

# 기울기 영상의 센서스 변환을 이용한 전역 기반 스테레오 정합 방법

고재런, 호요성  
광주과학기술원 전기전자컴퓨터공학부  
jrko@gist.ac.kr, hoyo@gist.ac.kr

## Global Stereo Matching Using Census Transform on Gradient Stereo Images

Jaeryun Ko, Yo-Sung Ho  
School of Electrical Engineering and Computer Science  
Gwangju Institute of Science and Technology (GIST)

### 요 약

센서스 변환을 정합 비용 함수로 이용한 스테레오 정합 방법은 별도의 매개변수 설정을 필요로 하지 않으므로 정합 장 기반의 다른 정합 비용 계산 방법보다 좋은 성능을 보이는 것으로 알려져 있다. 이 논문에서는 기존 밝기 값 기반의 센서스 변환과 더불어 각 화소의 기울기 값 기반의 센서스 변환을 포함한 정합 비용 함수 계산을 제안하였다. 실험 결과는 깊이 불연속 경계 영역에서 보다 정확한 변위 값이 할당되었음을 보여준다.

### 1. 서 론

스테레오 영상을 이용한 변위 지도 생성은 색상 영상 기반의 3차원 콘텐츠 생성을 위한 가장 기초적인 필요 과정으로써 이에 대한 다양한 방법이 연구되어왔다. 이를 위한 작업을 스테레오 정합이라 하며 서로 다른 시점을 갖는 스테레오 영상을 이용해 깊이를 측정하는 방법이다. 일반적으로 스테레오 정합 알고리즘은 크게 지역 기반 방법과 전역 기반 방법으로 나누어진다.

지역 기반 방법은 두 영상의 화소 간 정합 비용을 계산하여 변위 값을 결정하며, 구현이 간단하고 정합 속도가 빨라 실시간 응용에 적합하다. 정합 비용을 계산하기 위해 적절한 크기의 정합 창이 결정되어야 하는데, 이에 따라 정합 품질이 좌우되며 전체적으로 전역 기반 방법보다 품질이 낮은 편이다.

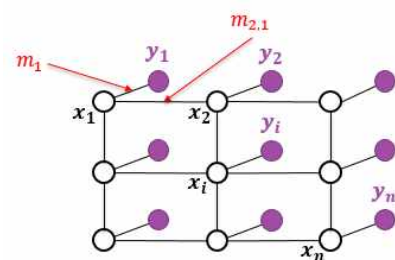
전역 기반 방법은 MRF(Markov Random Field) 기반의 에너지 함수를 최소화 시키는 값으로 적절한 변위 값을 추정하는 방법이며, 이를 위한 최적화 알고리즘으로 신뢰 확산(Belief Propagation)이나 그래프 컷(Graph Cut) 등의 방법을 사용한다. 에너지 함수의 최적화 과정에서 영상 전체 화소 간 관계를 고려하기 때문에 정합속도가 상대적으로 느리다.

이 논문은 기존의 밝기 값 기준의 센서스 변환(Census Transform)을 이용한 정합 비용 계산으로 전역 스테레오 정합을 실시한 결과에서 객체 경계 부분에서의 변위 값 할당의 정확성을 높이기 위해 기울기(Gradient) 영상을 이용한 센서스 변환을 추가하였다. 계층적 신뢰 확산 방법을 최적화 알고리즘으로 사용하여 깊이 불연속 경계 영역에서 결과가 개선되는 것을 확인하였다.

### 2. 전역 기반 스테레오 정합 방법

#### 2.1 신뢰 확산(Belief Propagation)

신뢰 확산을 이용한 스테레오 정합은 마르코프 네트워크(Markov Network) 상의 하나의 노드, 즉 화소에서 주변 화소로 메시지라 불리는 값들을 확산시키며 진행된다 [1]. 이때 각각의 화소는 주변으로부터 전달받은 모든 메시지 값들을 최대사후확률로 계산하여 변위를 할당하게 된다. 그림 1은 신뢰 확산을 이용한 전역 스테레오 정합에서의 마르코프 네트워크를 보여준다.



(그림 1) 영상에서의 마르코프 네트워크

화소에 대한 변위 값을 구하기 위해서 초기에  $y_i, i=1, \dots, n$ 은 각 변위 후보에 대한 정합 비용의 값을 보유하고 있다.  $m_i$ 는 노드  $y_i$ 에서  $x_i$ 로 보낸 메시지이며  $m_{ij}$ 는  $x_i$ 에서  $x_j$ 로 보낸 메시지를 의미한다. 신뢰 확산의 첫 단계로 이들 메시지를 모두 동일한 분포를 가지도록 초기화 한다. 다음으로 식 (1)을 통해 메시지가 수렴할 때까지 반복적으로 갱신하는 과정이다. 마지막으로 각 화소에서의 신뢰도  $b_i(x_i)$ 는 식 (2)을 통해 계산한다. 이때 최

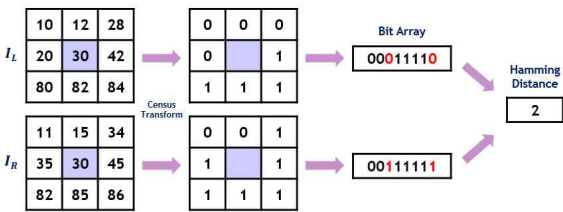
중적으로 구하고자 하는 값은 해당 신뢰도를 최소화하는  $x_i$ 을 선택한다.

$$m_{ij}^t = \min_{x_i} V(x_i, x_j) + D_i(x_i) + \sum_{s \in N(i) \setminus j} m_{sj}^{t-1} \quad (1)$$

$$b_i(x_i) = D_i(x_i) + \sum_{j \in N(i)} m_{ij}^t \quad (2)$$

2.2 기울기 영상에 대한 센서스 변환(Census Transform)

센서스 변환 방법은 입력 영상의 상대적인 밝기 값을 사용하며, 정합 비용 계산을 위한 함수로써 좋은 성능을 내는 것으로 알려져 있다 [2]. 일정한 크기의 변환 창을 통해, 창 중앙의 화소를 제외한 나머지 화소에 대하여, 중앙 화소와의 밝기 값의 크기 비교를 통해 비트 열로 나타낸다. 이를 비트열의 해밍 거리(Hamming Distance)를 계산하여 정합 비용으로 사용하게 된다. 그림 2는 센서스 변환의 과정을 간략하게 도식화 한 것이다.



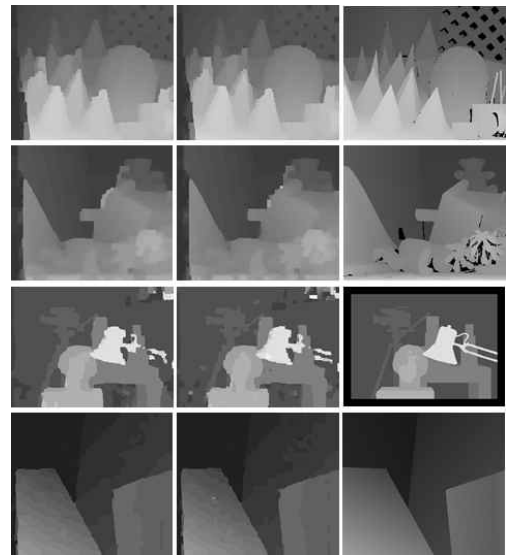
(그림 2) 센서스 변환의 전체 과정

기존의 센서스 변환을 이용한 정합 비용 계산에 덧붙여 깊이 불연속 영역에서의 정확한 변위 값 할당의 가능성을 향상시키기 위해, 기울기 영상에 대하여 센서스 변환을 추가함으로써 정합 비용 계산에 적용하였다 [3]. 먼저 입력 스테레오 영상에 대하여 가로 방향 소벨(Sobel) 연산자를 적용하여 얻어낸 기울기 영상  $G_x$ 을 획득한다. 밝기 값에 대해 적용한 방식과 동일하게 센서스 변환을 적용하여 이를 통해 얻어낸 해밍 거리를 추가하여 식 (3)을 정합 비용으로 사용하였다.

$$Cost(x_i, d) = \alpha Hamming_I + (1 - \alpha) Hamming_{G_x} \quad (3)$$

3. 실험 결과

기존의 센서스 변환만을 정합비용에 사용한 결과를 비교 평가하기 위하여 Middlebury Stereo Dataset에서 제공하는 스테레오 영상인 Cones, Teddy, Tsukuba, Venus를 실험 대상으로 사용하였다 [4]. 센서스 변환 창의 크기는 5x5로 고정하였으며, 신뢰 확산의 구현 중 하나인 계층적 신뢰 확산(Hierarchical Belief Propagation, HBF) 방법을 통해 전역 스테레오 정합을 실시하였다. 그림 3은 기존의 센서스 변환만을 사용한 정합 비용 계산 방법에서 어느 정도 향상 되었는지 보여준다. 미약하게나마 깊이 불연속 경계 영역에서 실제로 가까운 변위 값이 할당되었음을 확인할 수 있다.



(a) 기존 센서스 변환 (b) 기울기 추가 센서스 변환 (c) 실제 변위 지도

(그림 3) 최종 실험 결과 비교

4. 결론

이 논문에서는 기울기 영상의 센서스 변환을 이용하여 정합 비용 계산을 수정함으로써 개선된 전역 스테레오 정합 방법을 제안하였다. 깊이 불연속 경계 영역에서 보다 정확한 변위 값을 할당하기 위해 가로 방향 소벨 연산자를 적용한 기울기 영상을 이용하여 센서스 변환을 적용, 추가된 해밍 거리에 의한 정합 비용 계산식을 수정하였다. 제안하는 방법을 통해 미약하게나마 깊이 불연속 경계 영역에서 실제로 가까운 변위 값을 할당할 수 있었다.

감사의 글

본 연구는 미래창조과학부 ‘범부처 Giga KOREA 사업’의 일환으로 수행하였음. [GK15C0100, 기가급 대용량 양방향 실감 콘텐츠 기술 개발]

참고 문헌

[1] S. Yedidia, W. T. Freeman, and Y. Weiss, “Understanding belief propagation and its generalizations,” Exploring Artificial Intelligence in the New Millenium, pp. 239-269, 2003.  
 [2] S. Patil, J. S. Nadar, J. Gada, S. Motghare and S. S. Nair, “Comparison of Various Stereo Vision Cost Aggregation Methods,” International Journal of Engineering and Innovative Technology, vol. 2, pp. 222-226, 2013.  
 [3] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, “Fast Cost-Volume Filtering for Visual Correspondence and Beyond,” Computer Vision and Pattern Recognition, pp. 3017-3024, 2011.  
 [4] Middlebury Stereo Datasets, <http://vision.middlebury.edu/stereo/data/>