

움직임 예측 및 주변화소의 변위정보를 이용한 고속 스테레오 정합 방법

장용준 호요성

광주과학기술원 전기전자컴퓨터공학부

 {yjchang, hoyo}@gist.ac.krFast Stereo Matching Method Using Motion Estimation and
Disparity Information of Neighboring Pixels

Chang, Yong-Jun Ho, Yo-Sung

Gwangju Institute of Science and Technology (GIST)

요약

스테레오 정합은 촬영된 스테레오 영상 속 화소들의 대응점을 탐색한 후 대응점 사이의 변위차를 계산하여 깊이정보를 예측한다. 스테레오 정합에서 변위값을 계산하기 위해서는 스테레오 영상간의 대응점 탐색이 우선적으로 수행되어야 한다. 스테레오 영상의 변위값 범위를 모르는 경우 동일한 탐색선상에 있는 모든 화소들의 유사도를 비교한 후 최적의 대응점을 선택한다. 반면에, 변위값 범위가 제공되는 실험 영상으로 스테레오 정합을 할 경우 정해진 후보 화소들에 대해서만 대응점 탐색을 수행한다. 많은 스테레오 정합 논문들이 실험의 효율성을 위해 변위값 범위 정보가 제공되는 실험 영상으로 스테레오 정합을 수행한다. 하지만 실제 스테레오 정합 환경에서는 이와 같은 정보를 얻기가 힘들다. 이러한 문제점을 극복하기 위해 본 논문에서는 움직임 예측 및 주변화소의 변위정보를 이용한 고속 스테레오 정합 방법을 제안한다.

1. 서론

3차원 영상 콘텐츠는 지난 몇 년 동안 문화, 의료, 산업 등 다양한 분야에서 전문가 및 많은 대중들에게 개발되고 사용되어왔다. 최근에는 현실 세계를 기반으로 다양한 영상 콘텐츠를 결합하는 증강현실 (Augmented reality, AR)과 가상의 공간에서 다양한 영상 콘텐츠를 제공하는 가상현실 (Virtual reality, VR)이 대중들로부터 많은 관심을 받고 있다. 3차원 영상 및 AR/VR에 사용되는 많은 실감 영상 콘텐츠들은 객체의 깊이정보를 활용한다. 깊이정보 획득에는 다양한 방법이 사용되는데, 그 중 스테레오 정합을 이용한 깊이정보 획득은 깊이 카메라를 이용하는 방법과 비교하여 환경적 요인에 강인하다는 장점이 있다. 하지만 이 방법은 깊이정보를 얻기 위한 대응점 탐색 과정에서 많은 시간이 걸린다는 단점이 있다.

스테레오 정합 방법은 변위값을 예측하기 위해 정해진 변위값 후보들에 대해서만 대응점 탐색을 수행한다. 이러한 방법은 변위값 후보의 개수가 적을 경우 스테레오 정합을 빠르게 수행할 수 있다. 하지만 촬영에 사용된 스테레오 카메라의 간격이 클 경우에는 탐색에 사용되는 변위값 후보들이 많아지기 때문에 스테레오 정합에 걸리는 시간이 길어진다. 많은 스테레오 정합 논문들은 스테레오 정합 알고리즘을 평가하는 과정에서 실험의 효율성을 위해 실험 영상의 최대 변위값 정보를 활용해 스테레오 정합을 수행하는 방법을 사용한다. 하지만 실제 스테레오 영상 촬영 환경에서는 이러한 정보를 얻기가 쉽지 않다. 따라서 본 논문에서는 주변 화소의 변위정보를 활용해 초기 탐색 단계를 결정하고, 이 값을 움직임 예측 방법에 적용하여 변위값을 효율적으로 탐색하는 방법을 제안한다 [1].

2. 스테레오 정합의 깊이정보 탐색 방법

스테레오 정합은 스테레오 영상과 인간의 양안시차 특성을 이용해 깊이정보를 예측하는 방법이다. 따라서 스테레오 정합은 각 영상 속 화소들의 대응점을 탐색하고, 검출된 대응점들 사이의 변위 차이로 해당 화소의 깊이정보를 예측한다. 스테레오 정합에서 전역 탐색 방법은 각 화소의 대응점을 찾기 위해 탐색선상에 있는 모든 화소에 대해 대응점 탐색을 수행하는 방법을 뜻하며, 이러한 탐색 방법으로 인해 지역 최소화 문제에 강인하다는 장점이 있다. 하지만 색상 변화가 일정한 평탄한 영역의 경우 정확한 깊이정보를 예측하지 못한다는 단점이 있다.

많은 논문에서 스테레오 정합 알고리즘을 평가하기 위해 Middlebury stereo 실험 영상을 사용한다 [2, 3]. 이 실험 영상은 스테레오 정합 알고리즘의 효율적인 평가를 위해 사용자에게 최대 변위값 정보를 제공한다. 최대 변위값을 이용한 대응점 탐색은 잘못된 대응점을 탐색할 확률을 낮추기 때문에, 스테레오 정합 결과에서 변위값 오류를 줄여준다. 또한, 이 방법은 탐색선상의 모든 화소를 탐색하는 것이 아닌 정해진 후보들에 대해서만 탐색하기 때문에 전역 탐색 방법보다 속도가 빠르다는 장점이 있다. 하지만 실제 스테레오 정합 응용 분야에서는 획득한 영상의 최대 변위값 정보를 알 수 없다. 따라서 이러한 탐색 방법을 일반적인 대응점 탐색 방법으로 사용하는데 한계가 있다.

3. 고속 스테레오 정합 방법

제안한 방법의 대응점 탐색은 비디오 코덱에서 사용되는 움직임 예측 방법과 유사하다 [1]. 그림 1은 좌시점 영상에 있는 한 화소의 대

응점을 탐색하기 위해 우시점 영상의 탐색선상에서 일어나는 대응점 탐색 과정을 보여준다.

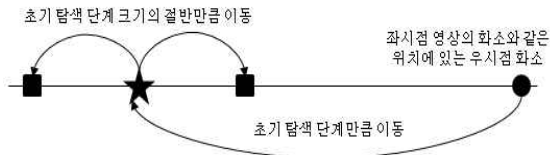


그림 1. 고속 스테레오 정합을 위한 대응점 탐색 방법

우선 변위값을 예측하려는 좌시점 영상의 화소와 동일한 위치에 있는 우시점 화소를 찾는다. 대응점 탐색을 위해 먼저 초기 탐색 단계 크기를 결정하고 그 크기만큼 우시점 화소로부터 좌측으로 이동한다. 이때, 이동한 위치의 화소와 좌시점 화소 사이의 유사도를 측정한다. 그 후, 초기 탐색 단계 크기의 절반을 새로운 탐색 단계 크기로 결정하고 최초 이동한 화소로부터 새로 결정된 탐색 단계 크기만큼 좌, 우측으로 이동하여 좌시점 화소와의 유사도를 측정한다. 측정된 유사도값 중에서 최적의 유사도를 갖는 화소로 이동하여 앞의 과정을 반복한다. 이러한 과정은 탐색 단계의 크기가 1보다 작아질 때 까지 지속된다.

초기 탐색 단계의 크기는 대응점 탐색 정확도에 큰 영향을 미치기 때문에 변위값이 이미 계산된 주변화소를 이용해 그 크기를 결정한다. 식 (1)은 제안한 방법의 초기 탐색 단계크기를 결정하는 수식이다.

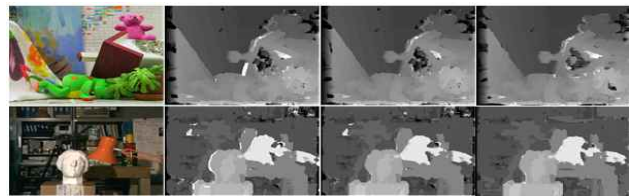
$$S_i = e^{-\frac{var(x,y)}{\alpha}} \cdot d_{(x-1,y)} + \left(1 - e^{-\frac{var(x,y)}{\alpha}}\right) \cdot d_{(x,y-1)} \quad (1)$$

식 (1)에서 S_i 는 초기 탐색 단계의 크기를 나타낸다. S_i 는 좌측에 위치한 화소의 변위값 $d_{(x-1,y)}$ 와 상단에 위치한 화소 $d_{(x,y-1)}$ 를 사용해 결정된다. 만약 현재 화소가 위치한 영역이 평탄한 영역이라면 현재 화소를 포함한 영역의 색상 분산값 $var(x,y)$ 이 작아진다. 따라서 이러한 경우 이전 화소의 변위정보에 큰 가중치를 준다. 반면에 $var(x,y)$ 이 큰 영역일수록 텍스처가 복잡한 영역일 확률이 높기 때문에 이러한 경우 이전 화소의 변위정보뿐만 아니라 상단에 위치한 화소의 변위정보도 사용한다. 만약 이전 화소의 변위값이 0에 가까운 매우 작은값일 경우, 그 화소의 변위값은 오류일 확률이 높다고 판단하여 특정 비율만큼 곱하여 초기 단계 탐색 크기로 활용한다.

4. 실험 결과

제안한 방법의 평가를 위해 Middlebury stereo에서 제공한 실험 영상을 사용했다 [3]. 그림 2는 실험에 사용된 영상과 서로 다른 탐색 방법으로 얻은 변위지도를 보여준다. 그림 2를 보면 제안한 방법의 실험 결과가 텍스처가 없는 평활한 영역에서 다른 방법보다 일관성있는 변위값을 갖는 것을 확인할 수 있다.

표 1과 2는 각각 대응점 탐색 방법에 따른 오차율 및 수행시간을 나타낸다. 표 1을 통해서 제안한 방법이 전역 탐색을 이용한 방법과 최대 변위값 정보를 이용한 방법보다 각각 1.24%, 0.89% 더 낮게 나오는 것을 확인할 수 있으며, 표 2에서는 스테레오 정합 속도가 기존의 방법보다 더 빠르다는 것을 확인할 수 있다 [4].



(a) 원본 영상 (b) 전역 탐색 (c) 최대 변위값 (d) 제안한 방법
그림 2. 스테레오 정합 결과 영상

	전역 탐색	최대 변위값	제안한 방법
	오차율 (%)		
<i>Teddy</i>	29.44	29.15	28.89
<i>Tsukuba</i>	10.78	10.37	8.84
평균	20.11	19.76	18.87

표 1. 대응점 탐색 방법에 따른 오차율

	전역 탐색	최대 변위값	제안한 방법
	수행 시간 (sec.)		
<i>Teddy</i>	66.26	18.63	3.49
<i>Tsukuba</i>	37.43	2.98	1.56
평균	51.85	10.8	2.52

표 2. 대응점 탐색 방법에 따른 수행시간

5. 결론

실제 스테레오 정합 환경에서는 스테레오 영상의 최대 변위값 정보를 획득하는 것이 쉽지 않기 때문에 변위값을 얻기 위해서 대응점 탐색선상의 모든 화소들과 유사도를 비교해야한다. 이러한 문제를 극복하기 위해서 본 논문은 현재 화소의 변위값이 주변 화소와 유사하다는 특성을 이용해 현재 화소가 위치하는 영역의 분산값에 따른 초기 탐색 단계 크기 결정 방법을 제안했다. 또한, 결정된 초기 탐색 단계의 크기를 움직임 예측 방법에 적용하여 최종 변위값을 효율적으로 획득하는 방법을 제안했다. 그 결과, 제안한 방법이 기존의 방법보다 빠르고 정확하게 변위지도를 획득한 것을 확인할 수 있었다.

감사의 글

이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2011-0030079).

참고 문헌

- [1] T. Koga, K. Inuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion compensated inter frame coding for videoconferencing," Proceedings of the National Telecommunications Conference, pp. 1-5, 1981.
- [2] 장용준, 호요성, "SAD 정합 알고리즘 수정을 통한 지역기반 스테레오 정합의 복잡도 감소기법," 한국방송공학회 추계 학술대회, pp. 33-36, 2014.
- [3] <http://vision.middlebury.edu/stereo/>
- [4] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," IEEE Workshop on Stereo and Multi-Baseline Vision, pp. 131-140, 2001.