 컴퓨팅의 실제 제7권 제3호, 2001.
- [13] 이용재, 이철우 "전역 및 부분 특징 정보를 이용한 제스처 인식", 정보과학회논문지: 소프트웨어 및 응용 제32권 제8호, 2005.

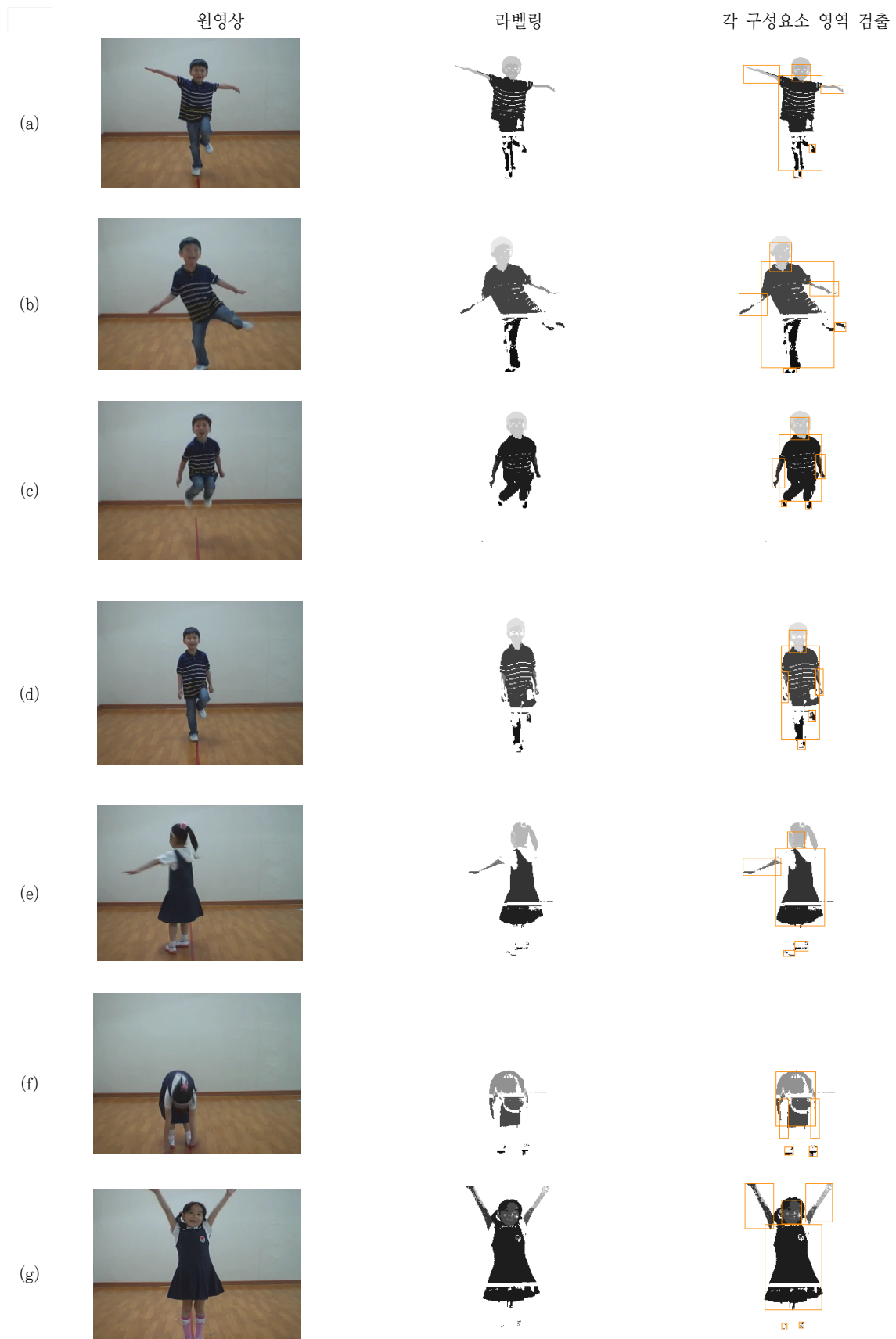
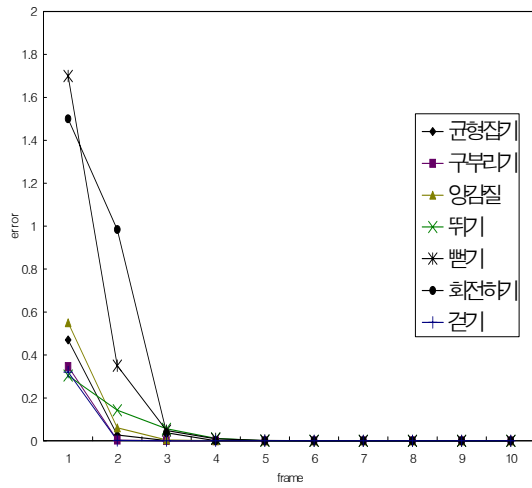
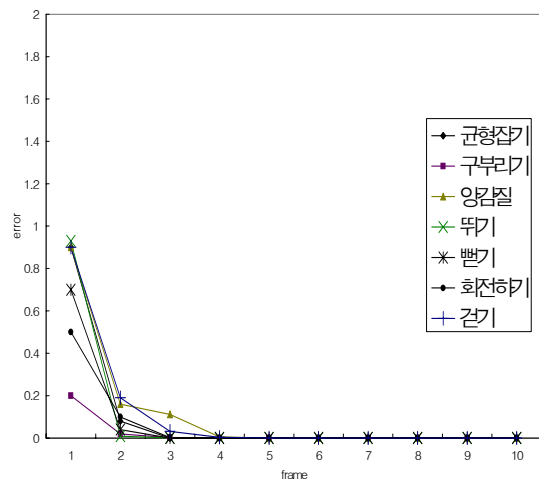


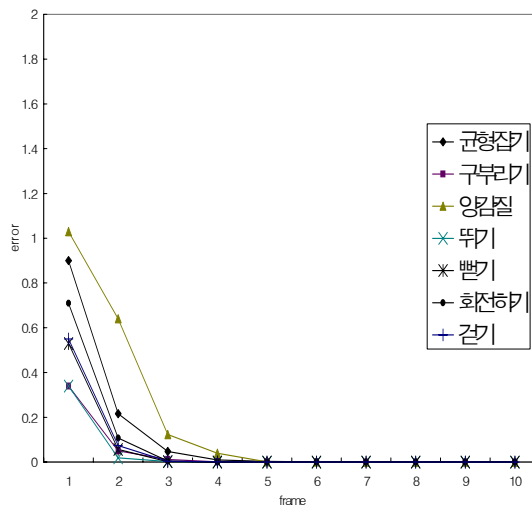
그림 3. 7가지 동작들의 시퀀스들. (a)균형 잡기, (b)양금질, (c)뛰기, (d)걸기, (e)회전하기, (f)구부리기, (g)뻗기



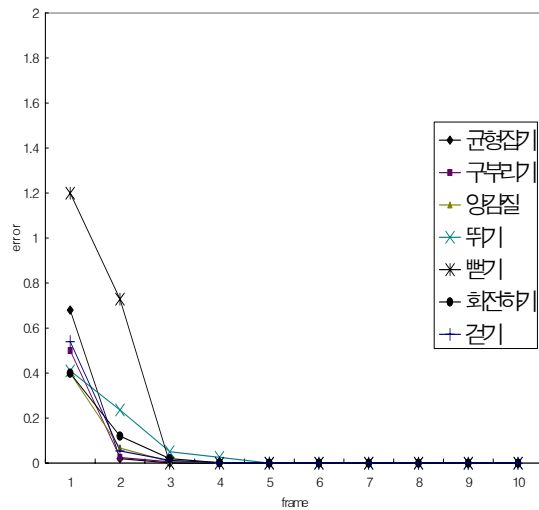
(a) 머리



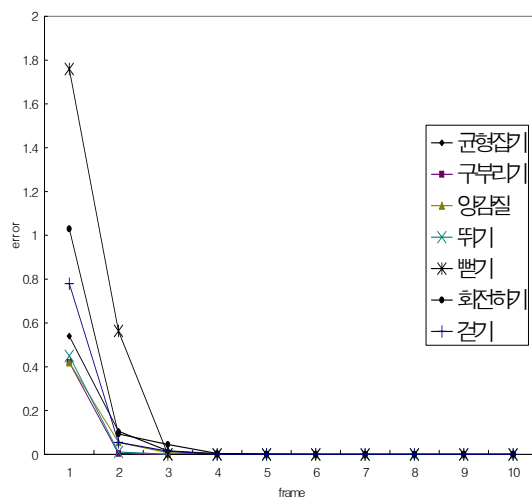
(b) 몸통



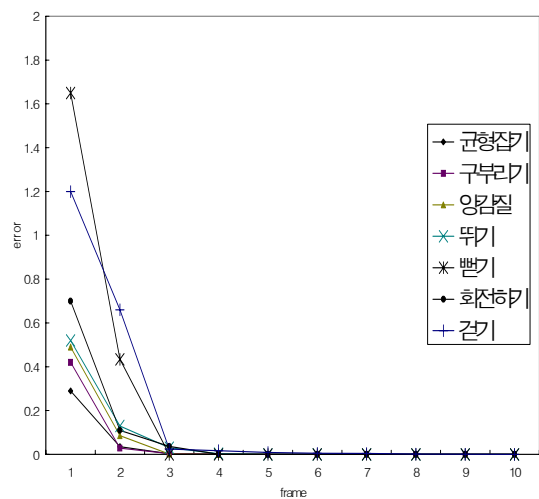
(c) 왼팔



(d) 오른팔



(e) 왼발



(f) 오른발

그림 4. 7가지 동작에 대한 구성요소별 2D 위치의 오차 분포