편향된 다양체 학습 기반 시점 변화에 강인한 인체 포즈 추정

(View-Invariant Body Pose Estimation based on Biased Manifold Learning)

허 동 철 † 이 성 환 ††
(Dong-Cheol Hur) (Seong-Whan Lee)

요 약 다양체는 고차원 표본 데이터들 사이의 관계를 표현하기 위해 저차원 공간에서 생성된 구조로서 고차원 데이터인 영상과 3차원 인체 구성 데이터를 처리하는데 많이 사용되고 있다. 다양체 학습은 이러한 다양체를 생성하는 과정을 말한다. 그러나 다양체 학습을 이용한 포즈 추정은 학습하지 못한 실루엣 변화에 취약하다. 실루엣 변화는 2차원 영상에서 시점 변화, 포즈 변화, 사람 변화, 거리 변화, 잡영에 의해 발생되며, 이러한 변화를 하나의 다양체로 학습하기란 어렵다. 본 논문에서는 실루엣 변화를 유발하는 문제중 하나인 시점 변화에 대한 문제를 해결하고자 한다. 종래에 시점 변화에 상관 없이 포즈를 추정하는 방법에서는, 각 시점마다 다양체를 가지거나 사상 함수에서 시점에 관련한 요소들을 분리하여 별도의 다양체로 학습한다. 하지만 이러한 방법들은 복잡하고, 추정 과정에서 어떠한 시점의 다양체를 통해 포즈를 추정할지 판단을 요구하며, 비교사 학습으로 인해 실루엣과 대응되는 3차원 인체 구성을 지정하기 어렵다. 본 논문에서는 시점 다양체, 포즈 다양체, 인체 구성 다양체를 편향된 다양체로 학습하여 사용하는 방법을 제안한다. 그리고 영상과 시점 다양체, 영상과 포즈 다양체, 인체 구성과 인체 구성 다양체, 포즈 다양체와 인체 구성 다양체 간에 사상 함수를 학습한다. 실험에서는 학습된 다양체와 사상 함수를 이용하여 24개의 시점에서 강인한 포즈 추정 결과를 보여주고 있다.

키워드 : 포즈 추정, 시점 추정, 다양체 학습, 교사 학습

Abstract A manifold is used to represent a relationship between high-dimensional data samples in low-dimensional space. In human pose estimation, it is created in low-dimensional space for processing image and 3D body configuration data. Manifold learning is to build a manifold. But it is vulnerable to silhouette variations. Such silhouette variations are occurred due to view-change, person-change, distance-change, and noises. Representing silhouette variations in a single manifold is impossible. In this paper, we focus a silhouette variation problem occurred by view-change. In previous view invariant pose estimation methods based on manifold learning, there were two ways. One is modeling manifolds for all view points. The other is to extract view factors from mapping functions. But these methods do not support one by one mapping for silhouettes and corresponding body configurations because of unsupervised learning. Modeling manifold and extracting view factors are very complex. So we propose a method based on triple manifolds. These are view manifold, pose manifold, and body configuration manifold. In order to build manifolds, we employ biased manifold learning. After building manifolds, we learn mapping functions among spaces (2D image space, pose

이 논문은 2006년도 교육과학기술부의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. KRF-2006-311-D00197)

† 학생회원 : 고려대학교 컴퓨터학과
dcheo@image.korea.ac.kr
†† 종신회원 : 고려대학교 정보통신대학 교수
swlee@image.korea.ac.kr
(Corresponding author임)

논문접수 : 2009년 8월 13일
심사완료 : 2009년 10월 12일

정보과학회논문지: 소프트웨어 및 응용 제36권 제11호(2009.11)

manifold space, view manifold space, body configuration manifold space, 3D body configuration space). In our experiments, we could estimate various body poses from 24 view points.

Key words : Pose Estimation, View Estimation, Manifold Learning, Supervised Learning

1. 서 론

다양체는 고차원 표본 데이터들 사이의 관계를 표현하기 위해 저차원 공간에서 생성된 구조로서 고차원 데이터인 영상과 3차원 인체 구성 데이터를 처리하는데 많이 사용되고 있다. 대다수의 다양체 기반 인체 포즈 추정 연구들은 다양체를 통해 고차원 데이터를 저차원의 공간에서 쉽게 다루고 빠른 포즈 추정 결과 및 정확한 포즈 추정을 얻기 위한 방법들에 집중되어 왔다. 하지만 이러한 다양체 학습 방법에는 예상치 못한 실루엣 변화에 취약한 단점이 존재한다. 실루엣 변화를 일으키는 대표적인 요인 중에 하나는 시점 변화를 들 수 있다. 같은 포즈라도 시점에 따라 다르게 보이기 때문에 시점에 상관 없이 포즈를 추정하기는 어렵다[1-4].

종래의 시점에 강인한 포즈 추정을 위해서는 시점마다 다양체를 학습하거나[2], 시점에 불변하는 인체 구성 데이터로부터 만든 다양체에서 여러 시점의 실루엣으로 사상하는 과정에서 시점에 관련된 요소들을 분리하는 방법을 사용하기도 한다[3]. 하지만 이러한 방법들은 실루엣과 대응되는 3차원 인체구성에 대해 지정하여 학습할 수 없으며, 특정 포즈들은 다양체 공간에서 같은 위치에 사상되기도 한다[1,4].

본 논문에서는 영상에서 실루엣 변화와 포즈의 변화를 별도로 표현하기 위해서 편향된 다양체 학습을 통해서 시점 다양체, 포즈 다양체를 학습하고, 입력된 포즈와 대응되는 3차원 인체 구성의 대응 관계를 만들기 위해 인체 구성 다양체를 학습한다. 학습된 다양체들과 영상 데이터 및 3차원 인체 구성 데이터를 가지고 일반화된 회귀 신경망[5]을 이용하여 사상 함수를 학습한다. 학습된 다양체와 사상 함수를 사용하여 입력되는 실루엣에 대응되는 3차원 인체 구성을 추정하게 된다.

2. 다양체 학습 및 인체 포즈 추정

2.1 제안한 방법의 구성

학습 단계에서 2차원 영상 데이터와 대응되는 3차원 데이터를 사용하여 그림 1과 같이 편향된 다양체 학습을 통하여 포즈 다양체, 시점 다양체, 인체 구성 다양체를 학습하고 일반화된 회귀 신경망[5]을 사용하여 공간 사이에 사상 함수를 학습한다. 학습된 다양체와 사상 함수를 사용하여 그림 2와 같이 추정 단계에서 입력 실루엣으로부터 시점 다양체와 포즈 다양체에서의 대응점을 찾고, 포즈 다양체의 대응점에 대하여 인체 구성 다양체에서 대응점을 찾는다. 찾은 대응점에 가장 근접한 3차원 인체 구성이 입력된 실루엣에 대응되는 3차원 인체 구성이 된다.

2.2 편향된 거리 행렬

편향된 다양체 학습에서는 기존에 다양체 학습에서 사용하는 학습 표본 간에 유클리드 거리 행렬이 아닌 편향된 거리 행렬을 구성할 필요가 있다[4]. 이 행렬은 표본 간의 유클리드 거리를 라벨 데이터를 사용하여 편향된 거리로 변형한다. 이 거리 행렬은 다양체 학습에서 이웃을 결정하는데 사용되어 편향된 거리가 가까운 표본끼리 다양체 공간에서 같은 지점으로 모이게 된다. 편향된 다양체 학습을 위한 편향된 거리 행렬을 생성하기 위해서는 식 (1)을 사용하여 원래의 유클리드 거리를 편향된 거리로 수정하게 된다[4].

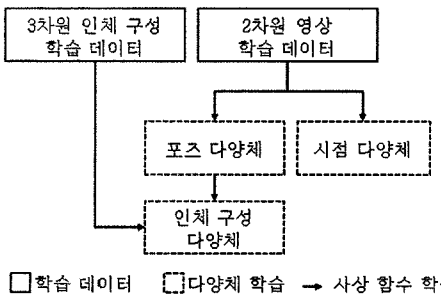


그림 1 학습 단계: 세 개의 다양체(포즈, 시점, 인체 구성)와 사상 함수를 학습

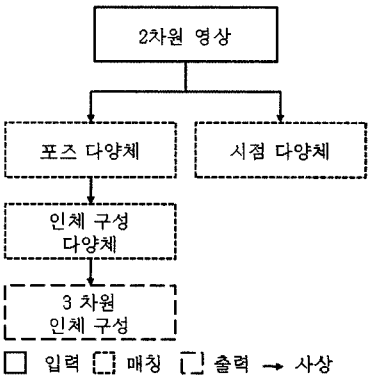


그림 2 추정 단계: 2차원 영상으로부터 3차원 포즈를 복원

$$\tilde{D}(i,j) = \begin{cases} \dfrac{a \times P(i,j)}{\max_{m,n} - P(i,j)} \times D(i,j) & P(i,j) \neq 0 \\ 0 & P(i,j) = 0 \end{cases} \quad (1)$$

$$P(i,j) = \sqrt{(L(i) - L(j))(L(i) - L(j))^T} \quad (2)$$

$$D(i,j) = \sqrt{(X(i) - X(j))(X(i) - X(j))^T} \quad (3)$$

$$\max_{m,n} = \max_{m,n} P(m,n) \quad (4)$$

식 (1)의 $\tilde{D}(i,j)$는 표본 i와 j간의 편향된 거리를 나타내며, 식 (2)의 $P(i,j)$는 표본 i와 j간의 라벨 거리를 나타낸다. $L(\cdot)$는 표본에 주어진 라벨이다. 그리고 식 (3)의 $D(i,j)$는 표본 i와 j간의 유클리드 거리를 나타낸다. 식 (4)의 $\max_{m,n}$은 표본 중에 최대 라벨 거리를 갖는 표본 m과 n 사이의 라벨 거리이며, a는 비례 상수이다. 그림 3, 4, 5에서 P와 V는 학습 전에 각 표본마다 주어진 라벨을 나타낸다.

그림 3(b)는 유클리드 거리를 사용하여 그림 3(a)에 대해 근접 포즈 및 시점 이웃을 결정한 결과이다. 그림 3(b)를 살펴 보면 비교사 학습으로 인하여 포즈 P와 시점 V가 얽혀서 결정된 결과를 얻게 된다.

그림 4(b)는 편향된 거리를 사용하여 그림 4(a)의 영상에 대해 근접 포즈 이웃을 결정한 결과이다. 그림 4(b)를 살펴보면 포즈 P가 같은 모습을 볼 수 있었으며, 시점 V도 일정하게 증가하는 모습을 볼 수 있다.

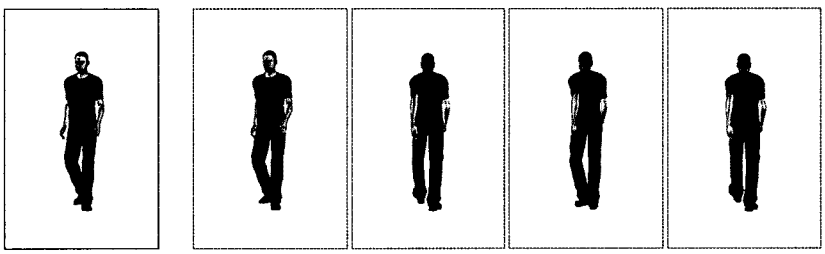
그림 5(b)는 편향된 거리를 사용하여 그림 5(a)에 대해 근접 시점 이웃을 결정한 결과이다. 그림 5(b)를 살펴보면 포즈 P는 다르지만 시점 V는 모두 같은 것을 볼 수 있다.

2.3 편향된 다양체 학습

제안한 방법에서는 시점 변화와 포즈 변화로 인한 실루엣의 변화 그리고 3차원 인체 구성 데이터의 변화를 표현하기 위해서 세 개의 다양체를 학습한다. 앞서 생성한 편향된 거리를 다양체 학습 알고리즘 중에 하나인 Locally Linear Embedding(LLE)[6]에서 이웃을 결정하는데 사용한다. 그림 6은 유클리드 거리로 학습한 다양체와 편향된 거리로 학습한 다양체를 보여준다. 그림 6(a)의 유클리드를 사용한 다양체와는 다르게 그림 6(b)의 편향된 거리를 사용하여 생성한 다양체의 경우 포즈 순서를 나타내는 점이 색에 따라 일정하게 정렬된 모습을 보여주고 있다.

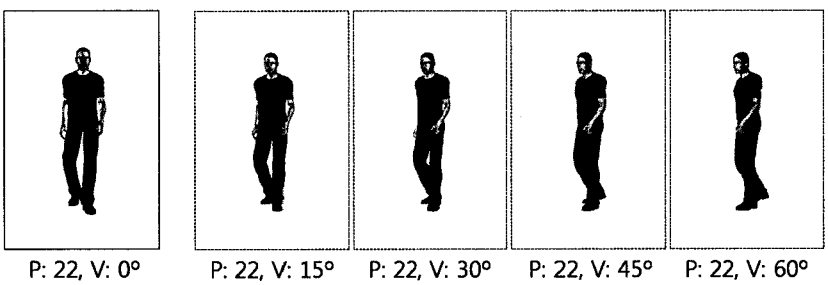
2.4 사상 함수 학습

3차원 포즈 복원을 위해 실루엣으로부터 포즈 다양체, 실루엣으로부터 시점 다양체, 포즈 다양체에서 인체 구성 다양체, 인체 구성 다양체로부터 인체 구성 데이터에 대한 사상 함수를 만들어야 한다. 기존 연구[3]에서는



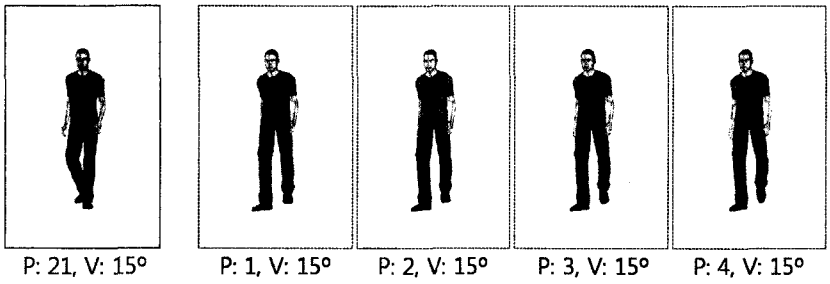
(a) 표본 (b) 유클리드 거리를 사용한 시점 및 포즈에 대한 근접 이웃

그림 3 유클리드 거리를 사용한 근접 포즈 및 시점이웃 결정



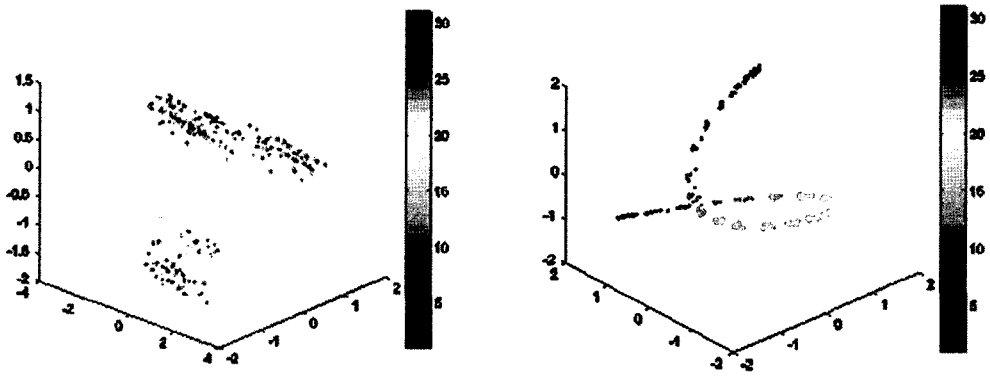
(a) 표본 (b) 편향된 거리를 사용한 근접 포즈 이웃

그림 4 편향된 거리를 사용한 근접 포즈 이웃 결정



(a) 표본 (b) 편향된 거리를 사용한 근접 시점 이웃

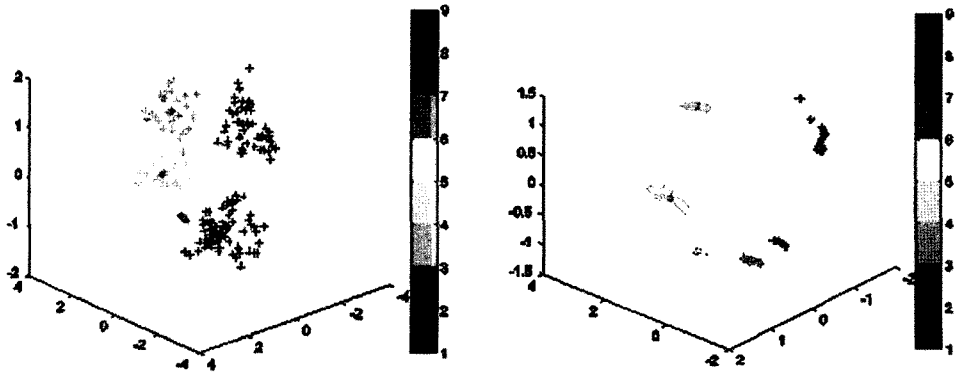
그림 5 편향된 거리를 사용한 근접 시점 이웃 결정



(a) 유클리드 거리를 사용한 다양체 학습 (b) 편향된 거리를 사용한 다양체 학습

그림 6 다양체 학습 비교

비선형 사상 함수를 만들기 위해서 다양체 공간에서 한 점과 그에 이웃한 점들을 결정하는 과정을 거치게 된다. 이 과정에서 편향된 다양체를 학습할 때 사용한 편향된 거리를 사용할 수 없어 원래 데이터와 차원 축소된 다양체 공간의 데이터 사이에 사상 결과가 그림 7(a)와 같이 좋지 않다. 그래서 본 논문에서는 회귀 방법을 사용하여 각 공간 사이에 사상 함수를 학습하기로 하였다. 회귀 방법으로는 일반화된 회귀 신경망[5]을 사용하였다. 회귀 신경망은 그림 7(b)와 같이 일정하게 분포된 결과를 보여준다.



(a) 비선형 사상 함수 (b) GRNN 회귀 함수

그림 7 비선형 사상 함수와 회귀 함수의 비교

$$\Psi_{ip} = \Psi\left(M_p, X_i\right), M_p \in R^3, X_i \in R^{2000} \quad (5)$$

$$\Psi_{iv} = \Psi\left(M_v, X_i\right), M_v \in R^3, X_i \in R^{2000} \quad (6)$$

$$\Psi_{pk} = \Psi\left(M_k, M_p\right), M_k \in R^3, M_p \in R^3 \quad (7)$$

$$\Psi_{bk} = \Psi\left(M_k, X_b\right), M_k \in R^3, X_b \in R^{62} \quad (8)$$

식 (5)의 $\Psi_{ip}: R^{2000} \rightarrow R^3$은 50×40 영상 데이터 X_i를 입력, 포즈 다양체 공간 좌표 M_p를 목표 값으로 하여 학습된다. 식 (6)의 $\Psi_{iv}: R^{2000} \rightarrow R^3$은 영상 데이터 X_i를 입력, 3차원 시점 다양체 공간 좌표 M_v를 목표 값으로 하여 학습된다. 식 (7)의 $\Psi_{pk}: R^3 \rightarrow R^3$는 포즈 다양체 공간 좌표 M_p를 입력, 인체 구성 다양체 공간 좌표 M_k를 출력으로 하여 학습된다. 식 (8)의 $\Psi_{bk}: R^{62} \rightarrow R^3$ 는 3차원 인체 구성(62 차원) $X_b \in R^{62}$를 입력, 인체 구성 다양체 공간 좌표 M_k를 출력으로 하여 학습된다.

2.5 3차원 인체 포즈 추정

앞서 만들어진 다양체와 사상 함수들을 이용하여 3차원 인체 포즈를 추정한다. 3차원 포즈를 추정하기 위해 입력된 영상에서 실루엣을 구하고 50×40 크기의 해상도로 변환한다. 변환된 하나의 실루엣 영상을 입력으로 아래의 식들을 이용하여 실루엣에 대응되는 3차원 포즈를 찾는다.

$$M_p^* = \arg\min_{M_p} \left\| M_p - \Psi_{ip}\left(X_{i^*}\right) \right\|^2 \quad (9)$$

$$M_k^* = \arg\min_{M_k} \left\| M_k - \Psi_{pk}\left(M_p^*\right) \right\|^2 \quad (10)$$

$$M_b^* = \arg\min_{M_b} \left\| M_k^* - \Psi_{bk}\left(X_b\right) \right\|^2 \quad (11)$$

$X_{i^*} \in R^{2000}$는 새로운 입력 실루엣 영상을 나타내며, 이 새로운 영상으로부터 식 (9)의 사상 함수 $\Psi_{ip}: R^{2000} \rightarrow R^3$를 이용하여 실루엣에 대응되는 포즈 다양체에서 한점 M^*_p를 찾고, 식 (10)의 사상 함수 $\Psi_{pk}: R^3 \rightarrow R^3$를 이용하여 인체 구성 다양체에서 한점 M^*_k을 찾는다. 그리고 식 (8)의 사상 함수 $\Psi_{bk}: R^{62} \rightarrow R^3$과 식 (11)을 이용하여 인체 구성 다양체 한점에 대응되는 3차원 인체 구성 $X^*_b \in R^{62}$을 찾아낸다.

3. 실험 및 결과 분석

본 논문에서는 학습과 실험을 위해 Poser 7 프로그램[7]을 통하여 24개의 시점에서 걷기 동작에 대한 영상 및 3차원 포즈 데이터를 획득하였으며, 30 프레임 길이의 걷기 동작과 24개의 시점을 포함하여 총 720프레임을 가지고 실험을 수행하였다. 그리고 Motion Golf 3D Swing Analyzer 프로그램[8]을 이용하여 골프 스윙 동작에 대해 8개의 시점에서 각 시점마다 60프레임 길이의 영상을 획득하였다.

그림 8과 그림 9는 각각 입력된 실루엣에 대응되는 포즈, 시점 파라미터에 대한 진리값과 비교한 모습을 보여준다. 30개의 포즈에 대한 파라미터 추정 결과에서는 포즈와 인체 구성에서는 좋은 결과를 보여주고 있지만 동작의 시작과 끝 부분에서 오류를 확인할 수 있었다. 그리고 시점에 대해서는 전반적으로 좋은 결과를 보여주고 있지만 간혹 중간에 시점 파라미터 추정에서 오차를 발견할 수 있었다.

그림 10은 시점을 고정한 채 포즈가 변화하는 모습을 보여준다. 그림 10(b)의 시점 다양체에서는 위치가 변하지 않지만, 그림 10(c)의 포즈 다양체에서는 위치가 변하는 모습을 보여준다. 그림 10(d)는 그림 10(a)에 대응하는 3차원 인체 구성을 보여준다.

그림 11은 포즈를 고정한 채 시점이 변화하는 모습을

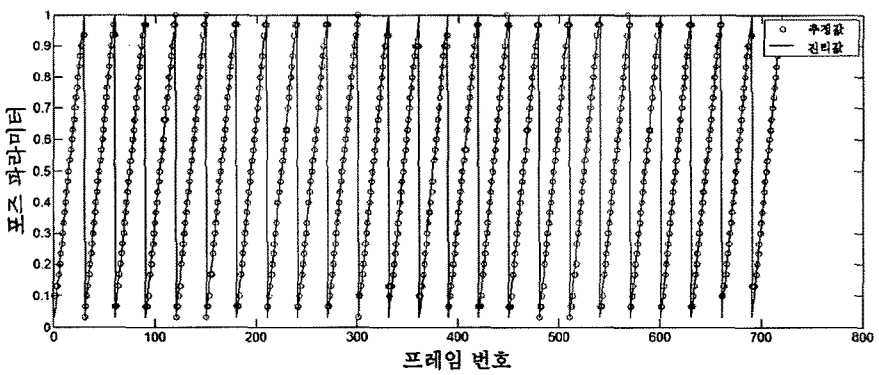


그림 8 포즈 파라미터 추정 결과(실선: 진리값, 원: 추정값)

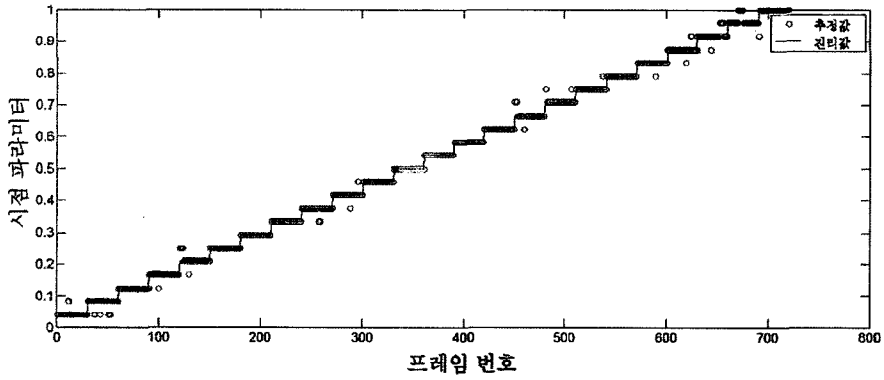
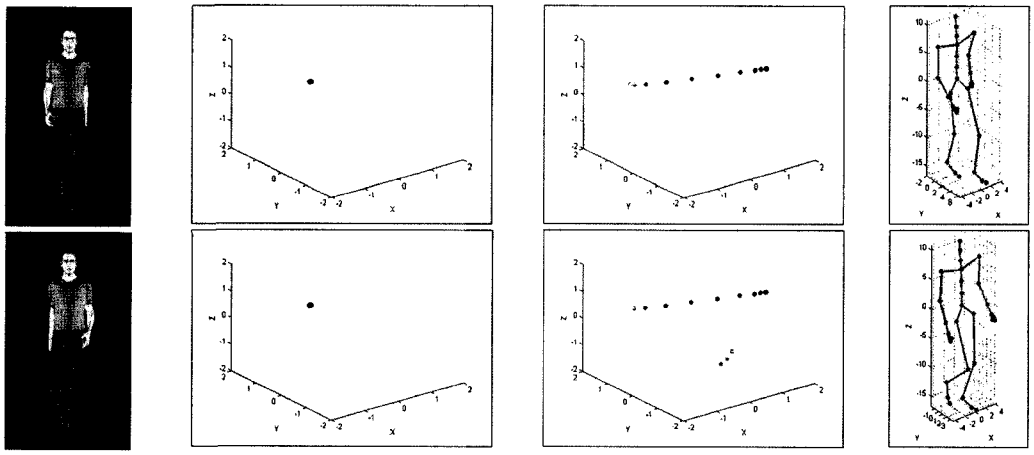
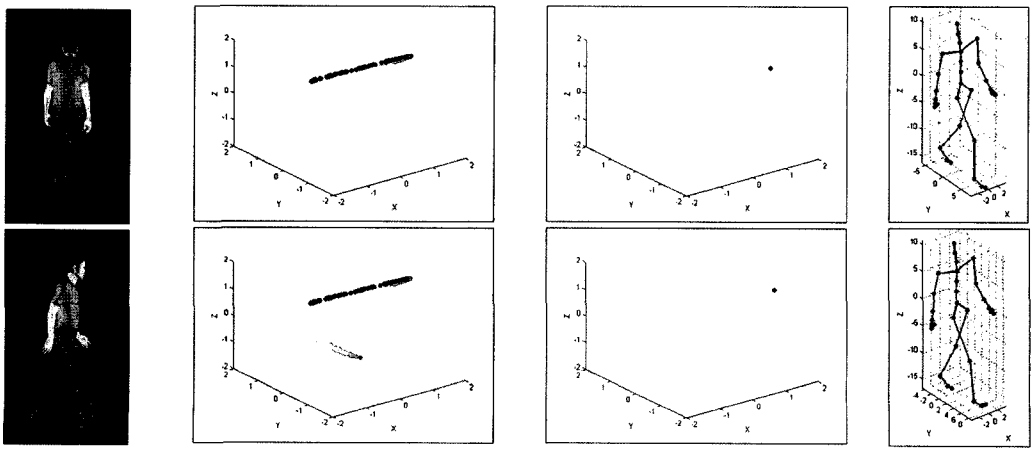


그림 9 시점 파라미터 추정 결과(실선: 진리값, 원: 추정값)



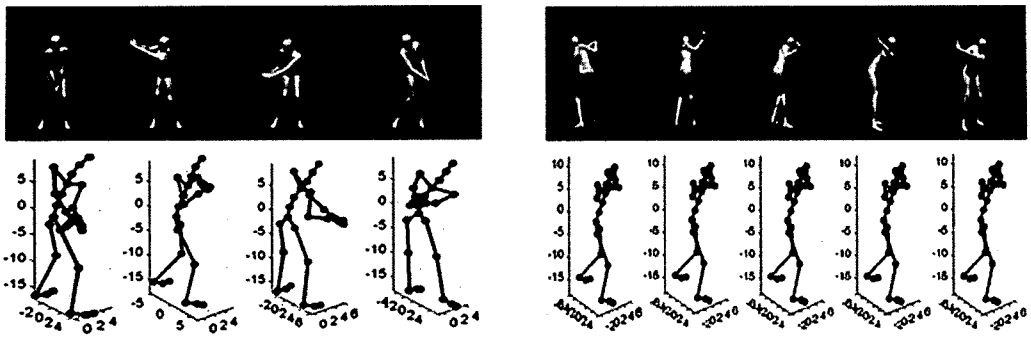
(a) 입력 영상 (b) 시점 다양체 (c) 포즈 다양체 (d) 추정 결과

그림 10 하나의 시점에서 다수의 포즈 복원 결과



(a) 입력 영상 (b) 시점 다양체 (c) 포즈 다양체 (d) 추정 결과

그림 11 다수의 시점에서 같은 포즈 복원 결과



(a) 한 시점에서 포즈 추정 (b) 여러 시점에서 포즈 추정

그림 12 골프 스윙 동작에 대한 포즈 추정 (첫째 줄: 입력 영상, 둘째 줄: 추정된 포즈 결과)

보여준다. 그림 11(c)의 포즈 다양체에서는 위치가 변하지 않지만 그림 10(b)의 시점 다양체에서는 위치가 변하는 모습을 보여준다.

그림 12는 Motion Golf 3D Swing Analyzer 프로그램[8]을 이용하여 생성한 영상에서 포즈를 추정한 결과를 보여준다. 그림 12(a)에서는 시점을 고정한 채 여러 포즈를 추정한 결과를 보여주고 있다. 그림 12(b)는 포즈를 고정한 채 여러 시점에서 획득한 영상에서 포즈를 추정한 결과를 보여주고 있다.

4. 결론 및 향후 연구 방향

실루엣 변화를 일으키는 요인에는 시점 변화, 포즈 변화, 사람 변화, 거리 변화, 잡영 등이 있다. 이러한 변화들을 하나의 다양체로 표현하기란 어렵다.

본 논문에서는 영상의 실루엣 변화에서 시점 변화와 포즈 변화, 3차원 인체 구성 데이터의 변화를 편향된 다양체로 학습하고, 이로부터 일반화된 회귀 신경망에 기반하여 각 데이터 공간 사이에 사상하는 함수를 학습하여, 학습된 다양체와 사상 함수를 기반으로 입력된 2차원 실루엣 영상로부터 3차원 포즈를 추정할 수 있었다.

기존의 다양체 학습에 기반한 인체 포즈 추정은 학습된 하나의 동작에 대해서만 포즈 추정을 할 수 있었다. 하지만 다양체 학습에 기반한 포즈 추정 방법이 널리 사용되기 위해서는 다양한 동작에 대해서 포즈 추정을 할 수 있어야 한다. 그러나 다양체 학습에서 여러 동작을 표현하기 위해서는 각 동작마다 다양체를 필요로 한다. 차후에 여러 동작을 적은 수의 다양체로 추정할 수 있는 연구가 필요할 것으로 보인다.

참 고 문 헌

[1] E. Murphy-Chutorian and M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.31, no.4, pp.607-626, April 2009.

[2] C. Lee and A. Elgammal, "Simultaneous Inference of View and Body Pose using Torus Manifolds," *Proc. IEEE/IAPR International Conference on Pattern Recognition*, Hong kong, China, pp.489-494, August 2006.

[3] C. Lee and A. Elgammal, "Modeling View and Posture Manifolds for Tracking," *Proc. IEEE International Conference on Computer Vision*, Rio De Janeiro, Brazil, pp.1-8, October 2007.

[4] V. Balasubramanian and J. Ye, "Biased Manifold Learning: A Framework for Person-Independent Head Pose Estimation," *Proc. IEEE Computer Vision and Pattern Recognition*, Minneapolis, USA, pp.1-7, June 2007.

[5] D. Specht, "A General Regression Neural Network," *IEEE Trans. on Neural Networks*, vol.2, no.6, pp.568-576, November 1991.

[6] S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol.290, no.5500, pp.2323-2326, December 2000.

[7] Poser 7: http://my.smithmicro.com/dr/poser.html.

[8] Motion Golf 3-D Swing Analyzer: http://motiongolf.com.

허 동 철

2007년 세종대학교 컴퓨터공학과(학사)
2009년 고려대학교 컴퓨터학과(석사)
2009년~현재 고려대학교 컴퓨터학과 박사과정. 관심분야는 컴퓨터 시각, 패턴인식, 얼굴 표정 분석, 포즈 추정 등

이 성 환

정보과학회논문지 : 소프트웨어 및 응용 제 36 권 제 1 호 참조