

스테레오 영상의 깊이정보를 이용한 3차원 얼굴추출방법

The Method of face detection using Disparity informations

*김세영¹, #김세영², 김경호², 김진대², 조지승²

*S.Y. Kim¹, #S.Y. Kim(pionir76@gmail.com)², K.H.Kim(robodr@dmi.re.kr)²,
J.D.Kim(jdkim@dmi.re.kr)², C.S.Cho(okrobo@dmi.re.kr)²

¹미디어네이처, ²대구기계부품 연구원, ²대구기계부품 연구원, ²대구기계부품 연구원

Key words : Stereo vision, epipolar geometry, 3d reconstruction, face-detection

1. 서론

최근 로봇 제어 기술의 발달과 더불어 인간의 사고나 촉각, 시각을 대신할 수 있는 다양한 센서 개발을 위한 많은 연구들이 수행되고 있다. 대표적으로 스테레오 비전(Stereo vision)은 실제 사람의 시각 모델을 그대로 반영한 기술로써 양안을 통해 투영되는 피사체의 시각차를 사람의 지각능력에 의해 깊이정보로 변환되는 과정을 2대의 비전 카메라를 통해 구현하는 기술이다. 본 논문에서는 기존 2차원 영상정보를 기반으로 연구되어 왔던 인체 얼굴 추적(Face detect)기술에 스테레오 영상으로부터 얻어진 변이 정보(Disparity Information)를 이용하여 배경과 전경을 분리하고 깊이정보의 형상분석(Profile)방법을 통해 인체의 얼굴영역을 추출하는 방법을 제안한다. 본 논문에서 제안하고 있는 스테레오 비전 기술을 이용한 얼굴 추출방법은 다음과 같은 단계를 수행한다.

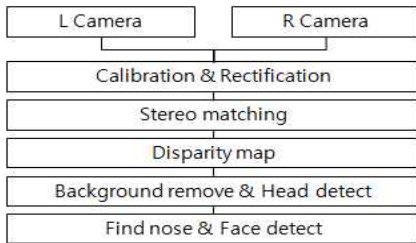


Fig. 1 The flow of face detect

2. 깊이영상 획득

일반적으로 스테레오 카메라는 평행 시각 스테레오 카메라 모델과 교차 시각 스테레오 카메라 모델로 나누어진다. 평행 시각 카메라 모델은 카메

라의 두 광축(Optical axis)이 평행이고, 교차 시각 카메라 모델은 두축이 고정점(Fixation point)에서 만나는 구조를 갖는다. 본 논문에서는 기하학적으로 해석이 용이한 평행 시각 스테레오 모델로 가정하여 변이 정보를 획득하고 피사체의 깊이맵(Depth map)을 생성한다.

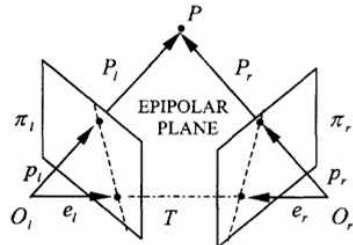


Fig. 2 The epipolar geometry

변이 영상은 양안 카메라로부터 획득되어진 좌, 우측 영상의 수평변이(Horizontal disparity)에 의한 화소별 대응점의 거리를 영상으로 표현한 것이다. 정확하고 잡음이 최소화된 변이 정보를 구하기 위해서는 카메라의 시각 오차를 보상하는 영상보정(Camera calibration) 및 좌, 우측 영상의 화소별 정합점을 찾는 스테레오 정합(Stereo Matching)과정이 선행되어야 하며 비교적 정확한 결과를 산출하지만 연산이 복잡하고 오랜 수행시간을 요구하는 전역정합(Global matching)과 연산이 간단하고 빠른 처리가 가능한 영역기반 정합으로 분류된다. 본 논문에서는 얼굴과 배경의 분리에 이용되는 변이영상의 빠른 생성을 위해 영역기반 스테레오 정합 알고리즘중 에피폴라선(Epipolar line)상에 있는 대응점의 이웃화소의

합의 차를 계산하여 최소화 되는 화소와 정합점을 찾아 내는 SAD 방법을 이용하였다.

$$\sum_{(u,v) \in W} I_1(u,v) - I_2(x+u,y+v) \quad (2.1)$$

3. 배경 분리

스테레오 영상으로부터 구해진 초기 변이정보의 거리정보를 이용하면 영상에서 사람과 배경을 쉽게 구별이 가능하다. 본 논문에서 이용하고 있는 스테레오 카메라 구조상 렌즈 중심을 포함하는 에피폴라(Epipolar line)상에 있는 수평 불균형(disparity)에 의한 차이값을 히스토그램으로 나타내면 그림 X와 같으며, 불균형 수치가 최소화 되는 지점을 임계값으로 선정함으로써 영상으로부터 배경을 제거하였다.

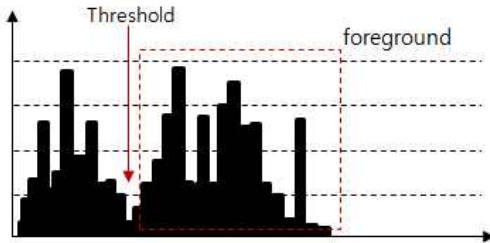


Fig. 3 Back ground remove

식(3.1)의 $g(x,y)$ 는 임계값을 기준으로 이진화된 영상이며, T는 임계값, $d(x,y)$ 는 변이 영상을 나타낸다.

$$g(x,y) = \begin{cases} 1 & \text{if } d(x,y) > T \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

4. 얼굴 검출

일반적인 사람의 얼굴 중 가장 높은 부위는 콧등에서 코끝으로 이어지는 영역이고 보통 얼굴의 가장 중심점에 위치하고 있다. 또한 사람의 두상과 몸영역의 구분은 다른 신체부위에 비해 비교적 좁은 영역을 차지하는 목 영역에 의해 구분될 수 있다. 본 논문에서는 분리된 전경 영역의 이진화된 영상 $g(x,y)$ 를 한 라인씩 스캔 하여 영역의 합이 최소화 되는 지점을 인체의 목 영역으로 판정하였다. 가장 큰 값을 갖는 두영역을 p, q라고 하면, p와 q사이의 각 x에 대한 영역은 식(4.1)과 같이

표현된다.

$$l(y) = \sum_{x=0}^{M-1} g(x,y), \quad y = p+1, \dots, q-1 \quad (4.1)$$

식(4.1)의 M은 전경 영상에서의 가로 크기를 나타내며 최소값을 가지는 목 부분을 구할 수 있다. 얼굴에서의 코 영역을 구하기 위해 앞서 구해진 목영역을 시작지점으로 데이터의 합이 0이 되는 지점까지 사전에 생성된 깊이 영상(Depth map)의 단면 프로파일을 분석하여 가장 높은 영역이 발생하는 곳을 얼굴의 중심으로 선정하였으며 중심점을 기준으로 목까지의 화소거리를 얼굴 전체영역으로 산정하였다.

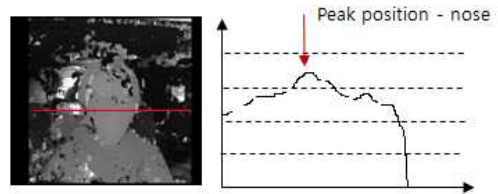


Fig. 4 Detect face center

5. 후기

본 논문에서는 양안식 스테레오 카메라 구조와 이를 통해 생성된 깊이차를 이용하여 영상으로부터 효과적으로 전경과 배경을 분리하였다. 또한 사람얼굴이 포함된 전경영역을 기준으로 인체의 목 위치를 검출함으로써 배경에 강인하고 실시간 얼굴 검출이 가능한 방안을 제안하였다. 그러나 단일 영상에 복수의 얼굴추출에 취약함을 보이므로 이는 지속적인 연구를 통해 보완이 가능할 것으로 기대한다.

참고문헌

1. R. Hartly, "Stereo from uncalibration camera" in European Conference on ComputerVision, 1992.
2. M. Hannah. Computer Matching of Areas in Stereo Images. PhD thesis, Stanford University, 1974
3. G. Yang and T. S. Huang, "Human face detection in complex background", vol. 27,1994