

사람 얼굴 영상 획득을 위한 영역기반 스테레오 매칭 기법

정창성⁰, 유재곤⁺, 황치정
충남대학교 컴퓨터학과
대덕대학 멀티미디어계열⁺
csjung@ipi.cnu.ac.kr

Area-based Stereo Matching Method for Extracting a Human Face Image

Chang-Sung Jung⁰, Chae-Gon Yoo, Chi-Jung Hwang
Dept. of Computer Science, Chungnam National University
Multimedia Department, Dae-Dok College⁺

요약

본 논문에서는 스테레오 영상에서 두 영상간의 깊이(Depth) 정보를 이용해서 영역을 기반으로 한 영상 매칭을 수행한 후 사람의 얼굴 영역을 결정한다. 영역 결정은 설정된 탐색 윈도우에 의해 좌, 우 영상간의 MSE(Minimum Square Error)를 검색하는 알고리즘을 제안한다. 매칭 과정에서 발생하는 영역 오류를 보정하기 위해서 타원 마스크를 이용하는 방법을 제안하며 실행 시간을 줄이기 위하여 2D 도메인에서 쿼드트리(Quad-tree) 윈도우를 사용한다.

1. 서론

컴퓨터 비전의 궁극적 목표는 인간 시각 시스템의 기능을 컴퓨터로 구현하는데 있으며 이를 구현하기 위해서는 입력된 영상 정보로부터 3차원 물체를 찾아내거나 인식하는 것이 기본적인 과제라고 할 수 있다. 2차원 영상으로부터 3차원 정보를 추출하는 과정은 매우 중요한 단계로서 하나의 카메라를 이용하는 단안 시법과 두 개의 카메라를 이용하는 양안시법이 있는데 후자를 스테레오 비전이라고 한다. 열악한 텍스처 데이터, 반복 패턴, 폐색(Occlusion) 등과 같은 문제점들 때문에 빌딩이나 탑과 같이 상대적으로 규칙적인 특징을 가지고 있는 대상에 대해서는 스테레오 매칭을 이용한 3D 재구성 등의 기법이 용이하게 적용되지만, 너무나 사람과 같이 랜덤한 물체에 대해서는 많은 어려움을 가지고 있다. 또한 영상처리의 적용에 있어서 이미지 데이터 자체의 애매함 또는 잡음, 처리 과정에서 발생하는 모호성, 그리고 인식과 해석 단계에서의 불확실한 지식 등으로 인해 많은 계산 비용에도 불구하고 좋은 결과를 보이지 못하고 있는 것이 현실이다. 본 논문에서는 스테레오 영상에서 두 영상간에 깊이 정보를 이용해서 영역을 기반으로 영상 매칭을 수행한 후 사람의 얼굴 영역을 결정한다. 영역 결정은 설정된 탐색 윈도우에 의해 좌, 우 영상간의 MSE를 검색하는 알고리즘을 제안한다. 이 과정에서는 실행 시간을 줄이기 위하여 쿼드트리 윈도우를 적용한다. 반복 패턴 등에 의해서 매칭 과정에서 발생하는 영역 오류를 보정하기 위해서 타원 마스크를 이용하는 방법을 제안한다.

2. 기존 연구

좌, 우 두 개의 카메라로부터 얻은 영상에서 3차원 상의 좌표를 구할 수 있으나, 이때 왼쪽 영상의 특정한 점이 오른쪽 영상의 어느 점에 대응되는지를 결정해야 되는 문제점이 발생한다. 스테레오 매칭은 좌우 영상간의 일치점을 찾아서 이 점들간의 차이를 구하는 과정으로써 3차원 정보를 획득하는데 있어서 가장 중요한 과정이라고 할 수 있다. 스테레오 매칭을 위한 방법으로는 영역기반 방법과 특징기반 방법 등이 있다.

2.1 영역기반 방법

영역기반의 스테레오 매칭방법은 매칭 기본 단위 표면 전체를 매칭하는 방법으로 이미지 밝기 자체를 기본 단위로 한다. 이 방법은 완전한 양안차 맵을 한번에 얻는 반면 요구되어지는 계산량이 매우 높다. 영역기반 방법의 일반적인 처리 방법은 최대 상관도(Cross-correlation)[1]나 최소 SSD(Sum of squared difference)[2]를 이용하는 것이다. 이것은 국부적으로 작은 영역을 매칭시킴으로써 조밀한 변이도를 구할 수 있지만, 매칭시키는 영역내의 모든 지점의 변이가 서로 유사하다는 가정을 하여야 하므로 경계부분에서 오차가 발생하며, 반복되는 무늬가 있는 경우 매칭의 정확도가 떨어지는 단점이 있다. 또한 상관관계를 계산하고 이를 중에서 최대화되는 영역을 찾는 과정은 설정된 전체 영역에서 처리되어야 하므로 많은 계산량이 요구되어 전체적인 처리속도가 떨어진다.

2.2 특징기반 방법

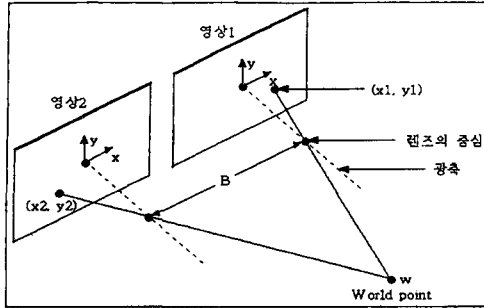
특징기반의 스테레오 매칭방법은 밝기 정보의 이미지로부터 매칭을 위한 기본 단위 집합을 추출하고, 만족할 만한 기본 단위를 가지고 매칭을 수행하는 방법이다. 양안차를 나타내는 결과가 이산적으로 나타나기 때문에 보간(Interpolation)이 필수적이고 기본적인 이미지 처리에 따라 특징이 상반되게 나타나는 특징이 있다. 이 방법에서는 먼저 영상특징을 추출하고 추출된 영상 특징 공간에서 정합시키는 것으로 처리속도를 향상시킬 수 있으나 영상의 전체 영역에 대한 조밀한 변이 추정은 불가능하며, 이를 위해서는 보간 과정을 거쳐야 한다. 스테레오 매칭을 위해 추출되는 영상 특징으로는 필터링된 영상의 경계점, 연결성을 가지는 윤곽선, 분할된 경계선, 모서리, 선소 등이 있다.[3]

3. 제안 방법

3.1 영역 기반 스테레오 매칭

스테레오 영상화 기법은 물체에 대해 두 개의 분리된 영상을 얻는 것이다. [그림 3-1]에서 두 렌즈 중심

간의 거리를 베이스 라인(Base line)이라고 한다.



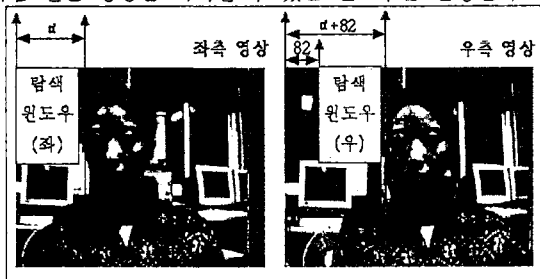
[그림 3-1] 스테레오 영상 획득 모델 (Fu, Gonzalez, and Lee[1987])

영상점 (x_1, y_1) 과 (x_2, y_2) 를 갖는 점 w 의 좌표 (X, Y, Z) 를 찾는 것이 스테레오 매칭의 목적이다. 카메라에 의한 두 좌표계는 각각의 원점이 다를 뿐 완전히 일직선상에 존재하게 된다. 카메라와 실좌표계를 일치시키면, 영상의 xy 평면은 실좌표계의 XY 평면과 일직선상에 있다. 그러므로 w 의 Z 좌표는 정확히 두 카메라 좌표계와 같다. 초점거리(Focal length), 베이스 라인, 사람 얼굴의 3차원 깊이의 관계를 기반으로 영상1에서의 사람 얼굴이 영상2에서는 어디에 위치하는가를 도출하면 [식 3-1]과 같다. 이를 이용해서 3D 공간상의 사람 얼굴 영역을 결정한다.

$$x_2 = x_1 + (\lambda - Z) / (\lambda B) \quad \text{[식 3-1]}$$

where, λ : 초점거리, B : 베이스 라인
 x_1, x_2 : 왼쪽/오른쪽 영상에서의 위치
 Z : 사람 얼굴의 3차원 깊이

두 영상간 픽셀의 차이와 실제 거리(mm)간의 관계를 알 수 없기 때문에, 실험 환경에서 사용된 거리를 사용하여 픽셀과 실제 거리와의 관계를 Calibration하여 도출한다. 본 논문에서는 700mm 전방에 있는 사람 얼굴을 매칭 대상으로 하고 있으며, Calibration 결과 두 영상에서의 사람 얼굴이 위치 차이는 82Pixel이다. 매칭을 위해서 사람 얼굴 영상을 커버할 수 있는 윈도우를 설정한다.



[그림 3-2] 영역기반 스테레오 매칭 .

좌측 영상을 기준으로 위치 차이 정보를 이용해서 우측 사람 얼굴 영상을 탐색한다. 이 때, 비교 방법은 [그림 3-2]와 같이 영상에 탐색 윈도우를 설정하고, 픽셀간의 MSE를 계산하여 최소 값의 위치를 우측 영상에서의 사람 얼굴 위치로 결정하도록 한다. 영역기반 매칭시 설정된 영역은 배경을 포함하기 때문에, 오차를 발생하지만 중앙 영역의 사람 얼굴이 좌우 영상간에 더욱 유사한 값을 차지하기 때문에, 직사각형 윈도우를 설정해도 된다.

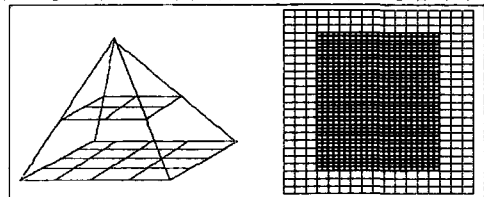
MSE는 [식 3-2]와 같이 계산되어진다.

$$MSE = \sqrt{\sum_{i=x}^{x+20} \sum_{j=y}^{y+10} (I(k, i) - Ir(j, i))^2}$$

where

$I(k, i)$: 왼쪽 영상에서의 픽셀의 밝기
 $Ir(j, i)$: 오른쪽 영상에서의 픽셀의 밝기
 [식 3-2] Minimum Square Error 계산식

영역기반 매칭 단계에서는 영상 비교의 속도를 높이기 위하여 쿼드트리 피라미드가 사용된다. 쿼드트리는 정보의 조밀여부에 따라 분할하는 방법이다.



[a] 쿼드트리 피라미드 [b] 쿼드트리 윈도우 [그림 3-3] 쿼드 트리

쿼드트리의 레벨은 요구되는 실행 시간에 따라서 조정된다. 각 영상의 서브 블록은 그레이 스케일 영상의 픽셀 밝기 상관계수를 이용하여 유사도를 측정한다. 본 논문에서 설정된 윈도우에서 가장자리 부분([그림 3-3] [b]의 가늘선 영역)은 사람 얼굴의 배경을 포함하므로 유사도 측정에 있어서 오류를 포함할 수 있다. 반대로 윈도우의 가운데 부분([그림 3-3] [b]의 굵은선 영역)은 전반적으로 중요도가 높다. 이와 같은 특성을 쿼드트리에 적용하여 얻은 [그림 3-3] [b]의 윈도우를 사용하여 유사도를 측정하는데 합리적으로 적용되도록 한다. 그 결과 [식 3-2]는 다음 [식 3-3]과 같이 변형되어 적용된다.

$$MSE = \sqrt{\sum_{i=-10}^{x+20} \sum_{j=-5}^{y+10} (I(k, i) - Ir(j, i))^2}$$

$i=-10$ $j=-10$ where, 윈도우 가장자리
 $i=-5$ $j=-5$ where, 윈도우 가운데 영역

[식 3-3] 개선된 MSE 계산식

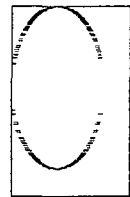
3.2 영역 보정

영역기반 매칭 단계에서 설정된 탐색 윈도우는 사람 얼굴이 큰 영역을 포함하기 때문에 대략적인 윤곽을 얻을 수 있지만 배경 영역을 포함하기 때문에 오차가 발생하게 된다. 이를 보정하기 위해 본 논문에서는 얼굴 모양의 타원 마스크를 영역기반 매칭에서 얻은 영상에 적용한다. [식 3-4]의 타원 방정식을 이용하여 [그림 3-4]와 같은 타원을 생성할 수 있다.

$$X^2/a^2 + Y^2/b^2 = 1$$

$$(c^2 = a^2 - b^2)$$

where
 초점: $F(c, 0), F'(-c, 0)$
 장축의 길이: $2a$
 단축의 길이: $2b$



[식 3-4] 타원 방정식 [그림 3-4] 타원 마스크

흰 영역은 0 그렇지 않은 곳은 1의 마스크를 갖는다. 생성된 타원은 설정된 탐색 윈도우보다 적은 영역을 형성하게 된다. 얼굴 영상 추적시 머리카락이나 턱에서 발생할 수 있는 오류를 줄이기 위하여 이와 같은 방법을 적용한다. 영상을 가로축과 세로축으로 이동하면서 [그림 3-4]의 타원 마스크를 씌운 다음 [식 3-5]에 의해서 표준편차 S를 구한다.

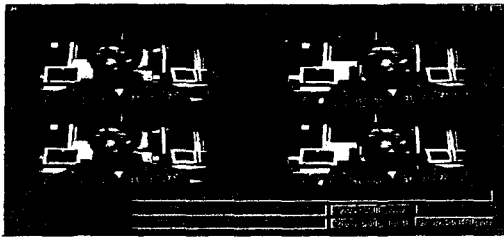
$$S = \sqrt{((I_0 - M)^2 + (I_1 - M)^2 + \dots + (I_n - M)^2) / N}$$

where

- I_0, I_1, \dots, I_n : 마스크가 1인 픽셀의 밝기
 - M : 마스크가 1인 픽셀들의 평균 픽셀의 밝기
 - N : 마스크가 1인 픽셀 개수
- [식 3-5] 표준 편차

그 결과 표준편차가 가장 작은 곳이 얼굴자 하는 영상의 X축, Y축 좌표가 된다. 표준편차가 작다는 것은 타원 영역 픽셀들간의 밝기 차이가 작다는 것을 의미한다. 그러므로 얼굴 영역에서의 표준편차가 가장 작게 되고 배경을 포함하게 되면 그 값은 커지게 된다.

4. 실험 결과



[그림 4-1] 스테레오 영상 캡처 프로그램

스테레오 영상은 640 x 480의 해상도를 갖는다. 스테레오 영상을 얻기 위해서 SAMSUNG MPC-C10 카메라를 사용한다. 이 모델은 기존의 입력장치와 달리 별도의 캡처보드 없이 직접 컴퓨터에 연결되어 사용될 수 있다.



[그림 4-2] 얼굴 영상 획득 프로그램

4.1 실험 예

[그림 4-3]은 스테레오 영상의 좌, 우측 영상이다. 두 영상에서 얼굴 영상이 추출 목표 대상이다.



[그림 4-3] 스테레오 영상

영역 기반 매칭 단계에서 사람 얼굴을 포함하는 대략적

인 윤곽을 얻을 수 있으며 그 결과 [그림 4-4] [a]와 같은 결과 영상이 획득된다.



[a]영역매칭 [b]상하보정 [c]좌우보정 [d]최종영역
[그림 4-4] 영역기반 매칭 및 영역 보정

다음 단계에서는 영역 오류를 보정하기 위해서 타원 마스크를 적용한다. 그 결과 위치를 보정한 [그림 4-4] [d]와 같은 목표 영상을 얻을 수 있다.

5. 결론 및 향후연구

실험 결과에서는 제안 방법이 목표영상을 렌더링 백그라운드로부터 분리하고 있음을 보이고 있다. 블루스크린과 같은 특정한 배경을 설정하지 않고도 사람의 얼굴 영상을 추출하는데 성공하였다. 사람 얼굴의 경우에는 경계선에서 머리카락의 모양과 색상 등이 불규칙하여 경계선 결정에서 부분적으로 오차가 발생하지만, 현재 결과로서 취득된 사람 얼굴 영역 영상의 경우로도 다양한 응용이 가능하다.

스테레오 매칭 방법으로 얻어진 영상을 이용하면 영상합성이나, 화상 전송 그리고 게임 시스템에서의 사용자 인식, 보안 시스템, 로봇틱스와 같은 컴퓨터 비전 분야에 적용될 수 있다.

향후에는, 영역 기반 방법과 특징 기반 방법을 함께 사용하여 사람 얼굴 영역 추출의 정확도를 높이는 방법에 대한 연구가 필요하며, 영역을 보정하기 위한 다양한 연구가 수행되어야 한다. 실험에서 발생한 불규칙한 모양과 색상을 가진 머리카락 부분의 경계선 처리 문제를 해결할 수 있는 방안에 대한 연구도 수행되어야 한다.

참고 문헌

- [1] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Addison-Wesley, 1993.
- [2] M. Okutomi, T. Kanade, "A Multiple-BaseLine Stereo," IEEE Trans. Pattern Anal. Machine Intell., Vol. 15, No. 4, pp. 353-363, 1993.
- [3] W. Eric, L. Grimson, "Computational Experiments with a Feature Based Stereo Algorithm", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-7, No. 1, pp. 17-33, 1985.
- [4] Luc Robert, "Camera Calibration without Feature Extraction", Computer Vision and Image Understanding, Vol 63, No. 2, pp. 314-325, 1996.
- [5] Davi Geiger, Bruce Ladendorf, Alan Yuille, "Occlusions and Binocular Stereo", International Journal of Computer Vision, Vol 14, pp. 211-226, 1995.
- [6] Ali Reza Mirhosseini, Hong Yan, "Human Face Image Recognition: An Evidence Aggregation Approach", Computer Vision and Image Understanding, Vol. 71, No. 2, August, pp 213-230, 1998.