

## 신뢰도 높은 변이추정을 위한 하이브리드 스테레오 정합 알고리즘

김득현 최진욱 오창재 손광훈

연세대학교 전기전자공학과

khsohn@yonsei.ac.kr

## Hybrid Stereo Matching Algorithm for Reliable Disparity Estimation

Kim, Deukhyeon Choi, Jinwook Oh, Changjae Sohn, Kwanghoon

School of Electrical and Electronic Engineering

Yonsei University

## 요약

본 논문에서는 다양한 변이 추정 방식 중 영역기반(Area-based) 알고리즘과 특징기반(Feature-based) 알고리즘을 결합한 하이브리드(Hybrid) 변이추정 알고리즘을 제안한다. 제안하는 알고리즘은 Features from Accelerated Segment Test (FAST) 코너 점 추출기[2]를 이용하여 좌, 우 영상 각각의 특징 점을 추출한 후, 특징 점들의 정보를 이용한 스테레오 정합을 통해 신뢰도 높은 초기 변이지도(Disparity map)를 생성하게 된다. 그러나 생성된 초기 변이지도는 조밀하지 못하므로, 조밀한 변이지도를 획득하기 위해 특징점이 추출된 영역에 대해서는 추정된 초기 변이 값을 이웃 픽셀과의 색 유사도를 고려하여 전파시키고, 특징 점이 추출되지 않은 영역에 대해서는 이전 윈도우(Binary window)를 활용한 영역기반 변이추정 알고리즘[1]을 이용하여 변이 값을 추정한다. 이를 통해, 제안 알고리즘은 특징 기반 알고리즘에서 발생할 수 있는 보간법 문제를 해결함과 동시에 신뢰도가 높은 초기 변이지도를 사용함으로써, 영역 기반 알고리즘의 정합 오차를 줄여 신뢰도 높은 변이지도를 생성할 수 있다. 실험 결과 추정된 초기 변이지도는 ground truth와 비교 시 약 99%이상의 정확도를 보이며, 특징 점이 추출된 영역에서 기존의 영역기반 알고리즘보다 더 정확한 변이 값이 추정되었음을 확인하였다.

## 1. 서론

최근 들어 3D 방송, 3D 영화 등 실감형 미디어에 대한 관심이 늘 어남에 따라 3D 콘텐츠의 품질과 다양성이 주요 이슈로 대두되고 있다. 고품질의 3D 콘텐츠 제작에 있어서 무엇보다 중요한 것은 정확한 깊이정보를 획득하는 것이며, 일반적으로 스테레오 정합 알고리즘이 널리 사용되고 있다. 스테레오 정합 알고리즘은 크게 지역적인 방법(Local method)과 전역적인 방법(Global method)으로 분류할 수 있으며, 지역적인 방법은 다시 영역기반(Area-based) 알고리즘과 특징기반(Feature-based) 알고리즘으로 분류할 수 있다.

영역기반 알고리즘은 일정 크기의 정합 윈도우를 형성하여 탐색 범위 안에 존재하는 모든 위치에 대해 Sum Absolute Difference (SAD), Sum Squared Difference(SSD), Normalized Cross Correlation(NCC) 등을 통한 추정 오차를 계산한 후, 오차가 가장 적은 윈도우의 위치를 정합 위치로 선택하여 변이 값을 추정한다. 이러한 영역기반 알고리즘에 의한 변이추정 결과는 조밀한 변이지도를 생성할 수 있지만, 정합 윈도우의 크기에 따라 계산복잡도와 품질이 좌우되는 단점이 있다.

특징기반 알고리즘은 좌, 우 영상에서 추출된 특징 점들 사이에서만 추정 오차를 계산하여 변이 값을 추정하게 된다. 이러한 특징 기반 알고리즘은 신뢰도 높은 변이 값을 빠르게 추정할 수 있지만 추출된

특징 점에 대해서만 변이 값이 구해지므로, 특징 점이 추출되지 않은 영역에 대해서는 단순히 보간법을 통해 변이 값을 추정 하게 된다.

전역적인 방법은 스테레오 정합 문제를 에너지에 대한 식으로 표현한 뒤, 그것의 전역적 최솟값(global minimum)을 구하여 변이지도를 생성하게 된다. 이러한 방법은 고품질의 변이지도를 생성할 수 있지만, 계산복잡도가 매우 높다는 단점이 있다. 전역적인 방법의 예로는 Belief Propagation(BP)와 Graph cuts 등이 있다.

이와 더불어, 최근 Ground Control Points(GCP)를 이용한 변이추정 알고리즘 [3][4] 또한 활발히 연구되고 있다. GCP를 이용한 알고리즘들은 신뢰도 높은 변이추정이 가능하지만, GCP를 구하는 과정에서 시간이 매우 오래 걸린다. [3]에서는 서로 다른 세 종류의 영역기반 변이추정 알고리즘을 이용하여 각각의 변이지도를 계산한 후에, 각 픽셀에서의 변이 값을 비교하여 신뢰도가 가장 높은 픽셀만을 GCP로 선택하는 알고리즘을 제안하였다. 또한, [4]에서는 탐색방향으로의 영역분할(segmentation)을 통해 초기 변이지도를 계산한 후, 가장 신뢰도가 높은 픽셀을 GCP로 선택하는 알고리즘을 제안하였다.

본 논문에서는 신뢰도 높은 변이지도를 빠르게 생성하기 위하여 특징기반 알고리즘을 통해 GCP 계산 시간을 단축시키고, 영역기반 알고리즘을 통해 그 값을 주변으로 전파시켜 조밀하면서도 신뢰도가 높은 변이지도를 생산함으로써, 특징기반 알고리즘과 영역기반 알고리즘의 장점을 결합한 하이브리드 변이추정 알고리즘을 제안한다.

논문의 구성은 2장에서 특징기반 변이추정 방식과 영역기반 변이 추정 방식을 결합한 하이브리드 변이추정 방식을 설명한다. 3장에서는 제안 알고리즘과 기존 특징 및 영역기반 알고리즘간의 비교 결과를 보여주고, 4장을 통해 결론을 맺는다.

## 2. 본론

본 논문에서는 기존 방법들의 단점을 극복하고자 특징기반 변이 추정 알고리즘을 이용하여 초기 변이지도를 생성하고, 이를 바탕으로 영역기반 변이추정 방법을 수행함으로써 신뢰도 높은 변이지도를 빠른 시간 안에 계산할 수 있는 하이브리드 변이추정 알고리즘을 제안한다. 이때 제안하는 알고리즘을 통해 획득한 초기 변이 값은 정확하다고 가정한다.

알고리즘 수행 과정은 다음과 같이 3단계로 나눌 수 있다. 1) 추출된 특징 점들의 변이 값을 추정하여 신뢰도 높은 초기 변이지도(GCP)를 획득한 후에, 2) 특징 점이 추출된 영역의 색상 정보를 이용하여 초기 변이 값을 이웃 픽셀로 전파시킨다. 3) 특징 점이 추출되지 않은 영역에 대해서는 [1]의 방법으로 변이 값을 추정하여 조밀한 변이지도를 생성한다.

영상 내의 특징 점을 추출하기 위한 대표적인 알고리즘으로는 Scale-Invariant Feature Transform(SIFT)[5], Speeded Up Robust Features(SURF)[6], FAST 코너 검출기[2] 등이 있다. 검출 성능 면에서는 [5][6]이 비교적 우수하지만, 검출 시간이 오래 걸린다는 단점이 있다. 따라서 제안하는 알고리즘에서는 특징 점 추출시간을 줄이고자 [2]에서 제안한 알고리즘을 이용하였다. 위와 같은 방법으로 추출된 특징 점들은 조명 변화에 강인한 Census transform을 통해 표현자로 나타낼 수 있다. 좌, 우 영상의 표현자를 탐색 방향으로 비교한 뒤, 정합 오차가 가장 작은 윈도우의 중심 픽셀을 정합 위치로 선택하여 각 특징 점들의 변이 값을 추정한다. Census transform은 식 (1)과 같이 표현할 수 있으며, 그림 1은 그 예를 보여주고 있다[7].

$$Census(p) = \bigotimes_{q \in w} T(I(p), I(q)) \quad (1)$$

$$T(\alpha, \beta) = \begin{cases} 1, & \text{where } \alpha > \beta \\ 0, & \text{otherwise} \end{cases}$$

⊗ 은 연속(concatenation)을 나타내며, q는 p 픽셀의 이웃 픽셀을 의미한다. 또한, I(p)와 I(q)는 p, q 픽셀에서의 밝기(intensity)이다.

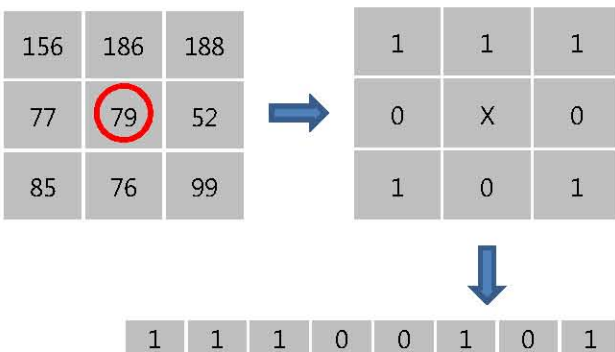


그림 1. Census Transform의 예

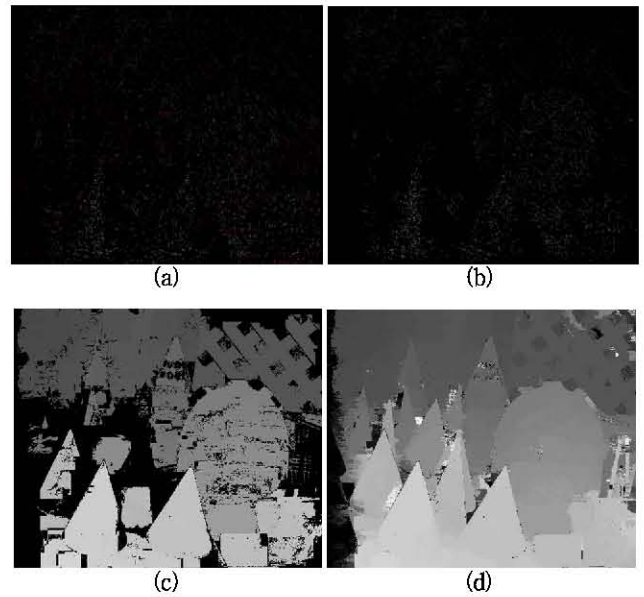


그림 2. 제안 알고리즘의 변이추정 과정  
(a) 필터링 전 초기 변이지도, (b) 필터링 후 초기 변이지도  
(c) 초기 변이 값 전파 후의 변이지도, (d) 최종 변이지도

추정된 초기 변이지도는 폐색 영역(occlusion region) 보정을 위해 상호 참조(cross-check)를 수행한다. 또한, 초기 변이지도의 신뢰도를 높이기 위해 식 (2)로 일정 크기의 윈도우 안에 있는 초기 변이 값의 빈도수를 계산한 뒤, 식 (3)에 의해 필터링 되어 신뢰도 높은 초기 변이 지도를 생성하게 된다. 그림 2(a)는 필터링 전의 초기 변이지도이며, 그림 2(b)는 필터링 후의 초기 변이지도이다. 빨간 점은 잘못 추정된 변이 값을 의미한다.

$$F(p) = \sum_{q \in w, q \neq p} A(D_i(p), D_i(q)) \cdot C_i(q) \quad (2)$$

$$A(\alpha, \beta) = \begin{cases} 1, & \text{where } |\alpha - \beta| < \lambda_{th} \\ 0, & \text{otherwise} \end{cases}$$

$$C_i(p) = \begin{cases} 1, & \text{if } F(p) \geq 1 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

F는 윈도우 내의 초기 변이 값의 빈도수를 의미하며, C<sub>i</sub>(p)는 좌 영상 내의 특징 점 위치(p)에서의 신뢰도를 의미한다. 만약 C<sub>i</sub>(p)의 값이 1이면, 해당 위치의 특징 점은 GCP 후보임을 의미한다. D<sub>i</sub>(p), D<sub>i</sub>(q)는 p, q위치에서 추정된 변이 값이며, A(α, β)는 두 변이 값의 차이를 의미한다.

그림 2(b)에서 알 수 있듯이 초기 변이지도는 조밀하지 못하다는 문제점이 있다. 이러한 문제점을 해결하기 위하여 본 논문에서는 [1]에서 제안한 알고리즘을 활용하였다. [1]에서는 윈도우 내의 색상 정보가 유사한 픽셀들을 하나의 집합으로 묶어 다양한 크기와 모양을 갖는 이진 윈도우를 생성한 뒤, 같은 집합에 속하는 픽셀들은 모두 동일한 변이 값을 대입하여 조밀한 변이지도를 생성하는 알고리즘을 제안하였다. 이때, 이진 윈도우는 식(4)와 같이 정의할 수 있다. 이에 따라 제안 알고리즘에서는 인접한 픽셀들 간의 색상 정보가 유사한 경우 동일한

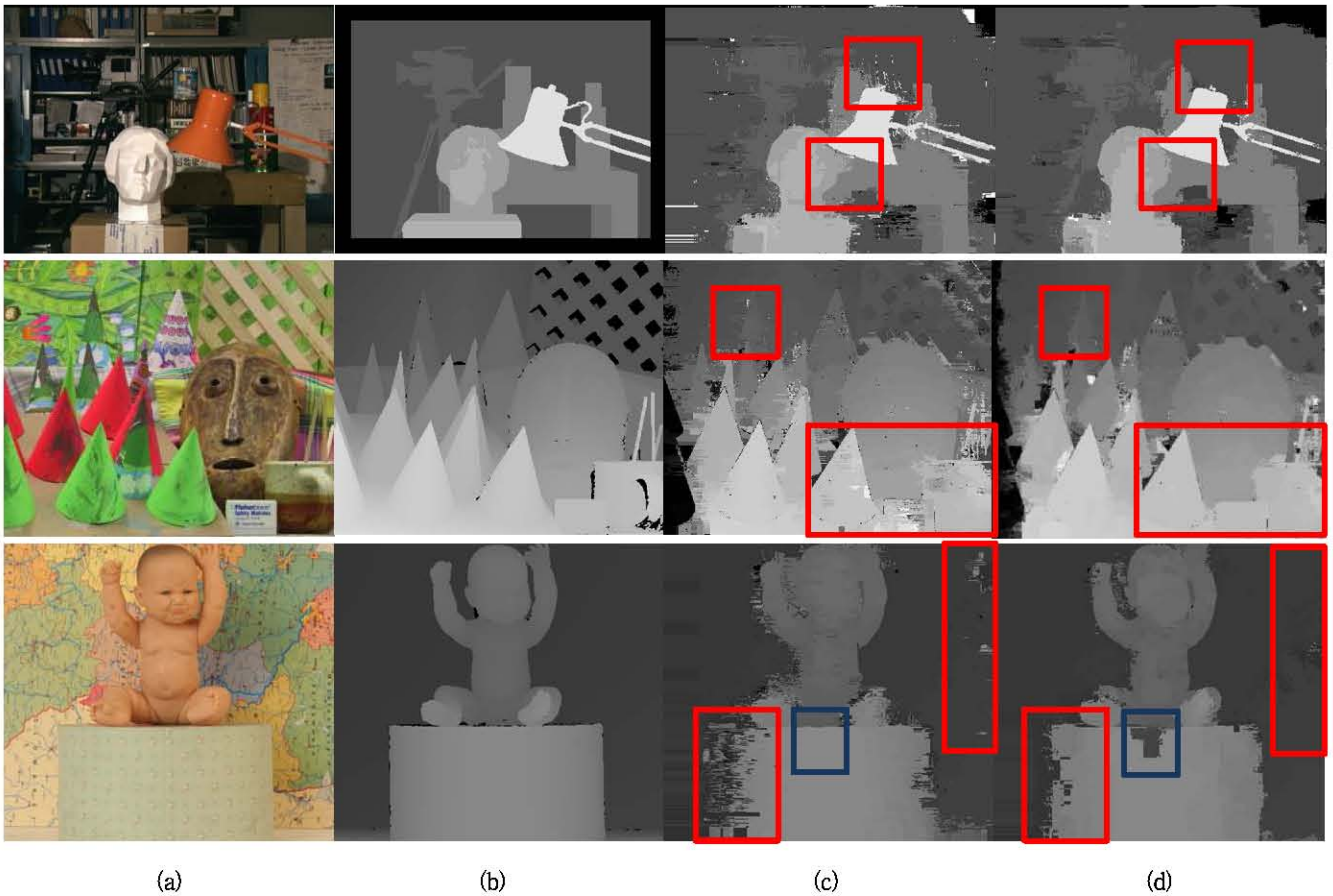


그림 3. 변이추정 결과 비교 (a) 기준 영상, (b) Ground truth, (c) [1]의 알고리즘 결과, (d) 제안 알고리즘 결과

물체로 판단하여 변이 값의 변화가 적다는 가정을 통해, 추정된 초기 변이 값을 중심으로 하는 이진 윈도우를 생성한 뒤, 같은 집합에 속하는 픽셀들에게 그 값을 전파시켜 그림 2(c)와 같이 부분적으로 조밀한 (semi-dense) 변이지도를 생성한다. 특이점이 추출되지 않은 영역에 대해서는 위의 과정으로 변이 값이 추정되지 않으므로, 이러한 영역에 대해서는 [1]의 알고리즘을 통해 변이 값을 추정한다. 이때, 일부 픽셀들은 여러 집합에 속하게 되어 모호한 변이 값을 갖게 된다. 이러한 픽셀들은 주로 물체의 경계면에 위치해 있으므로, 이진 윈도우의 크기를 절반으로 줄여 해당 픽셀에 대해서만 동일한 방법을 다시 적용하여 최종적인 변이지도를 획득하게 된다. 그림 2(d)는 제안 알고리즘을 통해 생성한 최종 변이지도이다.

$$W(p,q) = \begin{cases} 1, & \text{where } d_{pq} < \lambda_{th} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$d_{pq} = \sqrt{(L_p - L_q)^2 + (a_p - a_q)^2 + (b_p - b_q)^2}$$

식 (4)에서  $W$ 는 이진 윈도우이며,  $q$ 는  $p$ 픽셀의 이웃 픽셀을 의미한다. 또한,  $d_{pq}$ 는 Lab 색 공간에서의 색상 유사도를 나타내는 척도이며,  $\lambda_{th}$ 는 경계치를 의미한다.

### 3. 실험 결과

제안한 알고리즘은 Middlebury data sets[8] 중 Tsukuba, Cone,

|         | Total feature points | Error | Accuracy Rate(%) |
|---------|----------------------|-------|------------------|
| Cone    | 2894                 | 489   | 83.10            |
| Tsukuba | 1863                 | 297   | 84.06            |
| Baby1   | 2340                 | 198   | 91.54            |

표 1. 필터링 전 초기 변이지도의 정확도

|         | Total feature points | Error | Accuracy Rate(%) |
|---------|----------------------|-------|------------------|
| Cone    | 2250                 | 19    | 99.16            |
| Tsukuba | 673                  | 2     | 99.70            |
| Baby1   | 1541                 | 6     | 99.61            |

표 2. 필터링 후 초기 변이지도의 정확도

Baby1을 이용하여 실험하였다. 획득된 초기 변이지도(GCP)는 오차 경계치(error threshold)를 1로 하여 ground truth와의 오차를 계산해 보았을 때, 필터링 전에는 표 1에서와 같이 83%~91%의 정확도는 보인 반면, 필터링 후에는 표 2에서 보듯이 약 99% 이상의 정확도를 가짐을 확인하였다. 또한 그림 3에서 보듯이, 특징 점이 추출된 영역에서 [1]의 방법으로 획득한 변이지도보다 더 정확한 변이 값이 추정되었음을 확인하였다. 그림 3의 (a)와 (b)는 실험 영상 각각의 기준 영상 (reference image)과 ground truth 이다. 또한, (c)는 [1]에서 제안하는

알고리즘으로 추정된 변이지도이며, (d)는 본 논문에서 제안하는 알고리즘으로 구한 변이지도이다. 공정한 결과 비교를 위해 (c)와 (d) 모두 후처리(post-process)를 하지 않았으며, 빨간 영역은 신뢰도 높은 초기 변이지도를 이용함으로써, 추정된 변이 값의 정확도가 향상된 영역을 표시한 것이다. 하지만 파란 영역에서 알 수 있듯이, 초기 변이 값이 잘못 구해지게 되면 영역기반 변이추정 알고리즘을 수행 시 그 값이 주변으로 전파되어 결국 최종 변이지도에서도 오류로 남아있게 되는 문제점이 있다.

#### 4. 결론

기존의 변이추정 알고리즘 중 특징기반 변이추정 알고리즘은 빠르고 신뢰도가 높지만, 조밀하지 못한 변이지도를 생성한다는 특징이 있으며, 영역기반 변이추정 알고리즘은 조밀한 변이지도를 생성하지만, 계산 시간과 정합 오류가 정합 윈도우의 크기에 좌우된다는 특징이 있다. 따라서 본 논문에서는 각각의 장점을 결합한 하이브리드 변이추정 알고리즘을 제안하였으며, 기존 알고리즘에 비해 신뢰도가 높은 변이지도가 생성됨을 확인하였다. 하지만 잘못 추정된 초기 변이 값이 최종 변이지도 추정에 심각한 문제를 야기할 수 있다는 단점이 존재한다. 이를 해결하기 위해 초기 변이지도의 정확도를 높이는 연구를 추가적으로 진행할 계획이며, 더 나아가 초기 변이 값을 이용하여 영역기반 알고리즘의 탐색범위(search range)를 줄일 수 있는 알고리즘 연구를 진행할 수 있을 것이다.

#### 참고문헌

- [1] R. K. Gupta, and S.Y. Cho, "Real-time Stereo Matching using Adaptive Binary Window," *Proceedings of 3DPVT*, 2010.
- [2] E. Rosten, and T. Drummond, "Machine learning for high-speed corner detection," *Proceedings of ECCV*, 2006.
- [3] L. Wang, R. Yang, "Global stereo matching leveraged by sparse ground control points," *IEEE Proc. CVPR* 2011.
- [4] X. Sun, X. Mei, S. Jiao, M. Zhou, and H. Wang, "Stereo Matching with Reliable Disparity Propagation," *Proceedings of 3DIMPVT*, 2011.
- [5] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* vol. 60, pp. 91-110, 2004.
- [6] H. Bay, T. Tuytelaars, L. Van Gool, "Surf: Speeded up robust features," *Proceedings of ECCV*, 2006.
- [7] R. Zabih, J. Woodfill, "Non-parametric Local Transforms for Computing Visual Correspondence," *Proceedings of ECCV*, 1994.
- [8] <http://vision.middlebury.edu/stereo/data/>