

변위 공간상의 정합값 분포를 이용한 새로운 스테레오 정합 알고리즘

김재철*, 이경무**, 이상욱*

*서울대학교 전기컴퓨터공학부, **홍익대학교 전기전자공학과

A New Stereo Matching Algorithm Using Distribution of Match Values in Disparity Space

Jae Chul Kim*, Kyoung Mu Lee**, Sang Uk Lee*

*School of Electrical Engineering, Seoul National University

**Department of Electronics and Electrical Engineering, Hong-Ik University

Email : jckim@diehard.snu.ac.kr

요약문

본 논문에서는 영상을 특성에 따라 국부 영역으로 분류하고 변위 공간(disparity space)상에서의 특징을 분석하여 각각의 영역에 적합한 윈도우의 크기를 정하는 새로운 스테레오 정합 기법을 제안한다. 일반적으로 텍스처(texture)가 적은 영역이나 텍스처가 반복되는 영역, 그리고 깊이의 불연속선상에서는 고정된 크기의 윈도우를 사용하는 영역 기반 스테레오 기법은 잘 동작하지 않는다. 본 논문에서는 이러한 영역들의 변위 공간상에서의 정합 값 분포를 분석하여 스테레오 정합에 이용한다. 실험은 변위의 참값이 알려진 영상에 대해서 수행되었으며 기존의 방법에 비해 짧은 수행 시간 및 정확한 정합 결과를 보여 준다.

1. 서론

복수의 영상으로부터 3 차원 정보를 추출하는 기법은 컴퓨터 비전 분야에서 매우 중요한 위치를 차지하고 있다. 특히 두 장 이상의 영상을 이용하는 스테레오 정합 기법은 저렴한 비용에 비하여 양질의 3 차원 정보를 추출할 수 있기 때문에 이에 관하여 많은 연구가 수행되어 왔다 [1][2][3][4][5][6]. 여러가지 스테레오 정합 기법중에서 영역 기반의 스테레오 기법은 조밀한 변위 지도(dense disparity map)를 산출해 낼 수 있다. 이 때 변위 지도는 연속인 곡면에 대해서는 부

드러운 변위 값을 내면서도 그 경계선이 명확히 나타나야 한다. 그러나 고정된 크기의 윈도우를 사용하는 스테레오 기법에서는 이러한 두 가지 조건을 동시에 만족시키기가 쉽지 않다. 일반적으로 부드러운 변위 값을 낼 수 있는 경우에는 곡면의 경계선이 무너지고, 곡면의 경계선을 잘 살리는 경우 변위 값에 많은 잡음이 발생하게 된다. 이것은 잡음이 적은 변위 값을 위해서는 크기가 큰 윈도우가 필요하지만 경계선을 살리기 위해서는 작은 크기의 윈도우가 필요하기 때문이다.

이러한 문제를 해결하기 위해 많은 기법들이 제안되었는데, 이들은 영역에 따라 윈도우의 모양이나 크기를 바꾸는 적응적 윈도우 기법과 정합 과정에 연속성 및 유일성 제한을 전역적으로 적용하여 최적의 해를 구하는 기법으로 나누어 볼 수 있다. 전자의 대표적인 알고리즘이 Kanade 와 Okutomi 의 기법 [1]이며, 후자의 대표적인 알고리즘이 Marr 와 Poggio [3]에 의해 제안되고 Zitnick 과 Kanade [4][5]에 의해 보완된 협력적 스테레오 기법이다. 또한 최근에는 윈도우의 크기 뿐만 아니라 모양까지도 바꾸는 콤팩트 윈도우(compact windows)기법 [2]과, 최소 에너지를 구하는 방법의 일종인 그래프 컷(graph-cuts)방법을 적용하여 최적의 해를 구하는 기법 [6][7]이 소개되고 있다.

적응적 윈도우 기법들 [1][2]은 윈도우 내에서의 밝기값 변화와 각 단계에서 구해진 변위값 변화를 측정하여 최소의 불확실성을 보이도록

윈도우의 모양 및 변위값을 갱신한다. 적응적 윈도우 기법들은 갱신 규칙이 수학적으로 명료하게 유도되어 있기는 하지만 산출되는 결과는 고정 윈도우 방법에 비해 우수하지 못하다. 이는 알고리즘이 초기 추정치에 매우 민감하기 때문이다. 또한 적응적 윈도우 기법은 계산량이 많다는 단점이 있다. 전역적 제한을 이용하여 최적의 해를 구하는 방법들은 상대적으로 정확한 정합 결과를 보인다. 그러나 협력적 스테레오 기법[5]의 경우 텍스처가 적은 영역에서는 부정확한 결과를 보이며 깊이 불연속선 부근이 뭉개어지는 문제를 보인다. 한편 그래프 컷 기법들[6][7]은 다른 기법들에 비해 나은 정합의 결과를 보이지만 계산량이 많다는 단점을 보인다.

본 논문에서 제안하는 기법은 변위 공간 상에서의 초기 정합값의 분포를 분석하여 영역의 특징을 파악하고 그에 따라 영역의 특징에 맞는 윈도우의 크기를 정한다. 영역 기반 스테레오 기법은 텍스처가 적은 영역이나 텍스처가 반복되는 영역, 그리고 깊이의 불연속선 부근에서 일반적으로 정합 오차를 보인다. 한편 영상에 잡음이 존재하는 경우에도 정확하지 못한 결과를 보이게 된다. 그러나, 이러한 영역들의 변위 공간 상에서의 정합값 분포는 일정한 특징을 보인다. 이는 정합값의 분포가 정합 영역의 밝기값의 변동을 내포적으로 반영하고 있기 때문이다. 따라서 이러한 특징은 밝기값 차이만을 구해 정합을 하는 고정된 크기의 윈도우기법으로도 얻어질 수 있다. 따라서 고정된 크기의 윈도우로 초기 정합을 하여 영역의 특징을 내는 과정은 초기 정합에 민감하지 않다는 장점을 가진다. 초기 정합을 수행하면 영역은 정합값 분포의 특징에 따라 신뢰 영역, 잡음 영역, 에지 영역, 텍스처가 적은 영역, 텍스처가 반복되는 영역, 가리어짐 영역의 여섯 가지로 분류된다. 이러한 분류에 따라 적절한 정합이 수행되어 변위값 및 가리어짐 영역이 최종적으로 구해진다.

본 논문의 구성은 다음과 같다. 제 2 장에서는 스테레오 정합과 변위 공간에 관하여 간략히 설명하며 제안하는 영상의 영역별 분류 기법과 영역별 스테레오 정합 기법은 각각 제 3 장과 제 4 장에 기술한다. 제 5 장에서는 제안하는 알고리즘을 간략히 요약하고 제 6 장에서는 실험 결과를 제시한다.

2. 변위 공간

우선 본 장에서는 스테레오 정합에서 중요한 개념인 변위 공간에 대해 간략히 소개한다. 두 카메라가 평행하게 놓여 있다고 가정하거나 두

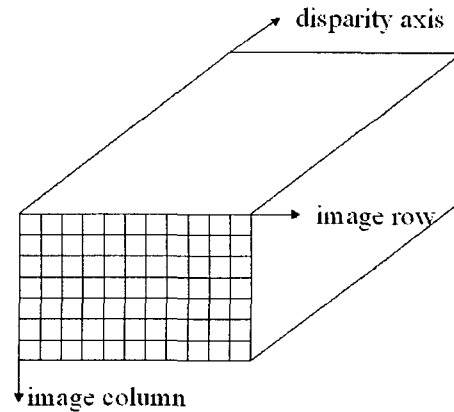


그림 1. 변위 공간

영상이 렉티피케이션 (rectification) 되었다고 가정을 한다면, 스테레오 정합은 기준 영상의 화소 $P(i, j)$ 에 대해 다른 영상의 화소 $P(i+d, j)$ 들을 정합하는 것이다. 이 때 d 를 기준 영상에 대한 변위(disparity)라고 부른다. 그렇다면 변위 공간은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \text{Disparity Space} = \\ \{(i, j, d) \mid 0 \leq i \leq \text{image width}, \\ 0 \leq j \leq \text{image height}, \\ 0 \leq d \leq \text{disparity search size}\} \end{aligned}$$

여기서 d 는 최소의 변위값을 0 으로 보고 각각의 변위값들을 최소의 변위값만큼 감소시킨 값이다.

변위 공간의 관점에서 보면 기준 영상의 화소 $P(i, j)$ 은 d 축 방향으로 각각의 d 에 대해 정합값을 가지고 있으며 이러한 d 축 방향의 정합값 분포를 구해 그 중 최대 혹은 최소의 정합값을 찾아내는 것을 스테레오 정합이라고 한다.

3. 영역 분류와 영역별 정합값 분포특징

제안하는 기법에서는 각 화소의 정합값 분포 특징에 따라 신뢰 영역, 잡음 영역, 에지 영역, 텍스처가 적은 영역, 텍스처가 반복되는 영역, 가리어짐 영역의 여섯 가지로 각 화소들을 분류한다.

신뢰 영역은 초기 정합의 결과만으로도 구한 변위값을 신뢰 할 수 있어 더 이상의 변위값의 갱신이 필요 없는 영역을 뜻한다. 신뢰 영역의 변위 분포는 변위의 참값을 중심으로 하여 참값에서 멀어지는 변위값 일수록 정합값이 줄어드는 참값 중심의 대칭적 형태를 보인다. 하나의 변위를 기준으로 거기에서 멀어질수록 감소하는 정합값을 가진다는 것은 그 하나의 변위를 신뢰

할 수 있음을 의미한다.

잡음 영역은 영상의 잡음에 의해 초기 정합의 결과를 신뢰 할 수 없는 영역을 뜻한다. 변위의 분포는 불규칙적인 특성을 보인다. 즉 정합값 분포의 증가, 감소가 불규칙적으로 자주 일어난다. 이는 잡음에 의해 참값이 아닌 변위에 서도 큰 정합값을 가질 수 있기 때문에 일어나는 현상이다. 이러한 잡음은 초기 정합 윈도우의 크기가 작기 때문에 일어나므로 이러한 영역은 윈도우의 크기를 일정 정도 이상 크게 하여 정합하면 정확한 변위를 얻을 수 있다.

에지 영역은 영상의 밝기 차이가 발생하는 경계 영역을 의미한다. 이 영역은 나머지 영역들과는 달리 정합값들의 분포로부터 구해지지 않고 단순히 영상으로부터 에지를 구해 얻는다. 이러한 에지 영역은 정합값 갱신 과정에서 깊이의 불연속선을 보존하고 가리어짐을 얻는데 이용된다. 구체적인 방법은 제 4 장에서 설명 할 것이다.

텍스처가 적은 영역은 밝기값의 변동이 거의 없는 영역을 의미한다. 밝기값의 변동이 없기 때문에 이런 영역에서는 정합값 분포 역시 변화가 없이 평탄하다. 잡음 영역의 경우는 일정 크기 이상의 고정된 윈도우를 쓰면 문제가 해결이 되지만 텍스처가 적은 영역의 경우 정합에 충분한 밝기값의 변동을 반영 할 수 있는 큰 윈도우를 사용하여 정합을 하여야 한다. 이는 윈도우 크기를 고정 할 수 없음을 의미한다. 제안하는 기법에서는 윈도우의 크기를 같은 물체 표면에 존재 하면서 텍스처가 적은 영역으로 분류된 모든 화소들로 잡았다. 다시 말해 텍스처가 없는 곡면 각각을 영역화(segmentation)하여 그 영역화 된 전체를 정합하는 방법을 쓰고 있다.

텍스처가 반복되는 영역은 밝기값의 변동이 주기적으로 일어나는 영역을 뜻한다. 이러한 영역은 정합값의 분포 역시 주기성을 지닌다. 이러한 주기적 특성을 보이는 영역은 주변 화소들로부터 정합값을 지원 받음으로써 정확한 정합의 결과를 얻을 수 있다 [4]. 지원 받는다는 것은 주기적 분포에서 최대값을 가지는 변위 각각에 대해 주변 화소들의 변위 각각의 근방에 위치한 변위들로부터 정합값을 지원 받아 그 정합값이 가장 큰 것을 참값으로 정한다는 것을 의미한다. 끝으로 가리어짐 영역은 한 쪽 영상에서는 보이지 않지만 다른 쪽 영상에서는 보이지 않는 영역을 의미한다. 이러한 영역에서는 정합값의 최대값이 매우 작다. 따라서 우선 일정 이하의 최대값을 보이는 화소들은 가리어짐 영역으로 분류한다. 그러나 잡음등에 의해 가리어짐 영역이라 해도 정합의 최대값이 클 수가 있다. 가리어짐은 깊이의 불연속선 근방에서 일어나므로 에지 화소 영

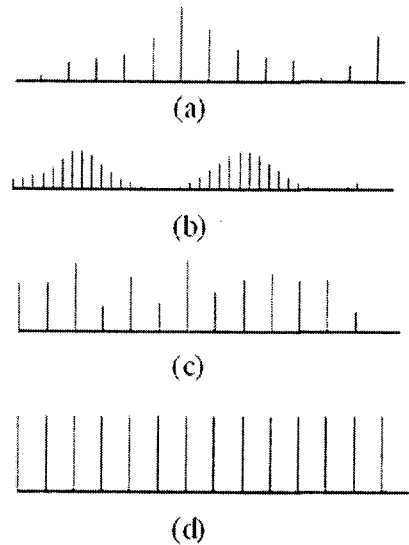


그림 2. 정합값 분포의 특징

- (a) 신뢰 영역 (b) 텍스처가 반복되는 영역
(c) 잡음 영역 (d) 텍스처가 없는 영역

역을 정합하고 여기서 제대로 정합되지 않은 화소들을 대상으로 가리어짐을 한 번 더 판정 한다 구체적인 방법은 제 4 장에서 설명 할 것이다.

4. 영역별 정합 알고리즘

잡음 영역 및 텍스처가 반복되는 영역은 이미 밝혔듯이 전자는 고정 크기의 큰 윈도우를 사용하고 후자는 주변 화소들로부터 지원을 받음으로써 정합이 가능하다. 본 절에서는 어떻게 에지 영역이 깊이의 불연속선을 보존하는데 쓰이며 또한 가리어짐을 판단하는데 쓰이는지에 관하여 자세히 기술한다.

제 3 장에서 기술한 바와 같이 텍스처가 적은 영역으로 분류된 화소의 경우 윈도우의 크기를 같은 분류의 화소들 중 같은 물체 표면에 존재 하는 모든 화소들로 잡는다. 이 때 문제가 되는 것은 텍스처가 없는 영역의 화소들 중 어떤 것들을 서로 같은 물체의 표면상에 있는 것으로 판단해야 하는가 하는 것이다. 보통은 범람 기법(flood-fill method)을 이용하여 자신의 근방에 있는 같은 분류의 모든 화소를 찾아 윈도우로 사용하면 된다. 하지만 이웃하는 서로 다른 깊이값을 가지는 두 물체가 서로 다른 밝기값을 지니서 서로 에지를 경계로 구별은 되지만 두 물체 모두 텍스처가 없는 특성을 가지는 경우가 생길 수 있다. 이런 경우에는 단순히 범람 기법을 사용하면 서로 다른 두 영역이 하나의 영역으로

잡히게 되어 결국은 잘못된 정합 결과를 나타내게 된다. 에지 영역은 이러한 문제를 바로 잡는데 이용된다. 앞의 경우 분명히 두 물체의 경계선에서는 에지가 발생할 것이며, 범람을 시켰을 때 범람이 에지에서 멈출 것이다. 따라서 서로 다른 깊이의 두 꼭면이 서로 다르게 영역화 될 수 있다.

한편 깊이의 불연속선은 정합값의 분포 특성이 뚜렷하지 못하다. 잡음이 없는 경우 경계의 양쪽에 해당하는 두 개의 변위에서 최대의 정합값을 가지는 분포를 보이지만 잡음이 있거나 경계의 양 쪽이 모두 텍스처가 적은 경우는 이러한 특징을 보이지 않는다. 따라서 단순히 정합값의 분포만으로 정합을 할 경우 불필요하게 큰 윈도우가 적용이 되어 깊이의 불연속선이 왜곡되게 된다. 따라서 깊이의 불연속선일 수 있는 가능성이 큰 에지 화소들에 대해서는 정합 과정에서 우선 제외를 하고 다른 영역의 정합을 모두 끝낸 뒤 후처리를 할 필요가 있다.

후처리 과정은 다음과 같다. 우선 에지 화소 근방에 3*3 윈도우를 적용한다. 윈도우 내에서 변위값의 변화가 존재하면 해당하는 에지 화소는 깊이의 불연속선으로 판정한다. 만약 그렇지 않다면 해당하는 에지 화소는 단순히 밝기값의 변동에 의해 생긴 것이므로 주변의 변위값으로 변위값을 준다. 끝으로 깊이의 불연속선으로 판단된 경우 에지가 완전히 수평 방향이 아니면 가리어짐 영역으로 판정한다. 이는 일반적으로 스테레오 영상은 두 대의 카메라가 평행한 상태에서 얻어지며, 이 때 가리어짐은 수평이 아닌 깊이가 불연속선상에서 발생 할 것이기 때문이다.

6. 실험 결과

실험은 츠쿠바 대학의 테스트 영상에 대해 수행되었다. 실험 결과를 관찰해 보면 카메라 부분이나 책상과 석고상과의 경계선, 그리고 책상과 배경과의 경계선이 기존의 적응적 윈도우를 이용한 스테레오 기법에 비해 보다 정확히 구해졌음을 알 수 있다.

7. 결론

본 논문에서는 국부 영역으로 영상을 영역화하고 변위 공간(disparity space)상에서의 특징을 분석하여 각각의 영역에 적합한 윈도우의 크기를 정하는 새로운 스테레오 정합 기법을 제안하였다. 본 논문에서는 이러한 영역들의 변위 공간상에서의 정합 값 분포를 분석하여 스테레오 정합에 이용하였다. 실험 결과 기존의 방법에 비해



그림 3. 실험 결과

제안하는 기법은 짧은 수행 시간 및 정확한 정합 결과를 보여 주었다.

8. 참고 문헌

- [1] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no 9, pp. 920-932, September 1994.
- [2] O. Vekler, "Stereo matching by compact windows via minimum ratio cycle," *Proc. IEEE International Conference on Computer Vision*, pp. 540-547, July 2001.
- [3] D. Marr and T. Poggio, "Cooperative computation of Stereo Disparity," *Science* vol. 194, pp. 209-236, 1976.
- [4] C. Zitnick and T. Kanade, "A volumetric iterative approach to stereo matching and occlusion detection," CMU Technical Report CMU_RI_98-30, 1998.
- [5] C. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *IEEE Trans Pattern Analysis and Machine Intelligence*, vol. 22, no. 7, pp. 675-684, July 2000.
- [6] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Proc. IEEE International Conference on Computer Vision*, pp. 377-384, 1999.
- [7] V. Kolmogorov, R. Zabih, "Computing visual correspondence with occlusion via graph cuts," *Proc. IEEE International Conference on Computer Vision*, pp. 508-515, July 2001.