

# Bilateral 필터를 이용한 그래프 컷 기반의 다해상도 스테레오 정합

홍석근\*, 김정엽\*\*, 원종운\*\*\*, 조석제\*\*\*\*

\*한국해양대학교 제어계측공학과, \*\*영산대학교 자유전공학부

\*\*\*한국철도기술연구원, \*\*\*\*한국해양대학교 IT공학부

e-mail:bassoflord@gmail.com

## A Multiresolution Stereo Matching Using a Bilateral Filter Based on Graph-Cut

Seok-Keun Hong\*, Kim Jeong-Yeop\*\*, Won Jong-Woon\*\*\*, Seok-Je Cho\*\*

\*Dept of Control & Instrumentation Engineering, Korea Maritime and Ocean University

\*\*School of Undeclared Majors, Youngsan University, \*\*\*Korea Railroad Research institute

\*\*\*\*Division Information Technology Engineering, Korea Maritime and Ocean University

### 요 약

본 논문은 스테레오 시각에서 3차원 정보를 얻기 위해 bilateral filter를 이용한 그래프 컷 기반의 다해상도 스테레오 영상 정합 방법을 제안하고자 한다. 제안한 방법은 변위도에서 물체의 경계를 명확히 구분하기 위해 bilateral filter를 이용하여 그래프의 각 노드들을 연결하는 경로의 가중치를 결정하고 정합 비용을 통합한다. 정합 과정에서 계산복잡도를 줄이기 위해 계층적 다해상도 구조를 적용하여 영상 피라미드를 만들고, 정합의 정확성을 향상시키기 위해 정합 영역에 변위 평활성과 같은 제약 조건을 적용하여 변위를 전파하는 방법을 사용한다. 실험을 통해 제안한 방법이 변위 탐색 시간을 감소시킬 뿐만 아니라 기존의 그래프 컷의 단점을 보완할 수 있음을 확인하고자 한다.

### 1. 서론

스테레오 시각(Stereo vision)은 한 장면을 담은 스테레오 영상에서 물체의 위치 차이를 계산하여 3차원 정보를 추출하는 과정이다. 스테레오 시각 연구에서 두 영상간의 대응점을 찾아내는 정합 과정(stereo matching)이 가장 어렵다. 그래서 대부분의 스테레오 시각에 대한 연구는 대응성 문제, 즉 정합점을 정확히 찾는 데 초점이 맞춰져 있다 [1].

스테레오 정합은 크게 두 부류로 나눌 수 있다. 하나는 정해진 크기의 영역에 스테레오 영상의 특성에 따른 제약 조건들을 적용하여 변위를 구하는 지역적 정합 기법(local matching method)이고, 다른 하나는 신경회로망(neural network), MRF(Markov Random Field) 이론에 기반한 신뢰 확산법(belief propagation)과 그래프 컷(Graph-Cut)[2,3] 및 진화 연산 알고리즘(evolutionary computation algorithm)과 같은 최적화 기법을 적용하여 전체 영상의 에너지 함수를 계산하여 최적의 변위 분포를 구하는 전역적 정합 기법(global matching method)이다 [2]. 이들 중 전역적 정합 기법이 지역적 정합 기법에 비해 정합의 정확도가 높다고 알려져 있으나, 지역적 정합 기법보다 적용하기 어렵고 계산복잡도가 높아 효율성을 향상시키는 방향으로 연구가 주로 이뤄지고 있다.

신뢰 확산법을 이용하는 스테레오 정합은 영상 전체의 화소를 이용하기 때문에 폐색 영역에 대한 오정합이 낮은

편이다. 그러나 여러 단계에 걸쳐 반복적으로 정합을 수행해야 하고 모든 화소에 대해 주어진 변위 탐색 범위 내의 변위들의 확률들을 모두 고려해야 하므로 계산 복잡도가 높고 무질감 영역에 있는 화소들이 주변 화소에 영향을 받아 오정합이 발생할 수 있다.

그래프 컷은 스테레오 영상의 모든 화소를 변위를 이용해 라벨링(labeling)하여 화소들의 라벨, 즉 변위가 같은지 여부를 확률로 판별해 에너지 함수를 최소화하는 라벨들을 연결하는 방식으로 신뢰 확산법보다 효율적인 방법으로 알려져 있다. 그래프 컷은 라벨이 다를 확률이 높은 화소는 끊어내어(cut) 해당 화소의 라벨은 다음 단계 처리에서 고려하지 않는다. 그래프 컷을 이용한 방법은 경계선에서 변위의 정확도가 높으나 물체의 경계가 아닌 단순한 에지 화소에서도 변위값을 다르게 구하기 쉽기 때문에 물체의 세세한 부분에서 정합 성능이 좋지 않고, 경계 영역에서도 연속성 제약으로 인해 뭉개지는 현상이 나타날 수 있다[3].

본 논문에서는 bilateral 필터를 이용한 그래프 컷 기반의 다해상도 스테레오 정합 방법을 제안한다. 그래프 컷을 이용한 최적화 과정에서 물체의 경계선을 명확히 보전하기 위해 bilateral 필터를 이용하여 그래프의 각 노드를 잇는 경로(path)에 가중치를 계산한다. 그리고 bilateral 필터 비용 통합(cost aggregation)을 수행하여 에너지 함수에 기하학적 차이와 광도적 차이를 반영한다. 정합 과정에서 증가하는 계산 복잡도를 줄이기 위해 다해상도 기법을 적

용하여 스테레오 영상을 정해진 단계까지 축소하여 영상 피라미드를 만든다. 이 때 근사영상, 예측차영상 및 에지 영상에 대한 계층구조를 만든다. 이러한 방법으로 최상위 계층에서 초기 변위도(initial disparity map)를 얻고, 각 계층별 변위도는 영상의 에지 정보와 변위 연속성(disparity smoothness)과 같은 제약 조건(constraints)을 이용하여 전파(propagation)하여 얻는다. 실험을 통하여 제안한 방법의 성능을 확인하였다.

**2. 제안한 방법의 개요**

제안한 방법은 효율적인 정합을 위해 다해상도 스테레오 정합 기법을 기반으로 하며 정합 비용 계산을 포함한 최적화 기법은 영상 피라미드 최상위 계층에서 수행한다. 최상위 계층에서 얻은 변위는 각 계층별 에지 영상과 지역 특징 벡터와 함께 하위 계층의 변위를 계산하는데 사용된다.

다해상도 스테레오 정합 제안한 방법의 블록 다이어그램은 그림 1과 같다. 주어진 스테레오 영상으로 피라미드 구조를 만든다. 원영상 크기를 레벨 0로 두고 단계별로 2 배씩 축소하여  $j+1$  계층의 피라미드를 만든다. 이 때 기준영상의 각 단계별 근사영상에서 에지 영상을 생성한다. 이는 원래 크기로 복원하는 과정에서 각 단계별로 복원될 변위도의 정확한 에지 위치를 보장하기 위함이다. 레벨  $j$ 의 에지 영상을 이용하여 각 화소에서의 윈도우 크기 및 모양을 결정하고, 스테레오 근사 영상에 적응적 윈도우를 적용하여 초기 변위도를 얻는다. 그리고 이 변위도를 단계별로 확대하며 전파 및 보간을 수행한다. 이 때 각 영역에 변위 제약 조건을 적용하여 변위를 정확히 전파하고 각 단계별로 얻어진 에지의 위치 정보를 참고하여 해당 단계의 변위도를 얻는다. 이러한 과정을 레벨 0까지 수행하여 원영상 크기의 최종 변위도를 얻는다.

**3. bilateral 필터를 이용한 비용 계산 및 통합**

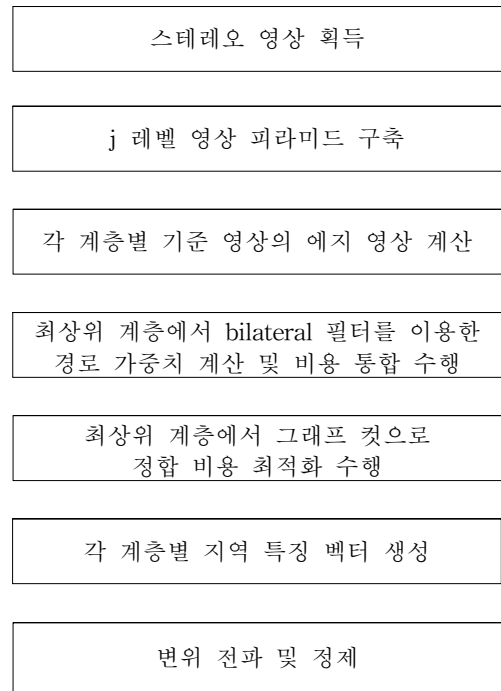
좌영상을 기준 영상으로 두고, 좌영상  $I_L$ 의 화소  $p$ 와 할당된 변위  $D_p$ 에 의해 대응되는 우영상  $I_R$ 의 화소에 대한 유사도는 식 (1)과 같이 AD(Absolute Difference)로 계산한다.

$$C_{AD}(p, D_p) = |I_L(p) - I_R(p - D_p)| \quad (1)$$

제안한 방법은 bilateral 필터를 이용하여 영상의 기하학적 차이와 광도적 차이를 기반으로 정합 비용에 가중치를 다르게 적용한다. 정해진 크기의 영역 A의 중심 화소  $p$ 와 영역 내 주변 화소  $q$ 에 대한 bilateral 필터[1]는 식 (2)과 같이 라플라시안 가중치 함수  $w(p, q)$ 로 표현할 수 있다.

$$w(p, q) = \exp\left(-\frac{\|p - q\|^2}{2\sigma_d^2} - \frac{\|I(p) - I(q)\|^2}{2\sigma_s^2}\right) \quad (2)$$

시작



최종 변위도

(그림 1) 제안한 방법의 블록 다이어그램

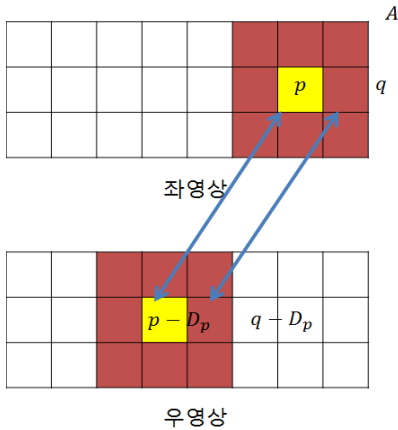
여기서  $I(p)$ 와  $I(q)$ 는 각각 화소  $p$ 와  $q$ 의 화소값이고,  $\sigma_d$ 와  $\sigma_s$ 는 각각 영역 A에서 계산된 위치상 근접성과 화소값 유사도의 표준편차이다.

좌영상의 화소  $p$ 와 우영상에서  $p$ 의 대응점  $p - D_p$ 을 중심으로 하는 영역에 대한 정합 비용 통합을 수행한다. 좌영상의 영역 A의 통합된 정합 비용(aggregated matching cost)  $C_{AD}^A(p, D_p)$ 는 식 (1)의 화소간 유사도와 식 (2)의 가중치를 이용하여 얻는다. 이 때, 가중치는 좌영상과 우영상에 대해 모두 구한다.

$$C_{AD}^A(p, D_p) = \frac{\sum_{q \in A} w_L(p, q) w_R(p - D_p, q - D_p) C_{AD}(q, D_p)}{\sqrt{\sum_{q \in A} [w_L(p, q)]^2 \sum_{q \in A} [w_R(p - D_p, q - D_p)]^2}} \quad (3)$$

여기서  $w_L(p, q)$ 와  $w_R(p - D_p, q - D_p)$ 는 각각 좌영상과 우영상의 중심화소  $p$ ,  $p - D_p$ 와 주변 화소와의 가중치이다.

제안한 방법의 정합 비용 계산은 그림 2와 같다. 영역 A의 크기를  $3 \times 3$  일 때, 좌영상의 중심화소  $p$ 와 우영상의 중심화소  $p - D_p$  뿐만 아니라 그 주변에 위치한 화소들도 영역내에서 동일한 위치에 있는 화소들끼리 가중치와 유사도를 계산한다.



(그림 2) 제안한 방법의 정합 비용 계산

4. 경로 가중치를 이용한 그래프

통합된 비용함수는 그래프 컷을 이용하는 최적화 기법에 적용된다. 그러나 기존의 그래프 컷은 변위 연속성을 이용한 비용이 대체적으로 크게 계산되는 경향이 있어서 물체의 경계선 부근이 뭉개지는 단점을 보인다[3]. 이를 개선하기 위해 제안한 방법은 좌영상을 실수 가중치를 갖는 그래프로 표현한다. 그래프  $G$ 는 식 (3)과 같다.

$$G = (V, E) \tag{3}$$

여기서  $V$ 는 그래프로 표현된 모든 좌영상의 모든 화소의 집합이고  $E$ 는  $V$ 에 속한 화소와 인접한 화소들 간의 에지를 나타낸다.  $V$ 는 공집합이 될 수 없지만  $E$ 는 가능하다.

가중치는 영상의 기울기(gradient)를 나타낸다. 화소  $p$ 와  $q$ 가 서로 인접해 있을 때, 두 화소간의 가중치  $w_g(p, q)$ 는 식 (4)와 같이 bilateral 필터의 화소값 유사성으로 계산할 수 있다.

$$w_g(p, q) = \exp\left(-\frac{|I(p) - I(q)|}{\sigma_s}\right) \tag{4}$$

가중치  $w_g$ 를 이용하여 그래프  $G$ 의  $E$ 에서 원하지 않는 에지를 제거할 수 있다. 이는 그래프 컷 수행 과정에서 임계치보다 낮은 가중치의 경로로 연결된 노드들이 서로 병합되는 현상을 방지할 수 있다. 최적화 기법과 변위 연속성 비용은 기존의  $\alpha$ -cut[2] 기법과 동일하다.

5. 영상 피라미드 구조와 변위 전파

제안한 다해상도 기반의 스테레오 정합을 그림 3에 나타내었다. 레벨 0의 스테레오 영상에 피라미드 기법을 적용하여 각각 레벨  $j$  단계의 피라미드 영상들을 얻는다. 레벨  $j$ 의 스테레오 영상을 정합하고 다음 단계인 레벨  $j-1$ 에 해당하는 크기로 확대하면서 지역적 특징 정보를 이용하여 이전 단계 변위를 정확한 위치로 전파하고 주변 화

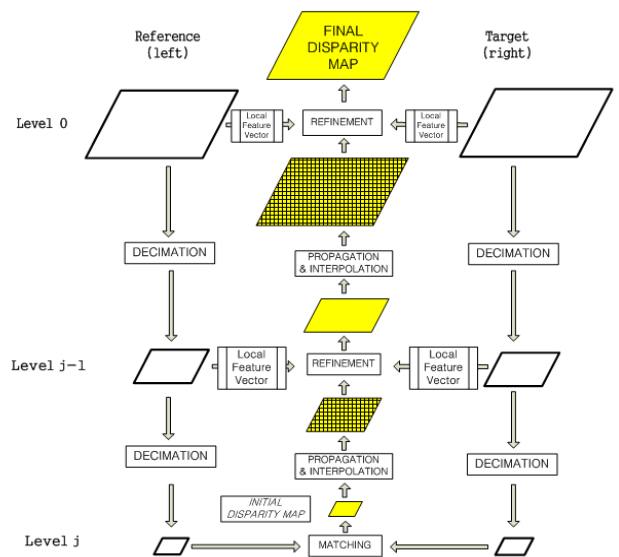
소들의 변위를 동시에 구한다. 그리고 레벨  $j-1$  단계의 스테레오 영상과 에지 영상을 이용하여 변위를 정제함으로써 현 단계 변위도의 정확도를 높인다. 이러한 일련의 과정을 거쳐 원래 크기인 레벨 0의 최종 변위도를 얻는다.

계층적 접근을 통해 스테레오 영상 각각에 대한 근사 영상 피라미드와 예측차 영상 피라미드를 얻을 수 있다. 레벨  $j-1$ 의 영상을 근사화 필터(approximation filter)처리하여 2배 축소하면 레벨  $j$ 의 근사 영상을 얻을 수 있다. 2배 확대한 영상과 레벨  $j-1$ 의 예측차 영상을 이용하여 레벨  $j-1$ 의 영상을 복원할 수 있다. 이런 과정을 반복하여 레벨 0의 원영상을 복원할 수 있다.

제안한 방법은 가우시안 저역통과 필터를 이용하는 가우시안 피라미드 기법을 이용하여 계층화된 영상들을 얻고 동시에 각 단계별 근사영상에 대한 에지 영상을 얻는다. 레벨  $j$ 까지 계층화가 이뤄지면 최상단 근사 영상에 대한 스테레오 정합을 실시하여 최초 변위도(initial disparity map)을 얻는다.

레벨  $j$ 의 최초 변위도를 얻은 후 변위도를 단계적으로 확대하며 변위를 전파한다. 최초 영상에서 화소  $(x, y)$ 가 2배 확대되면 이 화소는 2배 확대된 영상, 즉 레벨  $j-1$ 의  $(2x, 2y)$ ,  $(2x+1, 2y)$ ,  $(2x, 2y+1)$ ,  $(2x+1, 2y+1)$ 에 해당한다. 그리고 해상도가 두배씩 증가함에 따라 대응점의 상대거리도 2배 가까이 증가할 수 있다. 먼저 레벨  $j$ 의 변위는 레벨  $j-1$ 의 4개의 화소에 저장된다. 정확한 변위를 전파하기 위해 그림 4와 같은 Haar 웨이블릿 필터를 이용한 지역적 특징 정보를 사용한다. 레벨  $j-1$  스테레오 영상에  $2^2 \times 2^2$  크기의 Haar 웨이블릿 필터를 이용하여 4차원의 특징 벡터  $V$ 를 식(3)과 같이 구성한다. 이때 필터는 단계마다 승수를 1씩 증가시켜 레벨 0에서  $2^{j+1} \times 2^{j+1}$ 의 크기로 처리한다.

$$V = [\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|] \tag{5}$$



(그림 3) 제안한 다해상도 기반 스테레오 정합

여기서,  $d_x$ ,  $d_y$ 는 수평방향과 수직방향  $4 \times 4$  크기의 Haar 웨이블릿 필터로 얻은 응답을 나타낸다. 각 위치에 전과된 변위를 이용하여 식 (6)과 같은 특징차  $Dist(x,y)$ 를 구한다.

$$Dist(x,y) = |V_R(x,y) - (V_T(x-d_i(x,y),y))| \quad (6)$$

여기서,  $V_{Ref}(x,y)$ 와  $V_{Tar}(x,y)$ 는 각각 기준 영상과 대상 영상의 좌표  $(x,y)$ 의 Haar 웨이블릿 특징 벡터이고  $d_i(x,y)$ 는 변위이다. 이 때 변위  $d_i(x,y)$ 는 레벨  $j$ 에서 전과된 변위  $d_p(x,y)$ 부터  $(2d_p(x,y)+1)$ 까지 순차적으로 1씩 증가시켜 특징차를 가장 작게 하는 변위를 해당 좌표의 변위로 결정한다.

### 5. 실험 결과

제안한 방법의 성능을 검증하기 위해 Tsukuba 영상을 이용하여 실험하였다. 비교에 사용한 기존의 그래프 컷 기법들은 그래프 컷을 이용한 방법[2]과 칼라 세그먼트를 적용한 그래프 컷[3]이다. 각각의 변위도를 제안한 방법으로 얻은 변위도와 시각적으로 비교하였고 객관적인 성능 비교를 위해 비폐색 영역에 대한 오정합 화소 비율과 변위 불연속 영역의 오정합 화소 비율, 그리고 계산 시간을 계산하였다.

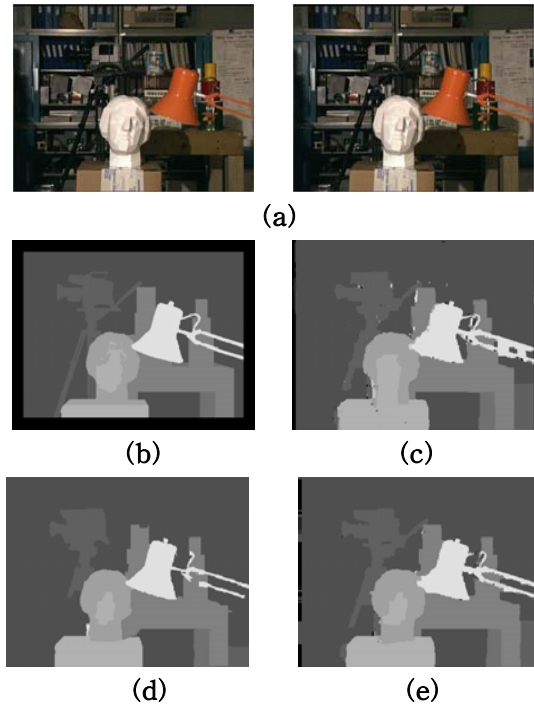
실험에 이용한 Tsukuba 좌, 우 영상은 그림 4의 (a)와 같다. 사실 변위도는 그림 4의 (b)이다. 기존의 그래프 컷으로 얻은 변위도는 그림 4의 (c)이고 칼라 세그먼트를 적용한 그래프 컷의 결과는 그림 4의 (d)이다. 그래프 컷만으로 얻은 변위도는 우측 스탠드 지지대 사이 공간의 변위가 제대로 반영하지 못했고 변위도 곳곳에 대응점을 찾지 못한 화소들이 보인다. 칼라 세그먼트를 적용한 경우 스탠드 지지대 부근의 오정합은 개선했으나 중앙의 캠코더에서 그 성능이 떨어지는 것을 볼 수 있다.

제안한 방법으로 얻은 변위도는 그림 4의 (e)이다. 앞선 변위도들보다 언급했던 영역의 경계선들이 비교적 잘 보전되었음을 알 수 있다. 이는 정합 비용 통합 과정에서 제안한 bilateral 필터와 그래프 컷 과정에서 적용한 경로 가중치로 인해 기존의 그래프 컷의 단점을 개선할 수 있음을 보여준다. 그리고 다해상도 처리 과정에서 최종 변위도를 얻기까지 에지와 지역 특정 정보를 이용하기 때문에 변위를 전과하는 과정에서 변위의 오차를 줄일 수 있음을 의미한다.

비폐색영역에서의 오정합 화소 비율은 그래프 컷이 1.94%, 칼라 세그먼트를 이용한 방법이 1.19%, 제안한 방법이 1.03%였다. 변위 불연속 영역에서는 각각 9.39%, 7.86%, 5.82%였다. 계산 시간은 2.31, 1.83, 0.66(sec) 소요되었다

### 6. 결론

본 논문에서는 bilateral 필터를 이용한 그래프 컷 기반의 다해상도 스테레오 정합 방법을 제안하였다. 영상 피라미드 기법을 이용하여 저해상도 영상과 에지 정보를 이용하여 적



(그림 4)

- (a) Tsukuba 영상
- (b) Tsukuba 사실 변위도
- (c) 그래프 컷의 결과
- (d) 칼라 세그먼트를 적용한 그래프 컷의 결과
- (e) 제안한 방법으로 얻은 결과

응적으로 정합 윈도우의 크기와 형태를 정의하였고, 변위도의 물체간 경계를 명확히 보전하기 위해 비용 통합 과정에서 bilateral 필터를 이용하였고 그래프 컷 수행과정에서 경로 가중치를 계산하여 적용하였다. 실험을 통해 제안한 방법이 경계선 영역이 뭉개지는 기존 그래프 컷 영역의 단점을 해결할 수 있음을 보였고, 다해상도 기법을 결합함으로써 bilateral 필터와 그래프 컷 방법의 계산 복잡도를 감소시킬 수 있음을 알 수 있었다.

### 참고문헌

- [1] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," International Journal of Computer Vision, pp. 7-42, 2002.
- [2] E. Gudis, G. van der Wal, S. Kuthirummal, and S. Chai, "Establishing Visual Correspondence from Multi-Resolution Graph Cuts for Stereo-Motion," IEEE 20th FCCM, pp. 29-32, 2012.
- [3] W. Chen, M. Zhang, and Z. Xiong, "Segmentation-based stereo matching with occlusion handling via region border constrains," pp. 70-76, CVIU 2009.