

컬러 마커를 이용한 3차원 모션 정보의 검출 및 추적

신수미, 이철우
전남대학교 컴퓨터공학과

3D Motion Information Detection and Tracking Using Color Marker

Su-Mi Shin, Chil-Woo Lee
Dept. of Computer Engineering, ChonNam National university
E-mail : smshin@image.chonnam.ac.kr, leecw@chonnam.ac.kr

요 약

본 논문에서는 두 대의 카메라로부터 입력받은 인간의 신체와 같은 관절체의 움직임을 Color Marker의 색상 특성을 이용하여 3차원 공간 정보를 인식하는 방법에 관해 기술한다. 이 방법은 인체에 물리적인 장치를 하지 않고 단순히 영상정보만을 이용하여 3차원 정보를 구하였다. 보다 정확한 인체의 특징점을 구하기 위해 본 논문에서는 신체의 관절각에 칼라 마커를 부착하고 칼라 마커들을 세그먼트하여 관절각의 2차원 정보를 구한 다음 스테레오 기하(Stereo Geometry)를 이용하여 3차원 정보를 계산하고 가상 공간상에서 인간의 움직임을 추적하는 방법을 제안한다. 제안하는 방법은 제스처 인식과 3차원 Virtual Reality 인터페이스 시스템 구성 등에 사용 될 수 있다.

1. 서론

최근 인간이 멀티미디어와의 통신활동에서 중심적인 요소가 됨에 따라서 인간과 컴퓨터 사이의 원활한 상호작용의 필요성이 커지고 있다. 그 예로써 무선 가상현실 인터페이스, 지능형 제품, 보안, 감시등과 같은 애플리케이션이 있다. 이와 같은 시스템 구성을 위해서 컴퓨터 비전(Computer Vision) 방법을 이용하여 영상으로부터의 3차원 정보를 이용하여 인간 행동을 추적하고 분석하는 연구들이 많이 수행되고 있다[3].

기존의 3차원 정보를 얻는 방법 중에는 인간과 컴퓨터 사이에 물리적인 연결을 하거나 Optical Sensor 나 Magnetic Sensor를 이용하는 방법이 있다. 그러나 이러한 방법들은 고가의 장비가 필요하고, 많은 물리적 제약이 따르게 되므로 인간과 컴퓨터 사이의 원활한 상호작용의 관점에서 볼 때 자연스럽지 못하다. 또 다른 방법으로는 신체의 위치 정보와 skin tone을 이용하는 방법이 있다. 위의 방법은 물리적 연결이나 별도의 장치는 필요로 하지 않는다. 그러나 인간의 움직임에 많은 제약 사항이 따르게 되고 신체의 정확한

특징점을 구할 수 없기 때문에 제스처 인식의 전처리로 사용할 경우 다양한 동작을 인식하기에는 부적절하다. 그러므로 본 논문에서는 두 대의 카메라로부터 입력된 좌우 영상에 대해서 칼라 마커를 기반으로 하여 각 신체 특징점 쌍을 구하고, 3차원 영상을 획득하기 위해 스테레오 기하를 통하여 신체 특징점들을 3차원 가상 공간상에 나타내는 것을 목적으로 하였다.

본 시스템은 크게 좌우 영상에서의 마커의 2차원상의 위치를 구하는 부분과 2차원 정보로부터 3차원 정보를 계산하는 부분 등 최종적으로 3차원 공간상에 Display하는 부분으로 나누어진다.

첫 번째로, 각 마커의 색상을 미리 정의된 수식과 이를 계산하기 위한 의사 역행렬(Pseudo-inverse Matrix) 방법을 이용하여 모델링한다. 그리고 입력받은 두 영상에서 모델링된 값에 의해 각 마커의 영역만을 세그먼트하여 각 영상에서의 마커들의 무게중심(Center of Gravity)을 구한다. 두 카메라로부터 얻어진 각 영상에 대한 마커들의 무게중심 즉, 신체 특징

점 쌍을 이용하여 스테레오 기하 기반의 3차원 모션 파라미터를 구한다. 그리고 마지막으로 신체 특징점의 움직임에 대한 3차원 정보를 가상 공간상에서 시각적으로 쉽게 확인할 수 있도록 OpenGL을 이용하여 나타내었다. 그림 1에 전 시스템의 구성을 보인다.

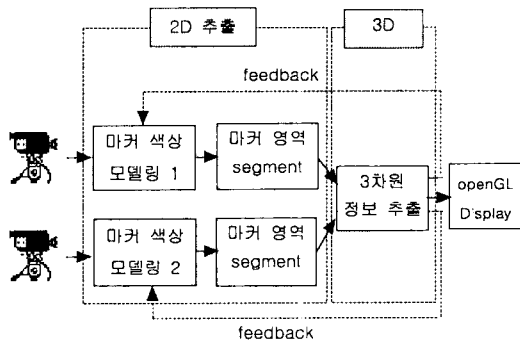


그림 1. 시스템 구성도

2. 컬러 마커의 2차원 정보 추출

마커의 2차원 정보를 추출하기 위해 본 논문에서는 신체 상반신(Upper Body)의 특징적인 머리, 좌우 어깨, 좌우 팔꿈치, 좌우 손목에 7가지 색상(주황, 하늘, 파랑, 연두, 자주, 노랑, 빨강)의 마커를 부착하였다. 그리고 두 카메라에서 각각의 영상을 획득하였다.

2.1. 마커들의 색상 파라메타 모델링

본 논문에서는 각 마커의 색을 정의하기 위해서 RGB Color Model을 사용하였다. 그림 2에서 확인할 수 있듯이 YUV Color Model은 조명에 덜 민감한 반면 다양한 색상을 구별하는 방법으로는 적합하지 않다. RGB Color Model은 RGB 값 자체에 색상 성분과 밝기 성분을 포함하고 있으므로 동일한 색상 특성을 가진 물체라도 조명의 영향에 따라 RGB값이 달라질 수 있다. 따라서 물체의 색상성분과 조명성분을 분리하여 다음과 같은 방법으로 정규화 하였다. 각 Pixel에서의 r, g, b의 값을 intensity i 의 2차 방정식으로 정의하는 식 (1)을 사용하였다[6].

$$\begin{aligned} \hat{r} &= R_2 i^2 + R_1 i \\ \hat{g} &= G_2 i^2 + G_1 i \\ \hat{b} &= B_2 i^2 + B_1 i \end{aligned} \quad \text{식 (1)}$$

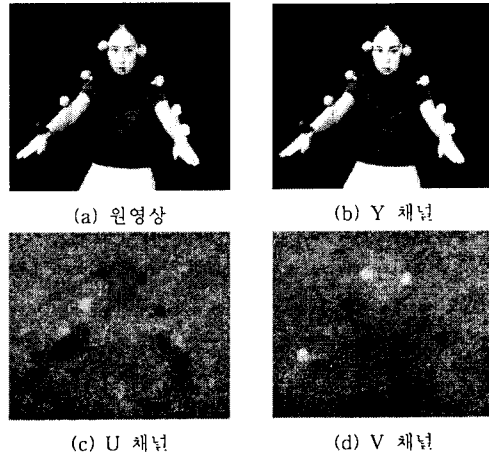


그림 2. YUV Color Model을 이용한 마커영역 검출

위의 식 (1)은 다음과 같은 행렬 식 (2)로 표현할 수 있다.

$$\begin{bmatrix} r_1 & g_1 & b_1 \\ \dots & \dots & \dots \\ r_m & g_m & b_m \end{bmatrix} = \begin{bmatrix} i_1^2 & i_1 \\ \dots & \dots \\ i_m^2 & i_m \end{bmatrix} \begin{bmatrix} R_2 & G_2 & B_2 \\ R_1 & G_1 & B_1 \end{bmatrix} \quad \text{식 (2)}$$

식 (1)을 이용하여 각 마커의 색상을 모델링 하기 위해서는 먼저 Training 이미지로부터 알 수 있는 $\hat{r}, \hat{g}, \hat{b}$ 값을 결정해야 한다. 여기서 $\hat{r}, \hat{g}, \hat{b}$ 값은 식 (2)에서의 $r_j, g_j, b_j, i_j (j=1, \dots, m)$ 값을 함축적으로 표현한다. $\hat{r}, \hat{g}, \hat{b}$ 값을 구하는 방법은 다음과 같다.

조명의 변화로 인하여 마커영역의 intensity가 다른 m개의 프레임에서 각 마커 영역만을 일정한 크기로 세그먼트 하였다. 이 영역의 모든 픽셀들의 r, g, b값과 intensity 값에 대한 평균을 다음 식 (3)와 같이 계산하여 r_j, g_j, b_j, i_j 를 구한다

$$\begin{aligned} r_j &= \frac{1}{n} \sum_{k=1}^n r_k, & g_j &= \frac{1}{n} \sum_{k=1}^n g_k \\ b_j &= \frac{1}{n} \sum_{k=1}^n b_k, & i_j &= \frac{1}{n} \sum_{k=1}^n i_k \end{aligned} \quad \text{식 (3)}$$

여기서, n 은 마커 영역의 모든 픽셀의 수이다.

이와 같이 m개의 r_j, g_j, b_j, i_j 값으로 Training 데이터가 주어지면, 각 마커의 컬러 파라메타 $R_1, R_2, G_1, G_2, B_1, B_2$ 를 다음과 같이 구할 수 있다.

식(2)는 m개의 선형방정식 $A_m x = c_m$ 과 같이 나타낼 수 있다.

여기서, A_m, x, c_m 은 다음과 같다.

$$A_m = \begin{bmatrix} i_1^2 & i_1 \\ \vdots & \vdots \\ i_m^2 & i_m \end{bmatrix}, x = \begin{bmatrix} R_2 & G_2 & B_2 \\ R_1 & G_1 & B_1 \end{bmatrix}, c_m = \begin{bmatrix} r_1 & g_1 & b_1 \\ \vdots & \vdots & \vdots \\ r_m & g_m & b_m \end{bmatrix}$$

계산하고자 하는 컬러 파라메타 행렬 x 를 구하기 위해서는 A^{-1} 을 구하여야 한다. 그러나 위에서 언급한 것과 같이 다양한 조명에서 색상을 Training했기 때문에 행렬 A 는 $m \times 2$ 행렬로써 A^{-1} 를 구할 수 없다. 즉, 연립선형방정식에 있어서 방정식의 수가 미지수의 수보다 크기 때문에, 방정식이 선형독립일 때 엄밀하게는 해가 존재하지 않는 경우와 같다.

따라서 A 의 의사 역행렬인 A^+ 을 구하여 행렬식(2)를 계산한다. 의사 역행렬은 평가함수 C 를 정의하고 이를 최소화 시켜서 대략적으로 A^{-1} 에 대응하는 값을 구하는 방법이다.

$$C = (Ax - c)^T (Ax - c) \quad \text{식 (4)}$$

x 에 대해서 미분을 취한 해는 식 (5)와 같다.

$$x = (A^T A)^{-1} A^T c = A^+ c \quad \text{식 (5)}$$

여기서, $A^+ = (A^T A)^{-1} A^T$ 이 이른바, 의사 역행렬(Pseudo-Inverse Matrix)이다.

위와 같이 의사 역행렬을 이용하여 위의 선형연립 방정식을 풀어 각 마커의 컬러 파라메타 $R_1, R_2, G_1, G_2, B_1, B_2$ 를 모델링 한다. 컬러 파라메타를 모델링 하기 위한 방법으로 SVD 방법을 사용하여도 유사한 결과를 얻는다.

2.2. 2차원 마커의 무게중심 추출

마커의 정확한 무게중심(Center of Gravity)를 찾는 과정은 신체 특징점의 움직임을 정확하게 추적하고, 또한 3차원 정보의 오차를 최소화하는데 중요한 역할을 한다.

무게중심을 구하기 위해서는 우선 각 마커의 영역을 세그먼트 하여야 한다. 이때 모델링 단계에서 정의된 식 (1)을 이용하여 원하는 마커 영역을 세그먼트한

다. 입력영상의 각 픽셀 (x, y) 의 intensity i 를 각 마커별로 모델링된 파라메타인 $R_1, R_2, G_1, G_2, B_1, B_2$ 와 함께 식 (1)에 적용하여 $\hat{r}, \hat{g}, \hat{b}$ 를 구한다. 이때, 마커의 색상 구분하는 에러 값은 다음과 같이 나타낸다.

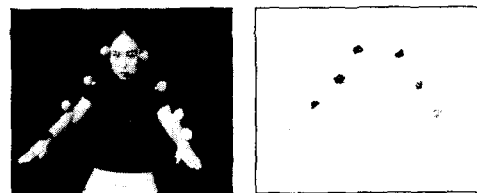
$$error = (r - \hat{r})^2 + (g - \hat{g})^2 + (b - \hat{b})^2 \quad \text{식 (6)}$$

여기서 r, g, b 는 입력 영상에서 각 픽셀의 RGB값이다. 만약 픽셀 (x, y) 의 intensity와 식 (6)에서 구해진 에러 값 모두 식 (7)의 조건을 만족하면 그 픽셀은 마커 영역이라고 정의한다.

$$\max < i < \min, \quad error < errorThreshold \quad \text{식 (7)}$$

여기서, \min 과 \max 는 intensity와 의 최소 값과 최대 값이고, $errorThreshold$ 는 마커 색상과 유사한 색을 구분하는 Threshold이다. 이렇게 구해진 칼라 특성에 의한 특징점 추출의 결과는 그림 3, 4와 같다.

그림 3-(b)에서 7가지 색상의 마커에 대한 영역이 정해지면 그림 3-(c)와 같이 각 색상별로 마커의 무게중심 값을 구할 수 있게 된다. 이렇게 구해진 좌우 영상의 각 마커의 무게중심 값들의 쌍을 이용하여 3차원 위치 정보를 계산 할 수 있다.



(a) 원영상

(b) 세그먼트 결과



(c) 마커 무게중심

그림 3. 마커 영역의 무게중심 추출

3. 3차원 마커의 위치 추적

두 대의 카메라로부터 입력된 각 영상에서 마커 영역을 기반으로 하는 각 신체 특징점의 무게중심의 쌍이 결정되면, 이를 이용하여 교정된 카메라의 정보와 스테레오 기하를 통하여 각 마커의 3차원 위치 정보를 구한다. 그리고 연속된 영상에서의 각 마커의 움직임의 3차원 정보를 추적한다.

그림 4에서 알 수 있듯이 (u, v) 는 기지패턴의 2차원 좌표이고, (X, Y, Z) 는 이에 대응하는 3차원 좌표이다.

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} = \begin{bmatrix} p_1^T & p_{14} \\ p_2^T & p_{24} \\ p_3^T & p_{34} \end{bmatrix}$$

3.1 카메라 칼리브레이션

스테레오 카메라를 이용하여 3차원 정보를 얻기 위해서는 먼저 2대의 카메라에 각각에 대한 카메라 칼리브레이션(Camera Calibration)이 필요하다. 카메라 칼리브레이션의 목적은 카메라의 내부적인 상태(내부 변수)와 카메라와 실세계 사이의 Rotation과 Translation(외부 변수)을 구하는 것이다. 본 논문에서는 Xu의 6개의 대응점 의한 카메라 칼리브레이션 방법을 사용하였다[7,8]:

위의 방법은 이미 알고 있는 6개의 대응점 즉, 칼리브레이션 패턴에서 동일하지 않은 평면에 위치한 6개의 점에 대한 2차원 영상의 좌표 (u, v) 와 이에 대응되는 3차원 좌표 (X, Y, Z) 로부터 내부변수와 외부변수를 포함하는 사영행렬 P 를 다음과 같은 수식에 의하여 계산할 수 있다.

$$s \tilde{m} = P \tilde{M} = A[R, t] \tilde{M} \quad \text{식 (8)}$$

여기서,

$$\tilde{m} = [u, v, 1]^T, \tilde{M} = [X, Y, Z, 1]^T$$

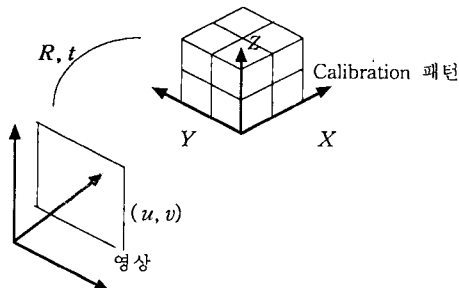


그림 4. 기지 패턴으로부터의 칼리브레이션

위의 식에 대하여 방정식 (9)를 얻을 수 있고, $B^T B$ 의 최소 고유치에 대응하는 고유벡터가 카메라의 사영행렬 p 이다.

$$Bp = 0 \quad \text{식 (9)}$$

여기서,

$$B = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1 X_1 & -u_1 Y_1 & -u_1 Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1 X_1 & -v_1 Y_1 & -v_1 Z_1 & -v_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_6 & Y_6 & Z_6 & 1 & 0 & 0 & 0 & 0 & -u_6 X_6 & -u_6 Y_6 & -u_6 Z_6 & -u_6 \\ 0 & 0 & 0 & 0 & X_6 & Y_6 & Z_6 & 1 & -v_6 X_6 & -v_6 Y_6 & -v_6 Z_6 & -v_6 \end{bmatrix}$$

$$p = [p_1^T, p_{14}, p_2^T, p_{24}, p_3^T, p_{34}]^T$$

이렇게 두 대의 카메라에 대한 사영행렬 P, P' 를 구하게 되면, 스테레오 비전에 의해 3차원 위치 정보를 계산할 수 있게 된다.

3.2 마커의 3차원 정보 추출

3차원 정보를 얻기 위한 마지막 단계로써 3.1의 카메라 칼리브레이션 과정에서 구한 두 카메라에 대한 사영행렬 P, P' 와 그림 5와 같은 스테레오 기하를 이용하여 3차원 위치 정보를 계산한다.

그림 5와 같이 좌우 영상에서 동일한 색상에 대한 마커의 무게중심인 대응점을 각각 $m = [u, v]^T$ 과

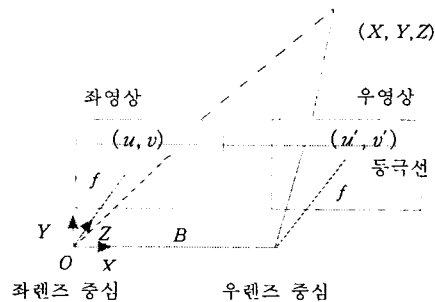


그림 5. 스테리오 기하

$m'=[u',v']^T$ 이라고 하면, 두 2차원 좌표에 대하여 식 (10)과 같은 식을 세울 수 있으며, 식 (11) 식을 유도할 수 있다. C 에 대한 의사역행렬 C^+ 을 이용하여 마커의 무게 중심에 대한 3차원 좌표 M 을 구할 수 있다.

$$s \tilde{m} = P \tilde{M}, \quad s' \tilde{m}' = P' \tilde{M} \quad \text{식 (10)}$$

$$CM = c \quad \text{식 (11)}$$

$$C = \begin{bmatrix} u\phi_{31} - \phi_{11} & u\phi_{32} - \phi_{12} & u\phi_{33} - \phi_{13} \\ u\phi_{31} - \phi_{21} & u\phi_{32} - \phi_{22} & u\phi_{33} - \phi_{23} \\ u'p'_{31} - \phi'_{11} & u'p'_{32} - \phi'_{12} & u'p'_{33} - \phi'_{13} \\ u'p'_{31} - \phi'_{21} & u'p'_{32} - \phi'_{22} & u'p'_{33} - \phi'_{23} \end{bmatrix}$$

$$c = [p_{14} - up_{34}, p_{24} - vp_{34}, p'_{14} - u'p'_{34}, p'_{24} - v'p'_{34}]$$

위의 방법으로 복원된 3차원 좌표의 각 (X, Y, Z) 축에 대한 오차는 다음 표 1 과 같다. 미리 정해진 21 개의 좌표에 대해 5번씩 3차원 복원을 수행하였다.

축	3차원 복원 오차
X	2.3 Cm
Y	3.9 Cm
Z	4.0 Cm

표 1. 각 축에 대한 3차원 복원 오차

위와 같이 구해진 각 영상에서의 7개의 마커들의 3차원 위치 정보를 가상공간 상에서 확인하기 위해서 그림 6, 7과 같이 openGL을 사용하여 나타내었다.

4. 실험 결과 및 향후 연구

본 논문에서는 인간 움직임의 3차원 정보를 추출하고 가상 공간상에서 나타내기 위한 7가지 색상의 컬러 마커 세그먼트 방법과 스테레오 카메라를 이용하는 방법을 제안하였다.

제안된 여러 가지 마커의 색상에 대한 세그먼테이션 방법의 정확도는 임의의 200개의 프레임에서 각각 마커의 무게중심을 계산하여 얻은 정보를 이용하여 확인한 결과 90%이상의 정확도를 보였다. 3차원 정보 또한 표 2에서 확인할 수 있듯이 각 마커의 3차원 복원률도 모두 90%이상의 정확도를 보였다. 그러므로 마커의 3차원 정보 추출에 대한 결과는 제스처 인식

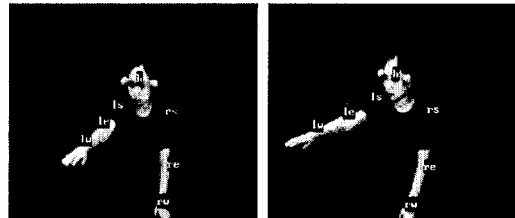
의 전처리로 쓰여도 좋을 만큼의 결과를 얻었다고 보여진다. 모든 마커의 움직임의 조합에 의해 나타나는 신체 동작에 대한 유사도는 3차원 공간상에서 확인한 결과 78%의 정확도를 보였다.

마커색상	신체 위치	3차원 정확도
주황색	머리	95.5 %
하늘색	오른쪽어깨	96.6 %
파랑색	왼쪽어깨	97.5 %
연두색	오른쪽팔꿈치	96.6 %
자주색	왼쪽 팔꿈치	95.8 %
노랑색	오른쪽 손목	93.3 %
빨강색	왼쪽	91.6 %

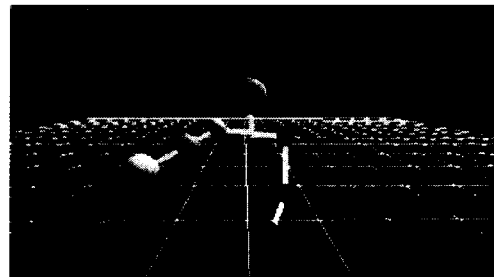
표 2. 각 마커에 대한 3차원 복원 정확도

그러나 동작도중 마커가 신체의 일부분에 의해 가려지면 그 마커에 대한 정보를 찾을 수 없게 된다. 이 문제를 해결하기 위하여 3대 이상의 다중 카메라를 이용하는 보다 안정적인 알고리즘을 개발할 계획이다.

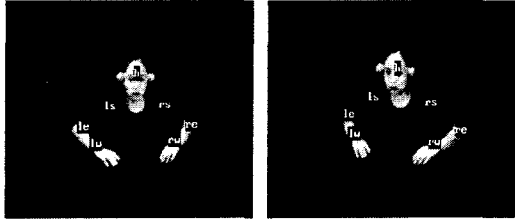
또한 현재는 상반신(Upper Body)의 신체 특징점에 대하여 실험하였으나, 향후에는 전신(Full Body)에 대한 특징점을 구하고, 모든 마커의 3차원 정보를 더 안정적으로 구하여 실시간 제스처 인식에 적용시킬 수 있는 방법을 연구할 계획이다.



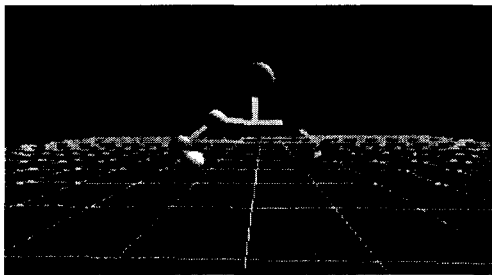
(a) 왼쪽 영상의 무게중심 (b) 오른쪽 영상의 무게중심 추출 결과



(c) 3차원 가상 공간에서 마커의 위치 그림 6. 신체 특징점 추출 결과 1



(a) 왼쪽 영상의 무게중심 추출 결과 (b) 오른쪽 영상의 무게중심 추출 결과



(c) 3차원 가상 공간에서 마커의 위치
그림 7. 신체 특징점 추출 결과 2

6th Korea- Japan Joint Workshop on Computer Vision Frontiers of Computer Vision 2000, 2000,2.
 [6] Y.Okamoto, R.Cipolla, H.Kazama, and Y. Kuno.; "Human interface system using qualitative visual motion interpretation.", IEICE, J76-D-II (8):1813-1821, 1993.
 [7] Xu Gang, 이철우 역, "3차원 비전"
 [8] Richard Hartley and Andrew Zisserman, "Multiple View Geometry in computer vision"
 [9] 오재용, 이철우; "3차원 시각정보를 이용한 VR 인터페이스 구현". 2001년 신호처리합동 학술대회 논문집

[참고문헌]

[1] Satoshi Yonemoto, Asuka Matsumoto, Daisaku Arita and Rin-ichiro Taniguchi Kyushu University; "A Real-Time Motion Capture System with Multiple Camera Fusion", IEEE ICIA, 1988.
 [2] Chistopher Wren, Ail Azarbayejani, Trevor Draeerl, and Alex Pentland. Pfinder; "Real-time tracking of the human body.", In Photonics East, SPIE Proceedings Vol. 2615 Bellingham, WA, 1995. SPIE.
 [3] Lee W. Campbell, David A. Becker, Ail Azarbayejani, Aaron F. Bobick, Alex Pentland.; "Invariant features for 3-D gesture recognition.", In Second International Workshop on Face and Gesture Recognition, Killington VT Oct., 1996
 [4] Daisaku Arita, Satoshi Yonemoto and Rin-ichiro Taniguchi; "Real-time Computer Vision on PC-cluster and Its Application to Real-time Motion Capture.", 2000 IEEE.
 [5] Chil-woo Lee, Yong-jae Lee; "Evaluation of Gesture Grade Using Sequential Motion Information on the Eigenspace", Proceeding The