

스테레오 영상의 영역별 특성에 따른 적응적 스테레오 정합 방법

장용준, 호요성
광주과학기술원 정보통신공학부
e-mail : yjchang@gist.ac.kr, hoyo@gist.ac.kr

Adaptive Stereo Matching Method Depending on Characteristic of Regions in Stereo Images

Yong-Jun Chang, Yo-Sung Ho
School of Information and Communications
Gwangju Institute of Science and Technology (GIST)

요약

스테레오 정합은 스테레오 카메라로 촬영된 서로 다른 두 시점의 색상 영상으로부터 깊이 값을 예측하는 대표적인 방법 중 하나다. 스테레오 정합을 통해 각 시점 영상의 변위 지도를 얻게 되는데, 이 지도는 색상 영상을 3차원 영상으로 만드는 중요한 단서로 사용 된다. 따라서 3차원 영상의 품질은 변위 지도의 정확도에 따라 결정된다. 특히, 영상의 평탄한 영역과 영상 속 객체의 경계 영역에 있는 화소들은 정확한 변위 값을 예측하기 어려워 변위 지도 품질에 많은 영향을 준다. 기존 논문은 영상 속 경계 영역의 변위 값 정확도를 높이기 위해 거리 변환 수식을 이용했다. 이 논문은 경계 영역의 변위 값을 빠르고 정확하게 얻기 위해 영상의 영역별 특성에 따른 적응적 스테레오 정합 방법을 제안한다.

1. 서론

깊이 값은 물체가 카메라로부터 얼마나 멀리 떨어져 있는지 나타내는 중요한 정보다. 많은 3차원 영상 콘텐츠들이 이러한 깊이 정보를 기반으로 만들어진다. 따라서 객체의 정확한 깊이 정보를 구하는 것은 3차원 영상 콘텐츠 생성에 있어서 아주 중요한 과정 중 하나다.

객체의 깊이 정보를 얻기 위한 방법 중 하나로 깊이 카메라를 이용해 깊이를 측정하는 방법이 있다. 깊이 카메라는 촬영하는 영상의 깊이를 빠르고 쉽게 얻을 수 있다는 장점이 있지만 깊이 측정이 가능한 거리가 제한적이며 햇빛의 영향을 받는 실외에서는 정확한 깊이 값을 측정하기가 힘들다는 단점이 있다.

반면에 색상 카메라로 촬영된 영상만 이용해 영상 속 객체의 깊이 정보를 예측하는 방법도 있다. 대표적인 예로 스테레오 정합이 있다. 스테레오 정합은 서로 다른 두 시점을 갖는 스테레오 영상을 이용해 깊이 값을 예측한다. 이 방법은 각 시점의 영상에서 서로 대응하는 점을 찾은 후 두 점 사이의 변위차를 계산해 깊이 정보를 얻는다. 이때 두 대응점은 화소의 유사도 비교에 의해 결정 되는데 비교 방법에 따라 연산의 복잡도가 크게 증가하기도 한다.

스테레오 정합을 통해 얻은 영상을 변위 지도라고 한다. 변위 지도는 객체의 깊이 정보를 담고 있으며 이를 이용해 3차원 영상 콘텐츠를 생성하기도 한다. 따라서 정확한 변위 값을 얻는 것이 고품질의 3차원 영상 콘텐츠를 만드는 데 있어서 아주 중요하다.

이 논문에서는 거리 변환을 이용해 변위 지도의 경계

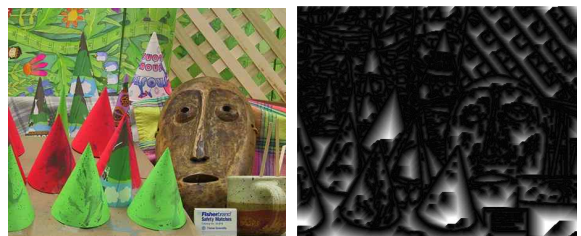
영역 변위 값 정확도를 높이는 방법을 제안한다. 제안한 방법은 색상 영상을 거리 변환을 이용해 바뀐 후 영역의 특성에 따라 서로 다른 정합 방법을 적용해 빠르게 변위 값을 탐색하도록 한다.

2. 거리 변환을 이용한 경계 영역 보존 방법

2.1 거리 변환

거리 변환은 영상 속에 있는 화소들이 객체의 경계로부터 얼마큼 멀리 떨어져 있는지 계산한다. 이 변환을 이용해 스테레오 정합에서 경계 영역의 변위 값 탐색 정확도를 높이는 연구도 발표되었다 [1]. 식 (1)은 거리 변환을 위해 사용되는 수식이다.

$$r_{i,j}^k = \min \begin{bmatrix} r_{i-1,j-1}^{k-1} + \beta & r_{i,j-1}^{k-1} + \alpha & r_{i+1,j-1}^{k-1} + \beta \\ r_{i-1,j}^{k-1} + \alpha & r_{i,j}^{k-1} & r_{i+1,j}^{k-1} + \alpha \\ r_{i-1,j+1}^{k-1} + \beta & r_{i,j+1}^{k-1} + \alpha & r_{i+1,j+1}^{k-1} + \beta \end{bmatrix} \quad (1)$$



(a) 원본 영상

(b) 거리 변환 영상

(그림 1) 거리 변환 결과

식 (1)에서 $r_{i,j}^k$ 는 (i,j) 에 있는 화소의 k 번째 거리 변환 값을 뜻한다. α 와 β 는 모두 거리 변환의 강도를 조절하는 역할을 한다. 거리 변환은 경계 영상에 적용되며 거리 변환 결과는 그림 1과 같다.

2.2 거리 변환을 이용한 스테레오 정합

거리 변환 영상은 경계 영역에 있는 화소뿐만 아니라 경계 영역 주변의 화소들까지 특정한 거리 값으로 나타낼 수 있다. 따라서 변환 영상의 이러한 특성을 이용해 경계 영역에 있는 화소들이 변위 값 정확도를 높일 수 있다 [2]. 식 (2)는 거리 변환을 이용한 정합 수식을 나타낸다.

$$D_s(d_s) = \frac{\sum_{t \in N(s)} W_{s,t}(dt_t) |I_L(x_t, y_t) - I_R(x_t + d_s, y_t)|}{\sum_{t \in N(s)} W_{s,t}(dt_t)} \quad (2)$$

식 (2)에서 $D_s(d_s)$ 는 변위 값 후보 d_s 에 따른 데이터 항의 에너지 값이다. 데이터 항은 좌시점 영상과 우시점 영상 화소들의 유사도 비교를 통해 계산된다. s 는 현재 화소의 위치를 나타내며 t 는 정합창 $N(s)$ 에 있는 s 의 이웃 화소들을 뜻한다. $W_{s,t}(dt_t)$ 는 두 화소의 유사도 비교에 사용되는 가중치 함수며 식 (3)과 식 (4)와 같이 정의된다.

$$W_{s,t}(dt_t) = f(dt_t) \cdot g(|I_{L,s} - I_{L,t}|) \quad (3)$$

$$f(dt_t) = 1 - e^{-\frac{dt_t^2}{2\sigma_f}}, g(|I_{L,s} - I_{L,t}|) = e^{-\frac{|I_{L,s} - I_{L,t}|^2}{2\sigma_g}} \quad (4)$$

식 (3)에서 dt_t 는 t 위치에 있는 화소의 거리 변환 값을 뜻한다. 거리 변환 가중치 함수는 식 (4)와 같이 거리 변환 결과 값을 이용한 가중치 함수인 $f(dt_t)$ 와 정합창 안에 있는 화소들의 색상 가중치인 $g(|I_{L,s} - I_{L,t}|)$ 로 구성된다.

따라서 식 (2)는 화소의 유사도 비교 과정에서 경계 영역 주변에 있는 화소에 더 많은 가중치를 두어 경계 영역 화소들의 변위 값 탐색 정확도를 높여준다. 식 (2)는 Markov 랜덤 필드를 이용해 최종 에너지 함수인 식 (5)로 정의할 수 있다. 식 (5)에서 $S_{s,t}(d_s, d_t)$ 는 평활화 항으로 현재 화소의 변위 값과 주변 화소의 변위 값을 비교하는 역할을 한다. 식 (5)는 여러 가지 최적화 방법을 통해 최종 변위 값을 결정한다.

$$E(d) = \sum_s D_s(d_s) + \sum_{s,t \in N(s)} S_{s,t}(d_s, d_t) \quad (5)$$

3. 영역별 특성에 따른 적응적 스테레오 정합

3.1 거리 변환을 이용한 스테레오 정합의 문제점

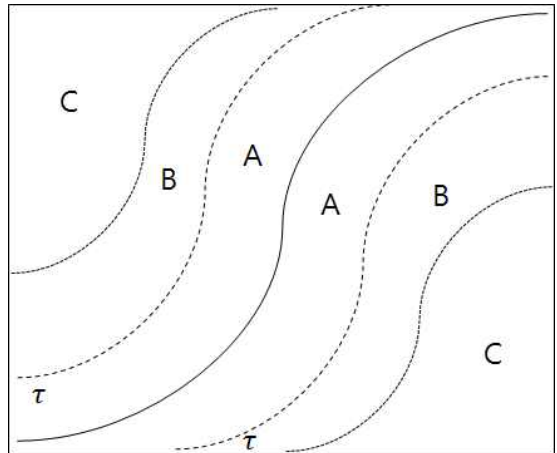
거리 변환 영상은 화소의 거리 값을 8비트 값으로 나타낸다. 따라서 화소 값은 0부터 255까지 총 256개의 값 안

에서 결정된다. 그림 1(b)에서 하얀색 화소들은 경계로부터의 거리 변환 값이 255 이상이 되는 것을 뜻한다. 거리 변환을 이용한 스테레오 정합은 이러한 영역들에 대해서도 식 (2)를 적용한다 [1]. 식 (2)의 정합창 안에 있는 거리 변환 값이 모두 255일 경우 가중치 함수 $f(dt_t)$ 는 모두 동일한 가중치 값을 갖게 돼 의미가 없어지게 된다. 또한 불필요한 연산이 증가하게 된다.

본 논문에서는 기존의 거리 변환을 이용한 스테레오 정합의 장점을 살리면서 연산 복잡도를 줄여 수행속도를 빠르게 하는 방법을 제안한다.

3.2 거리 변환을 이용한 영역의 구분

기존의 거리 변환을 이용한 방법에서 연산 복잡도를 낮추기 위해 제안한 방법은 영상의 영역별 특성에 따라 각각 다른 정합 방법을 적용한다. 우선 영상의 영역을 구분하기 위해 기존의 거리 변환 영상을 이용한다. 그림 2는 영상의 영역을 구분하는 방법을 나타낸다.



(그림 2) 거리 변환 영상을 이용한 영상의 영역 구분

그림 2는 거리 변환이 적용된 영상이다. 그림 2에서 실선은 영상의 경계 영역을 나타내며 점선과 파선은 사용자에게 의해 영역이 구분된 부분을 나타낸다. 거리 변환 값이 특정 임계값 τ 보다 작을 경우 그 값을 지닌 화소가 있는 영역은 A 영역으로 간주한다. 반대로 변환 값이 τ 보다 크고 255보다 작을 경우에는 그 화소들이 모인 곳을 B 영역이라고 정의한다. 마지막으로 변환 값이 255인 화소들이 있는 영역을 C 영역이라고 정의한다.

즉, A 영역은 경계를 포함한 경계와 가까운 주변 화소들이 모여 있는 영역을 뜻하며, B 영역은 경계로부터 멀리 떨어져 변환 값이 τ 와 255 사이인 화소들이 있는 영역을 뜻한다. 그리고 C 영역은 모든 화소의 변환 값이 255인 평탄한 영역을 뜻한다.

3.3 적응적 스테레오 정합 방법

3.2절에서는 거리 변환 영상을 이용해 영상을 세 가지 영역으로 구분했다. 그림 2에서 A 영역은 경계를 포함하

며 경계와 아주 가까이 있는 화소들을 포함하고 있다. 따라서 경계 영역의 변위 값 정확도를 높이기 위해서는 이 영역에 있는 화소들의 대응점을 정확히 찾는 것이 중요하다. 식 (2)를 사용하는 방법은 변위 지도에서 경계 영역을 잘 보존하는 역할을 하지만 우리는 연산의 복잡도를 낮추기 위해 색상 영상과 변화율 영상을 이용한 화소 기반의 스테레오 정합 수식을 적용했다 [2]. 변화율 영상의 화소 값은 객체의 경계 영역에서 굉장히 큰 화소 값 변화율을 보인다. 따라서 이 정보를 이용해 경계 영역의 정합 정확도를 높일 수 있다. 이 수식은 식 (6)과 같이 정의된다.

$$D_s(d_s) = k \cdot |I_L(x_s, y_s) - I_R(x_s + d_s, y_s)|, \text{ if } s < \tau \quad (6)$$

$$+ (1-k) \cdot |\nabla I_L(x_s, y_s) - \nabla I_R(x_s + d_s, y_s)|$$

식 (6)에서 ∇I 는 변화율 영상의 화소 값을 나타낸다. k 는 색상 영상의 정보와 변화율 영상의 정보를 합하기 위해 사용되는 가중치 값이다. 따라서 현재 화소 s 가 임계 값 τ 보다 작을 경우에 식 (6)을 적용해 데이터 항의 에너지를 계산한다.

B 영역에 있는 화소들은 거리 변환 값을 갖고 있지만 경계 영역으로부터 멀리 떨어져 있는 화소들이다. 이러한 경우 경계 영역과 같이 화소 값이 급격히 변화하는 부분이 없기 때문에 식 (6)과 같이 색상 영상과 변화율 영상만을 이용해 정확한 대응점을 찾는 것은 쉽지 않다. 따라서 현재 화소 s 가 임계값 τ 보다 크거나 같고 255보다 작을 경우 식 (2)를 이용해 에너지 값을 얻는다.

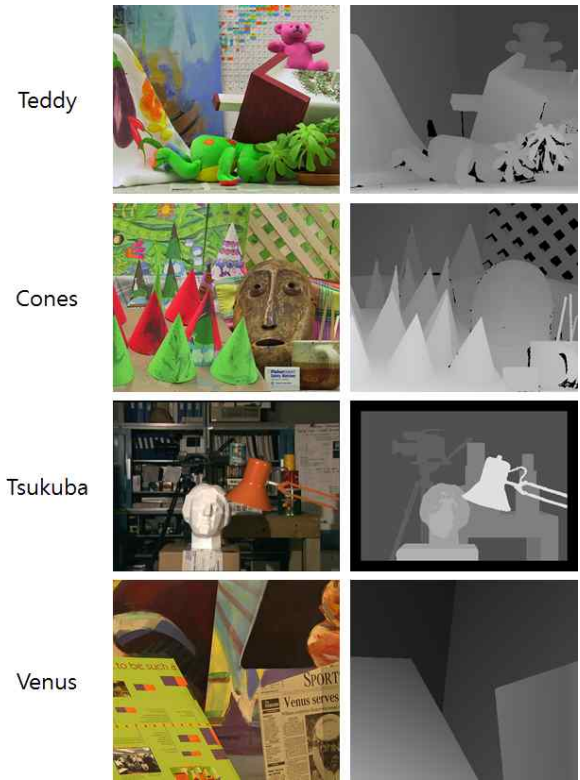
마지막으로 C 영역은 경계 영역으로부터 멀리 떨어져 있는 평탄한 영역이기 때문에 255이외의 특정한 거리 변환 값을 갖지 않는다. 또한 변화율 영상 및 색상 영상에서도 균일한 화소 값을 갖는다. 따라서 이 영역에서는 식 (2)를 이용한 정합 수식은 큰 영향을 주지 못하기 때문에, A 영역과 같은 식 (6)의 화소 기반의 정합 방법을 적용한다.

4. 실험 결과

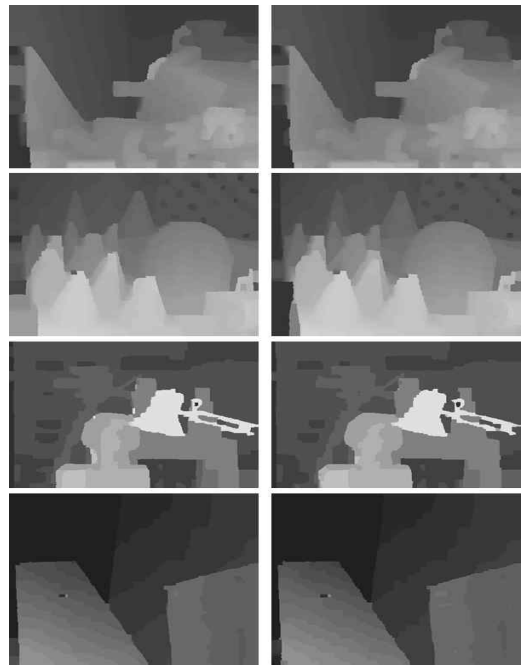
본 논문에서 거리 변환 영상을 구하기 위한 전처리 과정으로 *Canny* 경계 검출 방법을 적용해 색상 영상으로부터 경계 영상을 구했다 [3]. 또한 식 (2)와 식 (6)의 데이터 항이 사용되는 에너지 함수를 최적화 하기 위해 계층적 신뢰도 전파 방법을 사용했다 [4].

기존의 방법과 제안한 방법을 비교하기 위해 *Teddy*, *Cones*, *Tsukuba*, *Venus*, 총 네 개의 실험 영상을 사용했다. *Teddy*와 *Cones* 영상은 450×375 크기의 해상도를 갖고 *Tsukuba*와 *Venus*는 각각 384×288, 434×383 크기의 해상도를 갖는다. 각각의 스테레오 정합 방법을 통해 얻은 변위 지도를 객관적으로 비교하기 위해 원본 변위 지도 대비 실험 결과 영상의 오차율을 계산했다.

그림 3은 실험에 사용된 원본 영상과 각 원본 영상의 실제 변위 지도를 나타낸다.



(a) 원본 영상 (b) 실제 변위 지도
(그림 3) 원본 영상 및 실제 변위 지도



(a) 기존 방법 (b) 제안한 방법
(그림 4) 기존 방법 및 제안한 방법의 결과 영상

그림 4는 거리 변환을 이용한 스테레오 정합 방법과 제안한 방법의 실험 결과 영상을 보여준다. 그림 4(b)에서 *Teddy*와 *Tsukuba* 영상의 경우 일부 영역에서 기존 방법보다 경계 영역이 무너진 모습을 확인할 수 있다. 하지만

대부분의 실험 결과 영상이 거의 비슷한 변위 지도를 보여주고 있어 육안으로는 정확한 비교가 힘들다. 따라서 각 실험 영상의 오차율을 계산해 표 1에 정리하였다.

(표 1) 제안한 방법의 오차율 비교

Algorithm		기존 방법 [1]	제안한 방법
Teddy	nonocc.	10.68	10.37
	all	20.61	20.37
	disc.	27.28	27.16
Cones	nonocc.	6.69	6.22
	all	17.52	18.24
	disc.	17.71	17.77
Tsukuba	nonocc.	1.87	1.83
	all	3.98	3.72
	disc.	10.26	10.23
Venus	nonocc.	1.23	1.27
	all	4.16	4.20
	disc.	15.59	14.67
Average		11.47	11.34

실험 결과의 오차율은 실제 변위 지도와 결과 영상에서 같은 화소 위치의 변위 값을 비교했을 때 변위 차이가 1보다 클 경우 그 화소를 오류 화소로 판단하여 구한 백분율이다. 오차율은 비폐색 영역, 전 영역, 그리고 경계 영역에 있는 화소들을 비교해 구했다. 표 1을 보면 제안한 방법이 기존 방법보다 평균 오차율이 0.13% 더 낮은 것을 확인할 수 있다.

(표 1) 제안한 방법의 수행시간 비교

Algorithm	기존 방법 [1]	제안한 방법
Time (sec.)		
Teddy	39.03	13.29
Cones	40.79	9.00
Tsukuba	6.55	1.87
Venus	12.37	4.62
Average	24.69	7.20

기존 방법과 제안한 방법의 스테레오 정합 속도를 비교하기 위해 각 실험 영상의 변위 지도 출력 시간을 확인해 표 2에 정리하였다. 표 2의 수행시간 비교 결과 제안한 방법이 기존 방법보다 약 70%의 속도 증가를 보이는 것을 확인할 수 있다.

5. 결 론

본 논문은 영상의 영역별 특성에 따라 다른 스테레오 정합 방법을 제안했다. 변위 지도의 경계 영역 변위 값 정확도를 높이는 기존 방법으로 거리 변환을 이용한 방법이 있으나 이 방법은 복잡도가 높다는 단점이 있다. 이 문제를 극복하기 위해 제안한 방법은 거리 변환 영상을 이용해 영상의 영역을 구분하고 각 영역의 특성에 따라 다른 정합 방법을 적용했다. 그 결과, 제안한 방법의 평균 오차율은 기존 방법 대비 0.13% 감소했으며 평균 수행시간은 24.69초에서 7.20초로 약 70%의 속도 감소율을 보였다.

감사의 글

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

참 고 문 헌

[1] W.S. Jang, Y.S. Ho, “Discontinuity preserving disparity estimation with occlusion handling,” Journal of Visual Communication and Image Representation, Vol. 25 Issue 7, pp. 1595-1603, Oct. 2014.

[2] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, M. Gelautz, “Fast cost-volume filtering for visual correspondence and beyond,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 3017-3028, 2011.

[3] J. Canny, “A computational approach to edge detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, Issue 6, pp. 679-698, 1986.

[4] Q. Yang, L. Wang, R. Yang, S. Wang, M. Liao, D. Nister, “Real-time global stereo matching using hierarchical belief propagation,” British Machine Vision Conference, pp. 989-998, 2006.