

## 영역 분할을 통한 스테레오 정합 기법

\*안 재 우, \*유 지 상

\*광운대학교 전자공학과

\*skai9@kw.ac.kr, \*jsyoo@kw.ac.kr

### Stereo Matching Algorithm by using Segmentation

\*Jaewoo An, \*Jisang Yoo

Kwangwoon University

#### 요 약

본 논문에서는 인물 위주의 스테레오 영상으로부터 깊이 정보를 추출하기 위한 영역분할 스테레오 정합 기법을 제안한다. 제안한 기법에서는 두 대의 스테레오 카메라로부터 획득된 영상에서 임계값을 이용하여 전경을 제외한 배경을 먼저 제거하고, 배경이 제거된 영상으로부터 초기 변이지도(disparity map)와 R, G, B, white 4개의 색상 성분으로 분할한 영상을 생성하게 된다. 각 색상 정보로 분할된 영상의 경계(edge) 성분을 추출하고, 추출된 경계에서 정합 창을 이용하여 변이를 추정하고 각 색상 정보의 변이지도를 적절히 조합하여 최종 변이지도를 생성하게 된다. 실험 결과 제안한 기법이 기존의 영역기반(window based) 정합기법 등보다 인물 위주의 스테레오 영상에서 더 우수한 성능을 가지는 것을 확인하였다.

#### 1. 서론

최근 공간적으로 떨어져 있는 장소 또는 가상의 장소에서 사용자간의 현장감을 제공하는 tele-presence 시스템 등에서 자연감을 증대하기 위하여 eye contacted 영상 등을 생성하기 위한 가상시점 영상 생성에 대한 연구가 활발하게 진행되고 있다. 이러한 가상 시점 영상을 생성하기 위해서는 일반적으로 두 대의 카메라로 구성된 스테레오 카메라를 사용하여 먼저 깊이 정보를 추출하게 된다. 깊이 정보를 추출하기 위해서는 좌, 우영상의 정합점을 찾고 대응하는 변이를 구하는 것이 일반적이다.

스테레오 영상의 좌, 우 영상의 상관도를 이용하여 대응점을 찾아내는 과정을 일반적으로 스테레오 정합(stereo matching)이라고 한다. 또한 두 영상 간의 상관도에 따른 상대적인 거리를 표시한 것이 변이 지도(disparity map)이며, 변이 지도는 두 영상간의 상관도를 나타내는 중요 정보를 포함하고 있다. 변이정보를 이용하여 가상시점 영상을 생성할 수 있다. 이처럼 변이 지도는 매우 중요한 의미를 지니며 따라서 좌, 우 영상에서 최적의 정합점을 찾는 작업은 매우 중요하다[1,2].

본 논문에서는 인물 위주의 스테레오 영상에 적합한 스테레오 정합 기법을 제안한다. 제안한 기법에서는 두 대의 스테레오 카메라로부터 획득된 영상에서 비슷한 화소 값을 가지는 배경영역을 먼저 제거하게 된다. 배경이 제거된 영상을 이용하여 초기 변이지도와 영상의 색상 정보를 이용한 분할 변이지도를 구하고, 이들을 적당

히 조합하여 최종 변이지도를 획득한다.

스테레오 카메라를 통해 획득된 영상에서 배경을 제거한 후 남은 인물 위주의 전경을 하나의 정합 창으로 인식하여 초기 변이지도를 구하게 된다. 배경이 제거된 영상을 R, G, B, white의 4가지 색상 성분으로 분할하고, 각 색상 성분에서 독립적으로 분할 변이지도를 추정한다. 초기 변이지도는 최종 변이지도를 생성하기 위하여 분할 변이지도 획득 시 초기 값으로 이용된다. 또한 분할된 변이지도는 계층적 스테레오 정합을 이용하여 R, G, B, white의 각 색상 성분으로 분할된 영역을 하나의 정합 창으로 사용하여 스테레오 정합 과정을 수행한다. 마지막으로 이렇게 생성된 각 색상 영역의 변이지도를 조합하여 최종 변이지도를 생성하게 된다.

#### 2. 제안하는 기법

본 논문에서 제안하는 스테레오 정합 기법은 크게 두 과정으로 이루어져 있다. 배경을 제거하여 초기 변이지도를 획득하는 과정과 영상의 색상 정보를 이용하여 영역을 분할한 후 변이지도를 획득하는 과정이다. 초기 변이지도는 전경을 제외한 배경 영역이 제거된 영상을 이용하여 획득하며, 배경이 제거된 영상에서 색상 성분을 추출하여 분할 변이지도를 생성할 때 발생하는 오류를 최소화 하는데 이용된다.

##### 가. 배경 제거 기법

제안하는 기법에서 정확한 분할 변이지도를 얻기 위해서는 인물 위주의 전경을 제외한 나머지 배경을 제거할 필요가 있다. 배경을 제거하기 위해서 먼저 스테레오 카메라에서 영상을 각각 획득한 후 미리 저장된 배경 영상과의 차이를 구한다.

카메라를 통해 획득한 영상에서 스테레오 정합에서 가장 불필요한 부분은 인물 위주의 전경을 제외한 배경 영역이다. 본 논문에서는 화상회의 시스템 등의 응용분야에 적용하는 경우를 목표로 하고 있기 때문에 획득된 영상은 대부분 배경의 변화가 크게 없는 상태이다. 따라서 스테레오 정합 기법을 적용하기 전에 배경 영역을 획득하여 미리 메모리에 저장하고, 식 (1)을 이용하여 새로 획득 되는 영상과 화소 단위로 비교하여 배경 영역을 제거하게 된다[3,4].

$$f(x,y) = \begin{cases} f(x,y), & (|R_b - R_f| + |G_b - G_f| + |B_b - B_f|) > K \\ 0, & \text{else.} \end{cases} \quad (1)$$

여기서 K는 비교를 위한 임계값(threshold value)이다. 배경 영역 내에도 시간에 따른 조명 변화 등에 의해 화소별 작은 오차가 발생할 수 있다. 하지만 임의의 객체가 카메라의 촬영 영역 내에 들어와 배경을 가리게 되면, 이때는 매우 큰 차이가 발생하게 된다. 이러한 조건을 이용하여 많은 최적의 임계 값을 선택하게 되고, 이 임계 값보다 차이가 클 때는 전경의 객체로 차이가 작을 때는 배경 영역으로 각각 판단하게 된다. 그림 1(e)와 (f)는 원 영상에서 배경을 제거한 결과 영상을 나타낸다.

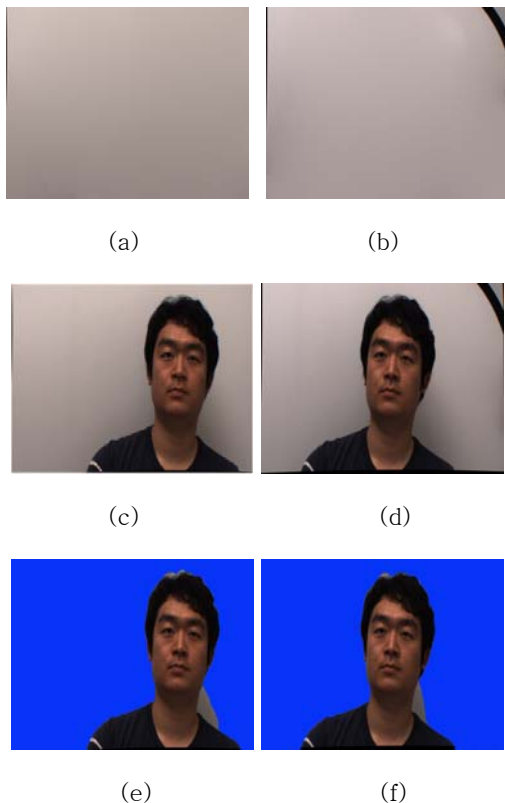


그림 1. 배경 영역을 제거한 결과

(a), (b) 좌, 우 배경 영상 (c), (d) 좌, 우 현재 영상  
(e), (f) 좌, 우 배경이 제거된 영상

나. 초기 변이지도 획득

초기 변이지도를 획득하기 위하여 앞 절에서 배경을 제거한 영상을 이용하여 마스크 영상을 생성 한다. 마스크 영상을 하나의 정합 창으로 인식하여 객체에 대한 초기 변이지도를 획득하게 된다.

그림 2는 배경이 제거 된 영상을 이용하여 마스크 영상을 생성한 것이다. 이렇게 생성된 마스크 영상을 하나의 정합 창을 사용하여 초기 변이지도를 생성한 결과를 그림 3에 보였다.

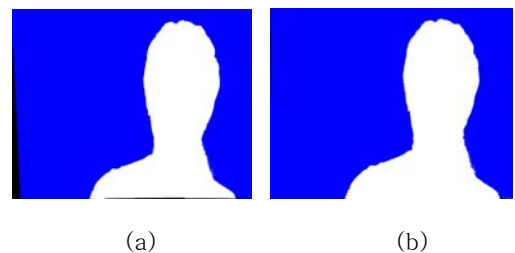


그림 2. (a) 좌 마스크 영상 (b) 우 마스크 영상

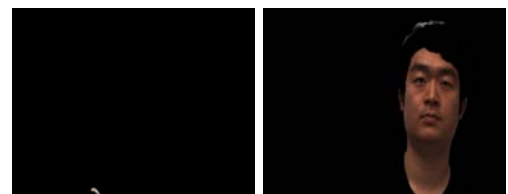


그림 3. 초기 변이지도

다. 영역분할

분할 변이지도를 생성하기 위하여 먼저 배경이 제거된 영상을 가지고 각 색상 성분으로 영역 분할을 하고, 각 분할된 영역을 가지고 경계 변이지도를 생성한다. 먼저 획득된 영상을 네 가지 색 성분으로 영역 분할하는데 RGB 영상에서 white 축을 중심으로 일정한 거리 안에 존재하는 값은 white로 간주하고, 나머지 영역들은 각 색상 성분으로 판단하게 된다. 획득된 영상을 white 성분을 제외한 R, G, B 세 개의 색상 성분으로 변환을 하면, 각 색상 성분 간의 경계가 모호함으로 인해 변이 생성에 있어 오히려 낮은 신뢰도의 결과를 가져올 수 있다[5,6].

그림 4는 두 대의 스테레오 카메라로 획득된 좌 영상으로부터 분할 변이지도를 획득하기 위하여 각 색상 성분으로 분할한 영상을 나타낸다.



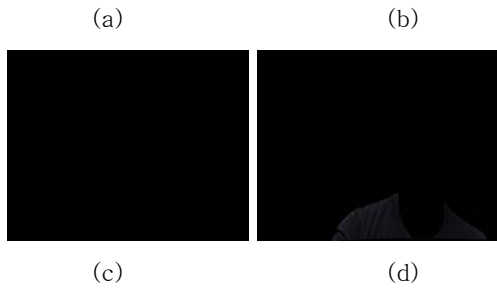


그림 4. 분할된 색상 영상(좌 영상)  
(a) white (b) red (c) green (d) blue

영상은 일반적으로 모서리, 경계 등에서 특징 점이 존재하며, 이러한 특징 점을 추출하기 위한 방법은 다양하다. 본 논문에서는 소벨(Sobel) 마스크를 이용하여 획득된 영상의 각 색상 성분으로부터 경계 성분을 추출하고, 이를 특징 점으로 이용한다.

그림 5는 소벨 마스크와 이를 이용하여 획득된 좌 영상의 각 색상 성분에 대한 경계 영상이다.

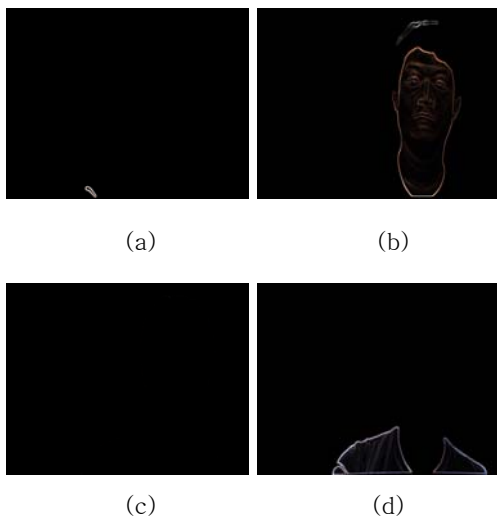


그림 5. 분할된 색상 영상의 경계 성분(좌영상)  
(a) white (b) red (c) green (d) blue

### 라. 변이지도 획득

실험을 통하여 3×3, 5×5, 7×7, 9×9, 15×15, 21×21 등의 정합 창 중에서 가장 성능이 우수한 9×9의 정합 창을 가지고 경계 변이지도를 생성한다. 이때 변이지도 추출을 위한 비용 함수는 식 (2)의 SSD(sum of square difference)를 이용한다.

$$d(x,y) = \underset{d_{\min} \leq d \leq d_{\max}, (i,j) \in N(x,y)}{\operatorname{argmin}} \{R(i+d,j) - L(i,j)\} \quad (2)$$

여기서 N(x,y)은 9×9 정합 창을 의미하고 (i,j)는 정합 창 내의 화소 위치를 나타낸다. R(x,y)와 L(x,y)는 좌, 우

영상의 화소 값을 나타내며, d는 범위 내에 존재하는 초기 변이 값을 나타내고, d(x,y)는 (x,y)에서 얻어진 최종 변이 값을 의미한다.

그림 6은 경계 변이지도 추출과 확산 과정을 통해 얻은 색상 영역 분할 영상의 변이지도를 나타낸다.

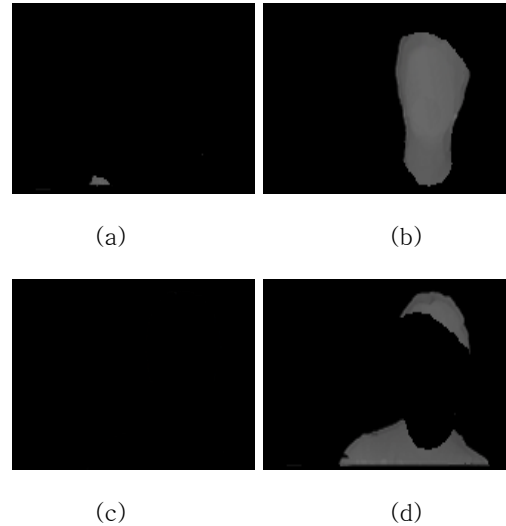


그림 6. 각 색상 영역의 변이지도(좌영상)  
(a) white (b) red (c) green (d) blue

위 과정을 통해 각 색상 성분에 대한 경계 변이지도를 획득이 완료되면 최종 변이지도를 생성한다. 확산을 통해 변이지도가 겹치는 영역이 발생할 수 있기 때문에 겹치는 화소 간의 평균값을 식(3)과 같이 변이지도 값으로 함으로써 변이지도가 연속성을 가지게 할 수 있다[7].

$$\begin{cases} D(i,j) = (d_w(i,j) + d_r(i,j) + d_g(i,j) + d_b(i,j))/K \\ \text{else. } D(i,j) = d_w(i,j) \text{ or } d_r(i,j) \text{ or } d_g(i,j) \text{ or } d_b(i,j) \end{cases} \quad (3)$$

여기서 D(i,j)는 최종 변이지도를 나타내며 (i,j)의 화소에 여러값이 존재할 경우 평균 값을 구하며, K는 (i,j)의 위치에서 white, R, G, B 값을 가지느냐에 따라 2-4까지의 값을 가지며 (i,j)에서 값이 하나만 있을 경우 그 값이 최종 변이지도 값이 된다.

### 3. 실험 결과

본 논문에서는 point gray 사의 모델인 Grasshopper2 카메라를 이용하여 획득된 영상을 실험영상으로 하였다. 두 대의 카메라 간의 거리는 6.5cm를 유지하여 촬영하였으며 카메라 간 동기와 보정을 미리 수행하고 본 논문의 실험을 수행하였다.

촬영한 영상의 대부분의 경우 중요 객체를 포함하는 전경과 배경 간의 경계에서 부정확성을 나타낸다. 이런

경우 블록 정합 기법을 사용하면 정합 창의 크기에 따라 다른 결과가 발생하고, 많은 영역에서 부정확한 결과를 얻게 된다. 반면에 본 논문에서 제안하는 기법을 이용하면 전경과 배경 간의 경계가 부정확하더라도 신뢰도 높은 변이지도를 생성할 수 있다. 그림 7에서 보듯이 획득된 여러 실험 영상에서 제안한 기법의 성능이 기존의

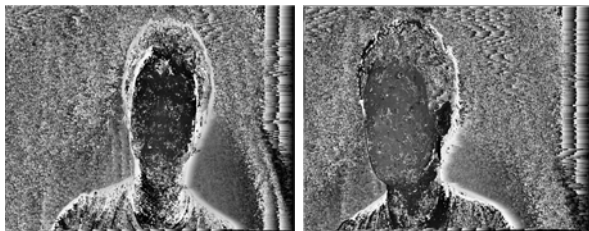
기법보다 우수한 것을 확인할 수 있다[1,2].



(a)



(b)



(c)

그림 7. 실험 영상의 최종변이지도

(a) 원 영상 (b) 제안한 기법 (c) 영역기반 정합기법

#### 4. 결론

본 논문에서는 인물 위주의 스테레오 영상에서 배경을 제거한 뒤 색상 정보를 이용한 새로운 분할 스테레오 정합 기법을 제안하였다. 배경 영상을 제거한 뒤 마스크 영상으로 변환하여 초기 변이지도를 생성하였고, 배경이 제거된 영상에서 색상 정보를 이용하여 각 색상 변이지도를 구하고 이를 잘 조합하여 최종 변이지도를 생성하였다.

영역기반 정합 기법에서 비슷한 화소 값이 많은 경우 발생하는 오류를 최소화 시켰으며, 각 색상 영역에서 정확한 변이지도를 검출함으로써 신뢰도 높은 최종 변이지도를 생성할 수 있었다.

#### <감사의 글>

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2011-0006791)

#### <참고문헌>

- [1] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient belief propagation for early vision," IEEE Conf. Computer Vision and Pattern Recognition, June 2004.
- [2] Scharstein, D., Szeliski, R., and Zabih, R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," IEEE Workshop on Stereo and Multi-Baseline Vision, 2001(SMBV 2001), pp. 131-140, Dec. 2001.
- [3] Taehee Lee, and Hollerer, T., "Hybrid feature tracking and user interaction for markerless augmented reality", Virtual Reality Conference '08, IEEE, pp. 145-152, March 2008.
- [4] 황선규, 영상처리 프로그래밍 by Visual C++, 한빛미디어, pp. 660-671, pp. 684-695, 2007년.
- [5] A. Klaus, M. Sormann, and K. Karner, "Segment based stereo matching using belief propagation and a self-adapting dissimilarity measure," ICPR '06, vol. 3, pp. 15-18, 2006.
- [6] Michael Bleyer and Margrit Gelautz, "A layered stereo algorithm using segmentation and global visibility constraints," ICIP '04. vol. 5. pp. 2997- 3000, 2004.
- [7] Brown, M.Z., Burschka, D., and Hager, G.D., "Advances in computational stereo," IEEE Transactions on PAMI, vol. 25, iss. 8, pp. 993 -1008, Aug. 2003.