

# 축소된 초기 변위지도를 이용한 지역 기반 스테레오 정합의 결과 보정

고재런, 호요성  
광주과학기술원 정보통신공학부  
e-mail : jrko@gist.ac.kr, hoyo@gist.ac.kr

## Disparity Correction for Local Stereo Matching Using Down-sampled Initial Disparity Map

Jaeryun Ko, Yo-Sung Ho  
School of Information and Communications  
Gwangju Institute of Science and Technology (GIST)

### 요 약

일반적인 지역 기반 스테레오 정합 방법은 정합 블록의 크기에 따라 정합 결과가 변동한다. 적용되는 정합 블록의 크기가 작을수록 객체 경계 부분에서 정확한 결과를 얻을 수 있으나 잡음에 민감하다. 반면 정합 블록의 크기가 커질수록 잡음에 강하지만 경계 부분에서 부정확한 정합 결과가 나올 확률이 높다. 이 논문에서는 기존의 지역 기반 스테레오 정합 방법을 사용하되 축소된 초기 변위지도를 적용한 정합 방법을 제안했다. 우선 축소된 원본 영상에서 좌우측 변위지도를 구하고, 교차 검사 방법을 통해 신뢰도 높은 초기 변위지도를 획득하여 원본 크기로 확대한다. 원본 영상의 스테레오 정합 과정에서 변위 값 계산 시 오차 범위를 참조하고 최종 변위지도를 획득했다. 세 가지 유사도 측정 방법을 사용해 실험을 수행한 결과, 전체적으로 잡음 제거에 효율적이며 개선된 정합 결과를 얻었다.

### 1. 서 론

20 초 다양한 스테레오 카메라 개발로부터 시작하여 3차원 입체 영상에 대한 연구는 상당히 오랜 역사를 가지고 있다. 최근 수년간 3차원 영화 흥행과 3차원 비디오 게임 장치 출시 등으로 3차원 엔터테인먼트 분야는 많은 관심과 더불어 활발한 연구가 이루어지고 있으며 이 중심에 3차원 입체 영상 분야가 핵심으로 자리하고 있다. 이러한 3차원 입체 영상을 만들기 위해서는 동일한 장면을 여러 시점에서 촬영한 영상들이 필요하다. 이를 위해 다양한 시점에 위치한 카메라로 촬영하여 영상을 획득하거나 기존에 촬영한 적은 수의 영상을 활용하여 중간 시점 영상을 합성하는 방법이 있다.

중간 시점 영상을 합성하기 위해서는 촬영한 영상의 깊이 정보가 필요하며, 깊이 정보를 측정하는 방법을 크게 두 가지로 나눌 수 있다. 능동적 깊이 측정 방법은 깊이 카메라 등을 이용하여 깊이를 직접적으로 측정하는 방법이며, 수동적 깊이 측정 방법은 공통된 장면에 대해 다양한 시점에서 촬영한 영상들 사이에서 각각의 변위차를 계산하여 깊이를 측정하는 방법이다.

능동적 깊이 측정 방법은 적외선이나 레이저를 이용하기 때문에 정확한 깊이 정보를 획득할 수 있으나, 이러한 방식을 지원하는 특수한 카메라가 필요하며 측정범위의 제한이 존재하므로 야외에서 촬영 시 여러 가지 한계가 존재한다. 반면 수동적 깊이 측정 방법은 일반적인 다시점의 컬러 영상을 획득할 수 있다면 공간의 제약을 받지 않는다. 다만 객체가 가진 무늬에 의해 정확한 변위차를 측정하는데 어려움이 존재한다.

이러한 수동적 깊이 측정 방법 중 두 개 시점에서 촬영한 영상을 이용하는 방법인 스테레오 정합 방법이 보편적으로 사용되고 있으며 다양한 연구가 지속되고 있다. 스테레오 정합 방법은 지역 기반 방법과 전역기반 방법 두 가지로 나눌 수 있으며 각각 장단

점을 가지고 있다. 지역기반 스테레오 정합 방법은 두 영상의 화소 간 정합 비용을 계산하여 변위 값을 결정한다. 정합 비용을 계산할 때 유사도 측정 방법인 SAD(Sum of Absolute Differences), SSD(Sum of Squared Differences), NCC(Normalized Cross Correlation) 등을 사용 한다 [1]. 지역기반 스테레오 정합은 전역기반 방법에 비해 정합 속도가 빠르지만 정합 정확도가 떨어지는 단점이 존재한다. 전역기반 스테레오 정합 방법은 MRF(Markov Random Field) 기반의 에너지 함수를 최소화 시키는 값으로 적절한 변위 값을 추정한다. 에너지 최적화 방법은 신뢰확산이나 그래프 컷 등의 방법이 있으며 이는 영상 전체 화소 간 관계를 고려하여 계산이 이루어지므로 높은 복잡도로 인해 정합 속도가 상대적으로 느리다 [2], [3].

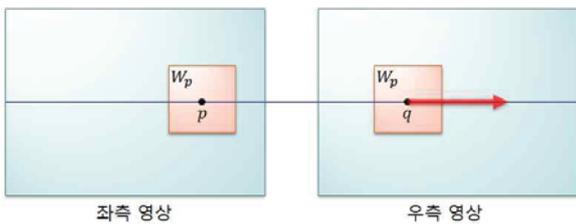
두 가지의 스테레오 정합 방법이 가진 각각의 장점과 단점을 극복하기 위한 많은 연구가 진행되고 있다. 이 논문은 지역 기반 스테레오 정합 방법에서 정합 블록의 축소에 따른 잡음 발생 문제를 해결하기 위해, 축소된 스테레오 영상에 대한 초기 변위지도를 이용하여 스테레오 정합 결과를 보정하기 위한 방법을 제안한다. 좌측 영상 기준 변위지도와 우측 영상 기준 변위지도의 교차 검사 방법을 통해 초기 변위지도의 신뢰성을 높이고 기존의 지역 기반 스테레오 정합 연산을 사용하여 결과가 개선되는 것을 확인하였다.

### 2. 지역기반 스테레오 정합 방법

지역기반 스테레오 정합 방법은 좌측 영상과 우측 영상 사이에 최대 유사도를 가진 화소 값을 찾기 위하여 정합 블록이라고 하는 화소 집합 단위로 정합비용을 계산한다. 집합 단위가 아닌 각각의 화소 값으로 정합 비용을 계산하게 되면 목표 화소 값과 유

값을 지닌 화소 값이 여러 위치에서 존재할 수 있으므로 모호한 정합 결과가 나타날 가능성이 매우 높다. 따라서 블록 단위의 화소 블록을 좌측 상단부터 우측 하단까지 수평 탐색선을 따라 이동시키며 중심 화소 값만이 아닌 주변의 화소 값들을 정합 비용 계산에 이용하는 방식으로 정합 정확도를 향상시킬 수 있다. 정합 블록의 크기는 3×3, 5×5 등 임의의 홀수 크기 정사각형으로 결정한다.

좌측 영상을 기준 영상으로 하여 우측 영상에 대해 지역기반 스테레오 정합을 수행하면 좌측 영상에 대한 변위지도도를 얻을 수 있다. 이때 그림 1과 같이 좌측 영상의 정합 블록  $W_p$ 을 고정시켰을 때 우측 영상의 정합 블록  $W_q$ 을 동일한 탐색선 상에서 이동시키며 정합 비용을 계산한다. 화소  $p = (x, y)$  위치에 대해서 화소  $q = (x - d, y)$  위치의 정합 블록이 최적의 정합 비용을 갖게 될 때 좌측 영상의 화소  $p$ 의 위치에서 변위 값  $d$ 를 갖는다.



(그림 1) 좌측 영상 기준의 스테레오 정합 과정

최적의 정합 비용을 계산하기 위해서 유사도 측정 방법을 사용하는데 이러한 예로 다음의 식 (1), 식 (2), 식 (3)은 각각 SAD, SSD, NCC 방법을 나타낸다.

$$SAD(d) = \sum_i \sum_j |I_L(x+d+i, y+j) - I_R(x+i, y+j)| \quad (1)$$

$$SSD(d) = \sum_i \sum_j [I_L(x+d+i, y+j) - I_R(x+i, y+j)]^2 \quad (2)$$

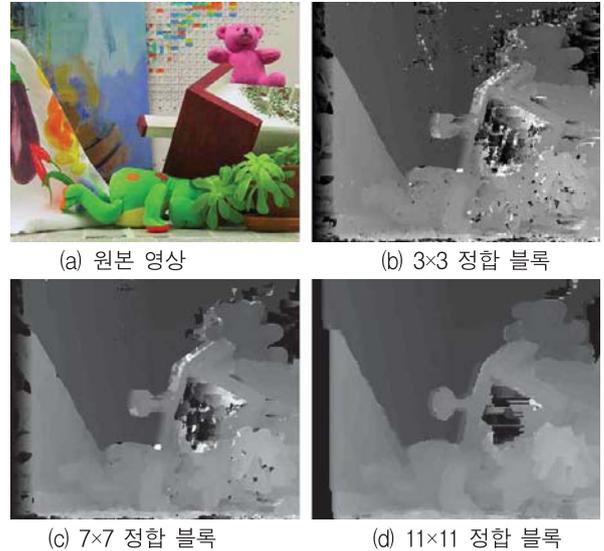
$$NCC(d) = \frac{Cov[I_L(x, y), I_R(x+d, y)]}{\sqrt{var_{I_L(x, y)} var_{I_L(x+d, y)}}} \quad (3)$$

위 식에서  $I_L, I_R$ 은 각각 좌측영상과 우측영상을 나타내며  $d$ 는 좌우 영상의 변위 값을,  $x$ 와  $y$ 는 영상 내 화소의 좌표 값을 나타낸다. 그리고  $i$ 와  $j$ 는 정합 블록 내의 정합 비용 탐색 범위를 표현한다. SAD와 SSD를 이용한 방법은 최소의 정합 비용을 갖는 위치가 최적 위치가 되며, NCC는 최대의 정합 비용을 갖는 위치가 최적 위치가 된다. 이 세 가지 방법의 장단점은 사용되는 영상이나 실험 환경에 따라 달라질 수 있으므로 특정한 방법이 뛰어나다고 말할 수 없다.

지역기반 스테레오 정합은 정합 과정에서 사용하는 정합 블록의 크기에 따라 변위지도도의 결과가 달라진다. 상대적으로 작은 크기의 정합 블록을 사용할 때는 영상 내의 객체가 가지는 다양한 패턴과 텍스처 등 세밀한 부분에 대해서 변위 값을 구하는 데 적절하지만 잡음 제거에는 약한 성능을 보인다. 반면에 상대적으로 큰 크기의 정합 블록을 사용할 시 비교적 객체의 경계 및 복잡한 패턴을 제대로 고려하지 못한다는 단점이 존재한다. 따라서 객체

경계 부분에 작은 크기의 정합 블록을 사용하고 나머지 부분에 대해서 큰 크기의 정합 블록을 사용하는 방법이 제시되었다.[4]

그림 2는 3×3, 7×7, 11×11 크기의 정합 블록을 이용하여 얻은 변위지도이다. 그림 2(b), 그림 2(c), 그림 2(d) 순으로 잡음이 사라지지만 원본 영상에 대조하여 확인 할 수 있는 객체 경계 부분은 두꺼워지거나 왜곡되고 있는 것을 확인할 수 있다.



(그림 2) 정합 블록 크기에 따른 변위지도 변화

표 1을 보면 정합 블록의 크기가 커질수록 전체와 텍스처 영역에 대한 BPR은 감소하지만 경계 영역에 대해서는 BPR이 증가하는 것을 확인할 수 있다.

(표 1) 정합 블록 크기에 따른 Bad Pixel Rate(BPR)

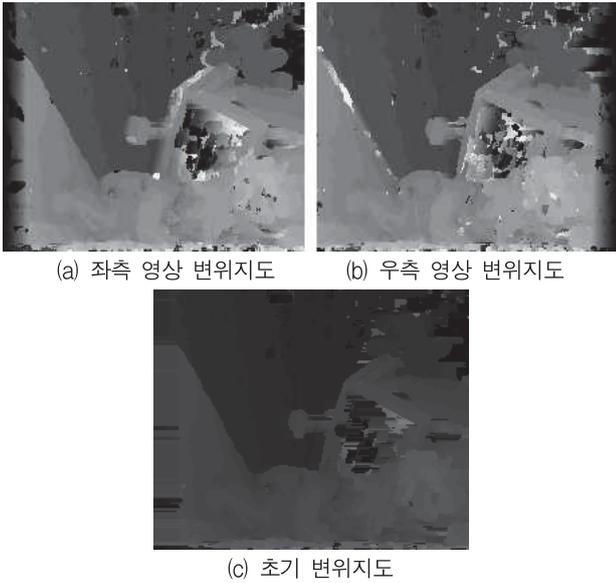
Block Size	BPR(%)			
	All	Non-occlusion	Textured	Discontinuity
3×3	31.34	23.77	20.56	33.64
7×7	27.75	19.84	17.40	33.59
11×11	27.29	19.35	17.11	36.75

### 3. 축소된 초기 변위지도를 이용한 정합 보정

그림 2와 표 1을 통해, 정합 블록의 크기가 커질수록 잡음에 대한 오류는 감소하나 경계 부분과 전체적인 BPR 값을 개선하는데는 효과적이지 못하다는 것을 알 수 있다. 때문에 이를 보완하기 위하여 정합 블록의 크기를 변경하여 얻은 변위지도도를 최종 변위지도도의 품질 개선에 사용하는 대신 원본 영상을 축소하여 얻은 초기 변위지도도를 이용하여 원본 영상에 대한 변위 값 계산 결과를 보정하는 방법을 제안한다.

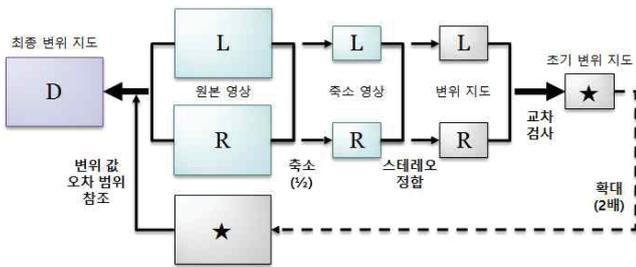
제안한 방법은 좌우 영상의 폭과 높이를 절반으로 축소한 영상으로부터 초기 변위지도도를 획득하는 것으로 시작한다. 좌우 원본 영상을 절반으로 축소한 영상들에 지역 기반 스테레오 정합을 실시하여 그림 3(a) 및 그림 3(b)과 같이 좌측 영상 기준 변위지도와 우측 영상 기준 변위지도도를 획득한다. 이렇게 획득한 좌측 변위지도에 대해 우측 변위지도가 가진 변위 값이 서로 일치하는지를 확인하는 교차 검사 방법을 실시한다 [5]. 이는 좌우 영상에 대해

변위 값의 오류를 상당 부분 상쇄시킴으로써 그림 3(c)과 같이 신뢰도 높은 초기 변위지도 생성한다.



(그림 3) 교차 검사를 통한 초기 변위지도 생성

초기 변위지도는 원본 영상에 비해 절반 크기이므로, 다시 2배로 확대하여 원본 영상에 대한 스테레오 정합 과정에서 참조할 수 있도록 준비한다. 원본 영상에 대하여 스테레오 정합을 실시하여 변위지도 생성 중 현재 좌표에서의 변위 값을 결정할 때, 초기 변위지도에서 동일한 좌표가 가지는 변위 값을 참조한다. 여기에 임의로 설정한 오차 범위를 적용하여 보다 정확한 변위 후보 내에서 변위 값을 결정하도록 한다. 이러한 과정을 통해 최적의 정합 비용을 갖는 변위 값을 획득하여 최종 변위지도를 획득한다. 이러한 전체 알고리즘 흐름은 그림 4와 같다.



(그림 4) 제안한 방법의 전체 흐름도

#### 4. 실험 결과

제안한 알고리즘을 실험하기 위해서 Middlebury Stereo Dataset에서 제공하는 Cones, Teddy, Tsukuba, Venus로 총 4개의 영상을 실험 대상으로 사용하였다 [6]. 3×3 크기로 고정된 정합 블록을 사용하고 초기 변위지도로부터 최대 ±5의 변위 오차 범위를 설정하여 지역기반 스테레오 정합을 실시하였다. 그리고 기존 SAD, SSD, NCC를 통해 정합 비용을 계산한 방법과 이들에 대해 제안한 알고리즘을 적용한 방법을 비교하였다.

그림 5와 그림 6은 4개의 실험 영상 중 Teddy와 Cones의 결과를 각각 나타낸 것이다. 그림 5(b), 그림 5(d)와 그림 6(b), 그림 6(d)는 각각 기존의 SAD와 SSD를 이용한 결과로 서로 큰 차이를

보이지 않는다. 하지만 그림 5(f)와 그림 6(f)는 기존의 NCC를 이용한 결과이며, 이는 SAD와 SSD보다 뚜렷한 차이를 보인다. 특히 다수의 미세한 잡음이 훨씬 두드러지게 나타났다. 제안한 방법은 각각 그림 5(c), 그림 5(e), 그림 5(g)와 그림 6(c), 그림 6(e), 그림 6(g)로 기존의 방법보다 잡음이 뚜렷하게 제거되는 효과를 보인다. 특히 미세한 잡음이 많이 나타나는 기존 NCC 결과에 대해서는 나머지 결과들에 비해 상대적으로 각각 그림 5(a)와 그림 6(a)에 더 가까운 상태를 보여준다.

(표 2) 실험 영상의 해상도

Test Images	Resolution
Cones	450×365
Teddy	450×375
Tsukuba	384×288
Venus	434×383



(a) Ground Truth



(b) 기존 SAD 결과



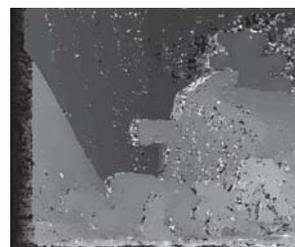
(c) 제안한 방법의 SAD 결과



(d) 기존 SSD 결과



(e) 제안한 방법의 SSD 결과

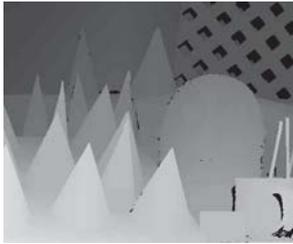


(f) 기존 NCC 결과



(g) 제안한 방법의 NCC 결과

(그림 5) Teddy 실험결과 영상



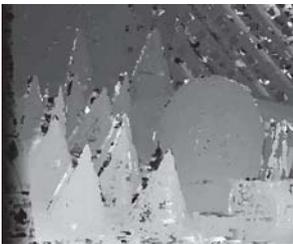
(a) Ground Truth



(b) 기존 SAD 결과



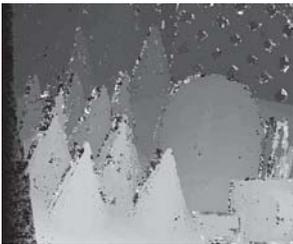
(c) 제안한 방법의 SAD 결과



(d) 기존 SSD 결과



(e) 제안한 방법의 SSD 결과



(f) 기존 NCC 결과



(g) 제안한 방법의 NCC 결과

(그림 6) Cones 실험결과 영상

(표 3) 폐색 제외 영역에 대한 BPR 결과 값 비교

Test Images	BPR_Non-occlusion(%)					
	SAD	Proposed with SAD	SSD	Proposed with SSD	NCC	Proposed with NCC
Teddy	23.77	23.06	22.37	21.83	16.34	13.28
Cones	21.82	20.47	18.18	16.99	10.07	9.91
Tsukuba	13.44	12.39	13.52	12.54	18.47	16.08
Venus	14.68	13.82	13.65	12.67	13.30	10.96
Average	18.43	17.43	16.93	16.01	14.55	12.56

(표 4) 전체 영역에 대한 BPR 결과 값 비교

Test Images	BPR_All(%)					
	SAD	Proposed with SAD	SSD	Proposed with SSD	NCC	Proposed with NCC
Teddy	31.34	30.62	30.11	29.53	24.69	21.73
Cones	30.75	29.43	27.60	26.37	20.43	20.07
Tsukuba	15.28	14.22	15.39	14.41	20.17	17.74
Venus	16.09	15.24	15.09	14.11	14.74	12.42
Average	23.37	22.38	22.05	21.11	20.01	17.99

2는 각 실험 영상의 해상도를 나타내며 Tsukuba의 경우 다른 세 실험 영상보다 확연히 낮은 크기의 해상도를 가지고 있다. 이 사실을 고려하여 표 3과 표 4를 통해 볼 수 있는 폐색 제외 영역 및 전체 영역에 대한 BPR 결과 값 비교를 분석하면, Tsukuba 영상에 대해서만 유독 NCC 방법이 SAD와 SSD보다 BPR이 높다는 사실을 알 수 있다. 하지만 나머지 세 영상에 대해서는 NCC 방법이 상대적으로 좋은 결과를 보이며 제안한 방법들을 통해서 기존의 방법보다 전반적으로 개선된 수치를 얻었다.

### 5. 결론

기존의 지역기반 스테레오 정합 방법은 정합 블록의 크기에 따라 다양한 변위지도들을 생성한다. 이 논문은 정합 블록 크기를 변경하여 적용하는 대신, 축소된 원본 영상에 대한 스테레오 정합과 교차 검사를 통해 획득한 초기 변위지도들을 참조하여 원본 영상의 스테레오 정합 결과를 보정, 개선된 변이 지도를 얻도록 하였다. 실험결과 전체 영역에 대하여 기존의 지역기반 방식의 스테레오 정합 방법을 사용하였다. 이를 통해 기존 방식에 비해 전체 영역과 폐색 제외 영역에 대해서 약 2%의 개선된 BPR 값을 얻었다. 또한 잡음 제거에 강한 특성을 지니며 NCC 방식의 유사도 측정을 사용할 경우 BPR 감소가 두드러졌다.

### 감사의 글

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

### 참고 문헌

[1] 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.

[2] 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.

[3] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," International Conference on Computer Vision, pp. 508-515, 2001.

[4] Y. J. Chang and Y. S. Ho, "Modified SAD Using Adaptive Window Sizes for Efficient Stereo Matching," International Conference on Embedded Systems and Intelligent Technology, pp. 9-11, 2014.

[5] W. S. Jang and Y. S. Ho, "Disparity Map Refinement using Occlusion Handling for 3D Scene Reconstruction," International Conference on Embedded Systems and Intelligent Technology, pp. 213-216, 2011.

[6] Middlebury Stereo Datasets, <http://vision.middlebury.edu/stereo/data/>