

정면 투영 환경에서의 자동 칼라 보정에 의한 손 영역 추적 알고리즘

고은영^o 남양희

이화여자대학교 디지털미디어학부

virginia8002@hanmail.net^o, yanghee@ewha.ac.kr

Bare-Hand Tracking Based on Automatic Color Calibration in Front Projection Environment

Jane Koh^o Yanghee Nam

Division of Digital Media, Ewha Womans University

요 약

최근 대형 디스플레이 및 웨어러블 컴퓨터의 등장과 함께 키보드와 마우스를 사용하는 일반 데스크탑 환경에서 벗어난 컴퓨터와의 자연스러운 상호 작용 연구가 활발히 진행되고 있다. 본 논문에서는 스크린 정면에 놓인 프로젝터가 스크린과 그 위에 놓인 사용자의 손 위에 화면을 투영할 때, PC 카메라로 입력된 프레임 속에서 손의 영역을 인식하여 컴퓨터와 상호작용하게 하고자 한다. 이 경우에 투영된 빛이 사용자의 손 위에도 합쳐짐으로 인하여 피부의 고유색이 사라진다. 또한, 투영되는 화면이 사용자와 컴퓨터의 상호 작용에 따라 추정할 수 없이 변화에 따른 적응적 인식 방법이 필요하다. 따라서, 본 논문에서는 손을 인식하기에 앞서 스크린에 투영될 원본 이미지에 대해 칼라 보정을 수행하여 추정되는 카메라 입력 프레임으로 원본 값에 대한 카메라 반응 값을 측정 테이블에 저장한다. 이를 통해 원본 이미지에 대해 칼라 보정을 수행하고, 생성된 카메라 추정 프레임과 실제 카메라로 입력된 프레임 간 밝기 성분을 비교하여 손 영역을 검출한다. 실험 결과, 주변의 조명 상태나 프로젝터 및 카메라의 위치에 관계없이 안정적인 인식 결과를 보였다.

1. 서 론

최근 대형 디스플레이 및 웨어러블 컴퓨터의 등장과 함께 키보드와 마우스를 사용하는 일반 데스크탑 환경에서 벗어난 컴퓨터와의 자연스러운 상호 작용 연구가 활발히 진행되고 있다. 더욱이, 유비쿼터스 컴퓨팅 환경으로의 변화가 이루어지는 추세에 따라 사람이 특별한 장비를 착용하지 않은 채 자연스러운 행동 양식에 따라 특정한 작업을 수행할 수 있도록 하는 기술은 계속 요구되고 있다.

본 논문에서는 프로젝터에서 투영된 빛이 스크린을 비추고 사용자가 화면 위에서 취하는 동작을 통해 컴퓨터와 상호작용 할 수 있게 한다. 이 때 프로젝터는 스크린의 정면에 놓이게 되며 투영되는 빛이 더해진 사용자의 손이 포함된 스크린의 화면은 일반 PC 카메라를 통해 입력되어 분석된다. 이와 같은 환경에서는 정면에 놓인 프로젝터의 빛이 사람의 손에 합쳐짐으로 인해 고유의 피부색이 사라지는 문제점이 있다. 또한, 본 논문에서는 투영되는 화면이 사용자의 동작에 따른 컴퓨터와의 상호 작용에 의해 추정할 수 없이 동적으로 변화한다는 가정이 존재하므로 실시간 인식에 어려움이 있다. 위와 같은 인식 상의 어려움들을 해결하기 위하여 본 논문에서는 손을 인식하기에 앞서 스크린에 투영될 원본 이미지에 대해 칼라 보정을 수행하여 추정되는 카메라 입력 프레임을 생성한다. 이를 위해 우선 백색 영상을 투영하여 프레임 내의 밝기 오차 맵을 생성한 후 R,G,B 채널 별로 원본 값에 대한 카메라 반응 값을 측정 테이블에 저장한다. 이를 통해 원본 이미지에 대해 칼라 보정을 수행하고, 생성된 카메라 추정 프레임과 실제 카메라로 입력된 프레임 간 밝기 성분을 비교하여 손 영역을 검출한다.

2. 관련 연구

카메라를 통해 입력되는 프레임 내에 존재하는 스크린에 맺힌 화면의 좌표는 원본 이미지의 좌표와 다르므로 먼저 원본 이미지 좌표로의 기하학적 변환이 요구된다. 두 평면의 기하 정보가 다른 경우 한 쪽의 기하 정보를 다른 한 쪽의 기하정보로의 변환을 평면 사영 변환(Planar Projective Transformation)이라고 한다[1]. 기존의

SmartPresentations이라는 연구에서는 스크린에 비스듬히 맺히는 화면에 대해 기하학적으로 보정된 화면을 생성했다. 먼저 스크린에 맺힐 영상과 원본 영상 간의 좌표 관계를 구하기 위해 변환 행렬인 호모그래피(homography) 행렬을 구한 후, 투영할 이미지에 미리 역으로 왜곡을 취해 스크린에 투영함으로써 왜곡을 보정하였다[2].

칼라 보정 연구의 경우 Jaynes의 논문에서는 프로젝터에 의해 투영될 원본 칼라와 카메라를 통해 입력되는 칼라 값들 간에 나타나는 관계를 R,G,B 채널마다 4개씩의 샘플 대응 쌍을 이용해 비선형적 관계식을 도출하였다[3]. Hilario와 Cooperstock 역시 원본 칼라와 카메라를 통해 입력되는 칼라 값들 간의 대응 관계를 구하되, 이들 관계에 대한 소수의 샘플 대응 쌍을 이용하여 비선형 식을 도출함으로써 정확한 결과를 내는 것은 어렵다고 보고, 선형최소사승법(Linear Least Squares Method) 및 대응쌍 정보와 검은색의 오프셋 값을 저장하는 측정 테이블을 사용하였다[4]. Kale은 정면 투영 환경에서 카메라 프레임 내에 존재하는 손에 의해 발생한 그림자의 영역을 먼저 찾은 후 Epipolar 정보를 사용해 손의 영역을 검출하였다[5].

본 논문에서는 카메라 입력 화면에 대한 기하학적 보정 시 카메라 내부 파라미터 특성에 의해 발생한 방사 왜곡을 보정한 뒤 homography 행렬을 구하여 평면 사영 변환을 수행하였다. 또한, 칼라 보정 시 칼라 변환 정보 외에 프레임 내의 밝기 분포의 차이도 계산하여야 한다고 보고, 카메라 프레임 내의 밝기 오차 맵과 RGB 각 채널의 칼라의 변환 정보에 대한 측정 테이블을 생성한 후 실시간 칼라 보정에 적용하였다.

3. 자동 칼라 보정에 의한 손 영역 추적 알고리즘

본 논문에서는 손을 인식하기 위해 원본 이미지에 대한 칼라 보정을 수행하여 추정되는 카메라 입력 프레임을 생성한 후 이를 실제 카메라 입력 프레임과 비교하는 방법을 사용한다. 따라서, 실제 카메라로 입력된 프레임에 대해서 기하 보정을 수행하여 실제 카메라 프레임 내에 존재하는 스크린에 맺힌 투영 화면 영역을 원본 좌표와 동일하게

변환시켜야 하며, 실시간으로 변화하는 원본 이미지에 대해서는 칼라 보정을 수행한다. 아래의 그림 1.은 전체 구조도를 나타낸다.

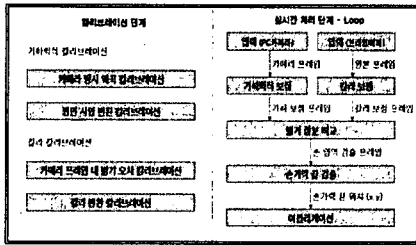


그림 1. 전체 구조도

3.1 기하 보정

3.1.1 카메라 방사 왜곡 보정

방사 왜곡(Radial Distortion)은 렌즈의 내부 파라미터 특성에 따라 발생하는 현상으로서 스테레오 카메라를 통해 체크판 무늬 이미지를 입력 받아 카메라의 내부 파라미터를 구해서 보정하는 방법이 가장 일반적으로 쓰이는 방법이다. 그러나, 본 논문에서는 하나의 카메라로 단일 영상에서 사용자가 마우스로 짚은 점들로부터 방사 왜곡 파라미터들을 구하는 방법을 사용한다[6]. 이는 사용자의 마우스 입력을 받아야 하는 불편함이 있지만, 별도의 캘리브레이션용 이미지 등이 필요하지 않다는 점에서 유용하다. 방사 왜곡 방정식은 식 (1)과 같다.

$$x_u = x_d + x_d \sum_{l=1}^{\infty} \kappa_l R_d^l, \quad y_u = y_d + y_d \sum_{l=1}^{\infty} \kappa_l R_d^l \quad \text{식 (1)}$$

x_u, y_u 는 왜곡이 보정되었을 때의 영상의 좌표, x_d, y_d 는 왜곡된 영상의 좌표, R_d 는 $x_d^2 + y_d^2$, l 은 직선들의 인덱스, κ_l 는 방사 왜곡 계수들을 의미한다. κ_l 들은 선들로 얻어지는 선형 방정식의 묶음을 풀어서 찾아낼 수 있다. 본 논문에서는 방사왜곡 계수를 2개까지만 구하는 경우를 위해서는 간략화 된 식 (2)를 사용한다.

$$\begin{pmatrix} (H_1 - G_1) \\ (H_2 - G_2) \end{pmatrix} \begin{pmatrix} \kappa_1 \\ \kappa_2 \end{pmatrix}^{-1} = \begin{pmatrix} G_2 & G_3 \\ G_3 & G_4 \end{pmatrix} \quad \text{식 (2)}$$

3.1.2 평면 사영 변환

방사 왜곡이 보정된 카메라 입력 프레임 내에 존재하는 스크린에 맺힌 화면의 영역은 평면 사영 변환에 의해 원본 이미지의 좌표와 동일하게 변환될 수 있다. x, y 는 카메라 프레임의 좌표이고, x', y' 가 원본 이미지의 좌표일 때, 둘 사이의 변환 관계는 최소 4개의 대응쌍 좌표들을 사용하여 식 (3)에서와 같이 3x3 호모그래피(homography) 행렬을 구함으로써 계산된다.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = h \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad \text{식 (3)}$$

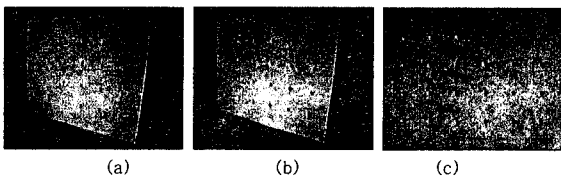


그림 2. 기하 보정 실험 단계별 결과 프레임

(a) 카메라 입력 프레임 (b) 방사 왜곡 보정 결과 프레임 (c) 평면 사영 변환 결과 프레임

3.2 칼라 보정

컴퓨터 화면 내의 임의의 색 값은 프로젝터에 의해 스크린에 투영된 후 카메라를 통해 입력받았을 때, 프로젝터의 휘도 특성, 실내의 조명 및 반사 환경, 카메라의 센서 특성 및 투영된 이웃 픽셀들 간의 상호 간섭 등으로 인하여 그 값이 변한다. 전처리 단계에서 미리 카메라 프레임 내의 밝기 분포와 원본 칼라의 색상 변환 정보들을 저장함으로써 임의의 원본 이미지에 대해 추정되는 카메라 프레임을 생성하기 위한 칼라 보정을 수행할 수 있다.

3.2.1 프레임 내부의 밝기 오차 맵 생성

스크린의 화면을 입력받은 카메라 프레임 내부의 밝기는 부위에 따라 차이가 발생한다. 따라서, 인식 전 단계에서 그림 3.에서와 같이 백색 영상을 투영하여 카메라로 입력받은 후 해당 프레임 내부에 따라 차이를 보이는 밝기의 분포에 대한 밝기 오차 맵을 생성한다. 각 픽셀 위치마다 획득되는 R,G,B 각 채널의 값을 최대값 255에 대해 감소된 비율을 산출하여 0.0에서부터 1.0 사이의 값으로 저장한다. 가장 밝은 지점을 1.0로 기준으로, 어두울수록 0.0에 가까운 값을 저장한다. 프레임 내부 밝기 오차 맵의 정보는 실시간 칼라 보정 시 실제 카메라 입력 프레임 내부의 밝기 분포를 표현하는 데 사용된다.

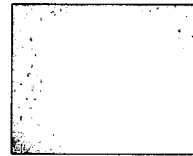


그림 3. 스크린에 투영하여 카메라로 입력받은 백색 영상

3.2.2 조명 상태에 적용적인 칼라 변환 값 획득

임의의 픽셀 값이 스크린에 투영된 후 카메라에 입력되면, 원본의 픽셀과는 다른 색상 및 채도 밝기 성분을 보인다. 또한, 실험 공간은 완전한 암실이 아니며 카메라의 센서 특성으로 인해 검은 색은 완전한 검은 색으로 입력되지 않으므로, 카메라로 검은 색을 입력했을 때의 오프셋 정보가 필요하다. 본 논문에서는 사용하는 RGB 색 공간은 한 채널이 256단계로 구성되며, 각 채널의 256단계의 원본 값에 대한 카메라 입력 값을 미리 저장한다. 이를 위해 빨강(R), 초록(G), 파랑(B) 세 가지 색 각각의 R 채널, G 채널, B 채널의 값을 256 단계로 투영하여 카메라로 입력된 값을 그림 4.에서 그래프로 나타내었다. 원본 값과 카메라로 입력된 값(Sensed Intensity)은 각 채널별 특업 테이블에 저장된다. (a)~(c)는 중간 조명, (d)~(f)는 어두운 조명 상태에서 얻어진 결과이다. 조명 상태에 따라 나타나는 검은 색의 오프셋 값 및 풀로링 곡선의 차이는 조명 상태에 적용적으로 칼라 보정을 가능하게 한다. 원본 이미지의 칼라 보정 시 R,G,B 각 채널마다 해당 특업 테이블을 통해 독립적으로 적용된다.

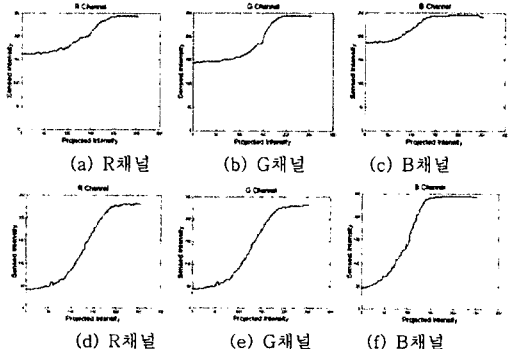


그림 4. RGB 채널별 원본 값에 대응하는 카메라 반응 값 풀로팅

3.3 손 영역 검출

실시간 손 영역의 검출은 칼라 보정을 거쳐 원본 이미지로부터 카메라 입력 프레임으로 추정된 프레임과 실제 카메라 입력 프레임 간의 밝기 성분을 차이를 구함으로서 이루어진다. 밝기 성분(Y)은 아래의 식 (4)을 사용하여 R, G, B 채널 값들의 조합으로 계산된다[7].

$$\text{밝기}(Y) = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad \text{식 (4)}$$

사람의 손과 팔 영역은 스크린에 비해 상대적으로 빛을 많이 흡수하여 위의 프레임들 간의 밝기 성분 차이가 검출된다. 이 때, 손의 영역이 아닌 노이즈 영역들을 분리하기 위해 검출된 영역들의 바운딩 박스들의 크기 및 위치 정보를 사용한다.

3.4 손가락 끝 검출

본 논문에서는 검출된 손 영역 중에서 손가락 끝부분의 위치를 검출함으로써 특정 어플리케이션과의 상호 작용에 필요한 위치 정보를 전달한다. 따라서, 앞서 검출된 손의 영역의 윤곽선 상에서 손가락의 위치를 검출한다. 검출 시에는 손이 들어온 방향에 따라 손가락 끝의 위치하는 윤곽선 상의 방향이 바뀌므로 먼저 손이 들어온 방향을 알아낸다. 그 후 손가락 위치 검출 대상이 되는 윤곽선의 영역 상에서 Segen과 Kumar가 제안한 k -curvature 알고리즘을 통해 검출한다[8]. 펼쳐진 손가락의 끝 부위는 손의 윤곽선 중에서 가장 굴곡이 급격한 부분이라고 보고, k 는 상수. $P(i) = (x(i), y(i))$ 는 윤곽선 점들의 리스트일 때, $[P(i-k), P(i)]$ 와 $[P(i), P(i+k)]$ 두 벡터 간의 최소 내적 값을 구함으로써 손가락 끝의 위치를 구한다. 본 논문에서는 실험을 통해 k 의 값을 30으로 결정되었다.

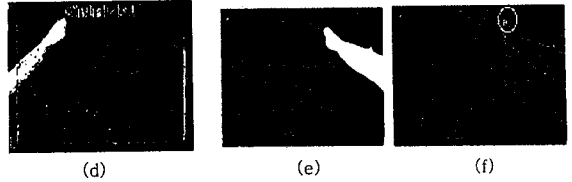


그림 6. (a)-(d) 손 영역 추적기 적용 예 (e),(f) 손가락 끝 검출

5. 결론 및 향후 연구

5.1 결론

본 연구에서는 일반적인 자연광 환경이 아닌 정면 투영 환경에서 화면과 사람의 영상을 카메라로 입력받아 인식하고자 하였다. 때문에, 기존에 많이 연구되었던 피부색의 정보를 사용할 수 없었으며, 원본 이미지를 카메라 입력 영상으로의 칼라 보정을 수행하여 실제 카메라 입력 프레임과의 비교를 수행하였다. 이를 통해 밝기 성분 비교를 통해 사용자 손의 영역을 분리해내었으며 이는 다양한 조명 환경에서도 적용적인 결과를 보인다. 이를 통해 실시간으로 원본 이미지의 칼라 정보를 안다는 가정 하에, 어떠한 동적 화면에 대해서도 카메라로 입력된 프레임 내의 손의 영역을 효과적으로 검출할 수 있게 되었다.

5.2 향후 연구

동일한 칼라라도 프레임 내 위치하는 부위에 따라 다른 색 정보를 나타낸다. 따라서, 프레임의 부위별 RGB의 카메라 반응 결과 값을 획득하여 적용함으로써 더욱 정확한 칼라 보정 결과를 보일 것을 기대한다.

6. 참고 문헌

- [1] R. Hartley & A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2000.
- [2] R. Sukthankar, R. G. Stockton, M. D. Mullin, "Smarter Presentations: Exploiting Homography in Camera-Projector Systems", *IEEE ICCV*, 2001.
- [3] C. Jaynes, S. Webb, R. M. Steele, "Camera-Based Detection and Removal of Shadows from Interactive Multiprojector Displays", *IEEE TOVCG*, 2004.
- [4] M. N. Hilario, J. R. Cooperstock, "Occlusion Detection for Front-Projected Interactive Displays", *International Conference on Pervasive Computing*, 2004.
- [5] A. Kale, K. Kwan, C. Jaynes, "Epipolar Constrained User Pushbutton Selection in Projected Interfaces", *IEEE CVPRW*, 2004.
- [6] S. B. Kang, "Semiautomatic Methods for Recovering Radial Distortion Parameters from A Single Image", *Cambridge Research Laboratory, Technical Report*, 1997.
- [7] D. Hearn & M. P. Baker, Computer Graphics, Prentice Hall, 1997.
- [8] J. Segen & S. Kumar, "Human-Computer Interaction using Gesture Recognition and 3D Hand Tracking", *IEEE ICIP*, 1998.

4. 구현 및 결과

4.1 구현 환경

PC 사양	Windows XP, P4 1.7Ghz, 512M RAM, NVIDIA GeForce FX 5600XT 비디오 가속기
프로젝터	InFocus사 DLP-650
PC 카메라	Fire-1400 CCD 센서, IEEE 1394 포트
실험 영상	320X256 size, 30f/s
구현 수단	Visual C++ 6.0, OpenCV 라이브러리, DirectX 8.0

4.2 실험 결과

본 논문의 손 인식을 일반 PC에서 이뤄지는 게임 프로그램에 적용하여, 스크린에 프로젝터를 통해 투영하고 카메라를 통해 화면 영역을 입력받아 사용자의 손으로 직접 상호작용이 이뤄지는 것을 실험하였다. 본 게임 프로그램은 DirectX 8.0을 사용하여 구현되었으며 그림 6.에서는 메뉴를 선택하는 장면이 보여진다. (a)의 원본 화면에 대해 칼라 보정을 수행한 결과가 (c)다. 이 카메라입력 추정프레임과 실제 카메라입력 프레임 (b) 간의 밝기 성분 차이를 통해 손 영역을 추출한 결과가 (d)이다. 이는 중간 조명 상태에서 실험한 결과로서 본 실험 환경에서 밝기 성분 차이의 임계값은 40.0이다. (e),(f)의 경우 검출된 손 영역의 윤곽선 상에서 손가락 끝의 위치를 찾은 실험 결과이다.

