

확산 거리 기반의 스테레오 비디오의 변이 탐색 범위 추정 방법

이예홍 함범섭 김봉조 강민성 손광훈

연세대학교 전기전자공학과

khsohn@yonsei.ac.kr

Diffusion Distance Based Disparity Search Range Estimation for Stereo Video

Li, Ruei-Hung Ham, Bumsub Kim, Bongjo Kang, Minsung Sohn, Kwanghoon

School of Electrical and Electronic Engineering, Yonsei University

요약

본 논문에서는 깊이 변화에 강인한 스테레오 비디오의 변이 탐색 범위 추정 방법을 제안한다. 정확한 스테레오 비디오의 변이 탐색 범위는 3D 영상 분야에서 다양한 응용으로 사용되는 변이 지도를 보다 정확하게 추정하는데 도움이 된다. 기존의 변이 탐색 범위 추정 방법은 인접한 프레임 간의 변이 히스토그램의 유사도를 이용함으로써 보다 안정적인 변이 탐색 범위를 추정할 수 있었지만, 시간의 흐름에 따라 깊이가 변하는 부분에서는 상당히 취약한 문제점을 가지고 있다. 이에 본 논문에서는 기존 방법의 이러한 문제점을 개선한 새로운 방법을 제안한다. 제안하는 방법은 변이 히스토그램의 유사도뿐만 아니라 프레임 간의 시간적 유사도를 고려하며, 비디오의 장면 전환에 의한 급격한 깊이 변화 또한 고려한다. 이에 추가적으로 변이 히스토그램의 유사도를 계산하기 위해 기존의 방법과는 달리 히스토그램 확산 거리를 이용하였으며, 서로 다른 개수의 대응점을 가지고 있는 프레임간의 변이 히스토그램이 대응점의 개수에 영향을 받지 않고 균일한 중요도를 갖도록 하였다. 실험 결과로 기존 방법과 제안한 방법의 변이 탐색 범위 추정 결과를 비교하였으며, 비교한 결과는 제안한 방법이 기존 방법에 비해 스테레오 비디오의 깊이 변화에 강인함을 보여준다.

1. 서론

최근 3D 영화와 TV 등의 보급으로, 3D 영상 분야에 대한 관심과 연구가 지속적으로 증가하고 있다. 이러한 시대적 흐름에 따라 다양한 3D 콘텐츠의 획득과 이러한 3D 콘텐츠의 질을 높여려는 시도 역시 활발히 진행되고 있다. 이러한 흐름 가운데, 스테레오 영상의 변이 지도(Disparity Map)는 2D-3D 변환(2D-3D Conversion), 중간 영상 합성(Intermediate View Rendering)과 같은 여러 가지 응용에서 사용된다. 따라서 스테레오 카메라로 획득된 좌, 우 영상에서 보다 정확한 변이 지도를 추정하는 연구가 활발하게 진행되어 오고 있다[1]. 이러한 변이 지도를 추정하는 수많은 알고리즘은 대부분 적절한 변이 탐색 범위(Search Range)를 알고 있으며, 모든 변이는 이 탐색 범위 안에 있다고 가정한다. 하지만 사용자에 의해 획득된 임의의 스테레오 영상은 이러한 탐색 범위가 주어지지 않는다.

가능한 탐색 범위를 지정해 주는 가장 간단한 방법으로는, 스테레오 영상이 에피폴라 구속조건(Epipolar constraint)을 충족한다고 가정했을 때, 가능한 변이 탐색 범위를 [-영상의 너비, 영상의 너비]로 지정해 주는 방법이 있다. 하지만 대부분의 변이 지도 추정 알고리즘에서 이렇게 거대한 변이 탐색 범위를 지정하는 것은 시간적, 계산적인 낭비를 유발하게 되며, 또한 오차가 나타날 확률을 크게 한다. 이와는 반대로 변이 탐색 범위를 지나치게 좁게 지정한다면 실제 변이 값이 지정된 탐색 범위 밖에 있을 수 있게 되어 또 다른 오차를 유발할 수 있게 된다. 따라서 임의로 획득된 스테레오 영상에서 적절한 탐색 범위를 추

정하는 것은 매우 중요하다.

스테레오 비디오에서 한 프레임에서의 변이 값은 연속된 다른 프레임에서의 변이 값과 가까운 값을 갖는다. 따라서 이러한 관계를 이용하여 스테레오 비디오의 탐색 범위를 추정하는 방법이 이전에 제안되었다[2]. 이 방법에서는 대응점 매칭 알고리즘으로 생성된 변이 값 히스토그램(Disparity Histogram)을 이용하여 각 프레임의 탐색 범위를 추정한다. 그 후, 보다 정확한 탐색 범위를 추정하기 위해, 이웃한 프레임의 변이 값 히스토그램의 유사도를 측정하여 가중치를 주어 합성한 후 적절히 설정된 임계 값을 기준으로 탐색 범위를 결정한다. 하지만 이 방법에서는 장면 전환이 고려되지 않았으며, 가중치를 계산할 때 이웃한 프레임간의 시간적 유사도를 고려하지 않은 점이 성능 면에서 악영향을 미친다. 또한 각 프레임에서 큰 편차로 추출되는 대응점의 수를 고려하지 않아 불균형적으로 계산된 변이 값 히스토그램이 생성되게 된다.

본 논문에서는 이러한 문제점을 해결하기 위해 개선된 방식의 스테레오 비디오의 변이 탐색 범위 추정 기법을 제안한다. 제안하는 방법에서는 위에서 언급된 기존의 방법에 장면 전환과, 프레임 간의 시간적 유사도를 고려하며, 서로 다른 대응점의 수에 영향을 받지 않도록 적절한 가중치가 적용된 변이 값 히스토그램 생성 방법과 그에 따라 다르게 적용해야 하는 임계 값을 결정한다.

2장에서는 위에서 언급한 기존의 탐색 범위 추정 방법을 자세히 소개하며, 해당 방법의 문제점을 분석한다. 3장에서는 2장에서 분석된 문제점을 해결하기 위한 방법을 제안하며, 4장에서는 제안된 방법의

실험 결과를 보인다.

2. 기존의 탐색 범위 추정 기법

기존의 탐색 범위 추정 기법은 대응점 매칭 알고리즘을 이용하여 매칭되는 대응점 쌍들의 변이 차를 이용해 변이 값 히스토그램을 만든다. 일반적으로 KLT(Kanade-Lucas-Tomasi) [3,4]나 SURF (Speeded Up Robust Features)[5]와 같은 알고리즘을 사용할 수 있으며, 본 논문에서는 SURF 알고리즘을 사용하였다. SURF 알고리즘으로 매칭된 대응점들을 이용해 변이 값 히스토그램을 구하는 식은 다음과 같다.

$$h[i] = \sum_{j=1}^N f(D(i), \left[\frac{d_j}{B} \pm 0.5 \right] \cdot B), \quad i = 1, 2, \dots, M \quad (1)$$

위의 식에서 $h[i]$ 는 i 번째 Bin(Bin)의 히스토그램 값을 나타내며 B 는 Bin의 크기, M 은 Bin의 총 개수를 나타낸다. $f(a,b)$ 는 $a=b$ 일 때 1이며 아니면 0인 함수이다. $D(i)$ 는 양자화 된 변이 값을 나타내며, 따라서 측정된 각각의 변이 값 d_j 는 가장 가까운 $D(i)$ 에 대해 그 히스토그램 값을 증가시킨다. 이렇게 구한 히스토그램을 이용하여 인접한 프레임 간의 깊이 유사도를 고려하기 위하여 가중치 w 를 다음과 같이 구한다.

$$w_{n,p} = \exp\left(-\sum_i^M |h_n^{nor}(i) - h_p^{nor}(i)|/\sigma_S\right) \quad (2)$$

여기서 $w_{n,p}$ 는 n 번째 프레임과 p 번째 프레임간의 유사도 값을, σ_S 는 가중상수를 나타내며, 가중치의 민감도를 조절한다. 또한 $h^{nor}(i)$ 는 정규화 된 히스토그램을 나타낸다. 이러한 가중치 $w_{n,p}$ 가 적용된 p 번째 프레임의 히스토그램 $h_p^w(i)$ 는 다음과 같다.

$$h_p^w(i) = \sum_{n \in N(p)} w_{n,p} h_n(i) \quad (3)$$

$N(p)$ 는 p 의 인접한 프레임의 범위를 나타낸다. (3)에서 구한 히스토그램에 한계치를 적용하여 최종적으로 변이 탐색 범위를 구하기 위해 다음과 같이 한계치를 정의한다.

$$T_h = \begin{cases} 2B & \text{if } d < 0 \\ \lfloor B/2 \rfloor + 1 & \text{otherwise} \end{cases} \quad (4)$$

정의된 한계치를 이용해 변이 탐색 범위를 다음과 같이 추정할 수 있다.

$$R = \{k | D(i) - B/2 \leq k \leq D(i) + B/2, h[i] > T_h\} \quad (5)$$

이 기법의 문제점으로 가장 먼저 장면이 전환됨으로 인한 깊이 값의 불연속적인 변화를 고려하지 않은 점을 들 수 있다. 이 경우, 장면

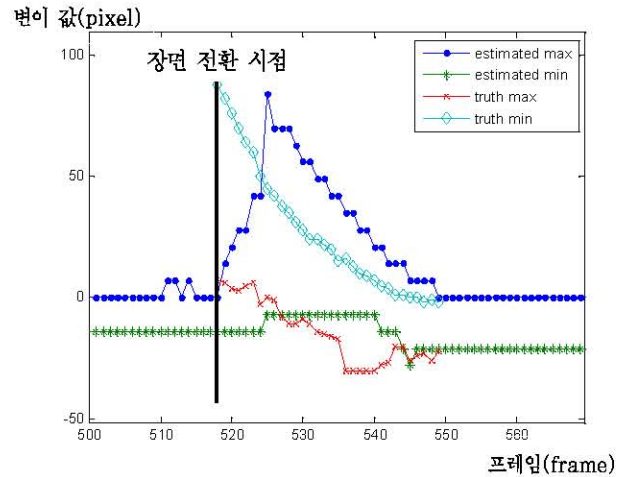


그림 1. 기존 방법의 문제점

전환 전의 히스토그램이 전환 후의 히스토그램에 영향을 미치게 되며 그 반대로 가능하다. 이로 인해 장면 전환 전, 후의 깊이 차이가 크다면 성능에 부정적으로 작용할 수 있으며, 따라서 장면 전환 시점을 찾아내어 그 시점 기준으로 전 후의 영상이 서로 영향을 미치지 않게 해야 한다.

두 번째 문제점으로는 프레임간의 시간적인 유사도가 가중치 $w_{n,p}$ 에 고려되지 않았다는 점이다. 이에 대한 영향으로, 상대적으로 가까이 있는 프레임과 멀리 있는 프레임이 시간적으로 같은 가중치를 갖게 된다. 이는 탐색 범위 추정에 악영향을 미칠 수 있다.

마지막으로 프레임 당 추출할 수 있는 대응점 수의 차이가 클 수 있다는 점을 고려하지 않았다는 문제점을 들 수 있다. 이에 따라 히스토그램을 정규화 하지 않고 가중된 히스토그램을 구하거나 한계치를 단순히 Bin의 크기 B 에 따라 적용한 점은 상대적으로 대응점이 적게 추출되는 프레임에서 주변의 프레임의 영향을 너무 많이 받게 되며 이 또한 성능에 부정적인 영향을 미칠 수 있다. 결과적으로 [2]의 성능은 시간 변화에 대해서 깊이 변화가 크지 않은 비디오에 관해서는 좋은 성능을 보일 수 있지만, 시간 변화에 따라 깊이 변화가 큰 비디오에 관해서는 좋은 성능을 내지 못한다.

그림 1은 장면이 바뀌고 난 후 깊이가 크게 변화했다가, 시간의 흐름에 따라 다시 점진적으로 변하는 영상에 대한 기존 기법의 실험 결과이다. 여기서 truth max과 truth min는 각각의 프레임에서 직접적으로 구해본 탐색범위의 실제 최대 값, 최소 값이며, estimated max와 estimated min은 기존 알고리즘을 이용하여 추정된 탐색범위의 최대 값과 최소 값이다. 이 둘을 비교해 보면 상당히 다름을 알 수 있다. 따라서 이 기법은 연속적으로 깊이 분포가 거의 변하지 않는 영상에서는 상당히 좋은 성능을 낼 수 있지만, 그림 1의 예와 같이 깊이 분포가 시간적으로 상당한 변화를 보이는 영상에 대해서는 좋지 못한 성능을 낼 수 있다. 따라서 다음 장에서 우리는 이러한 문제점을 해결하기 위하여 개선된 기법을 제안한다.

3. 제안하는 탐색 범위 추정 기법

제안하는 기법에서 변이 값 히스토그램은 기존의 기법과 마찬가지로 SURF 알고리즘을 이용하여 (1)과 같은 식을 통해 구한다. 제안



(a) 520번째 프레임



(b) 3532번째 프레임
그림 2. 실험 영상 예시

하는 기법에서의 가중치 $w_{n,p}$ 는 다음과 같다.

$$w_{n,p} = \exp\left(-\sum_i^M d_{diffusion}(h_n^{nor}(i), h_p^{nor}(i))/\sigma_S\right) + |n-p|/\sigma_T$$

(6)

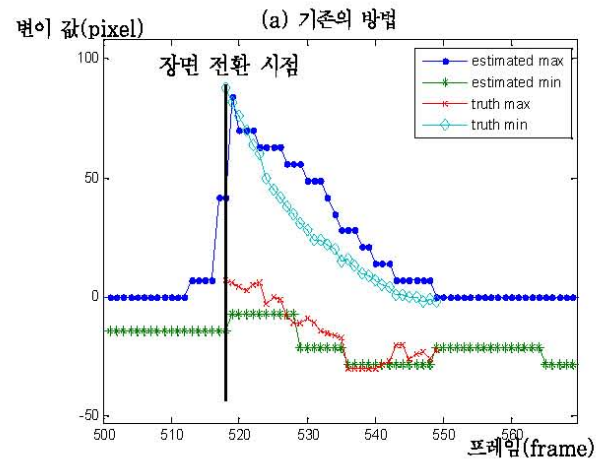
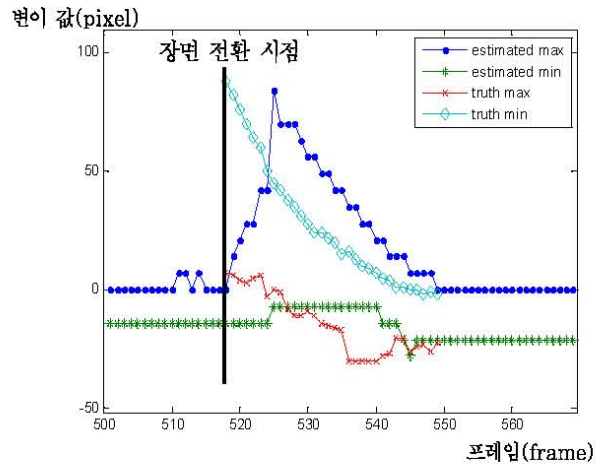
우리는 보다 더 세밀하게 히스토그램의 유사도를 측정하기 위해 히스토그램 확산 거리(Diffusion distance for histogram comparison) [6]을 사용하였다. $d_{diffusion}(h_n^{nor}(i), h_p^{nor}(i))$ 는 $h_n^{nor}(i)$ 과 $h_p^{nor}(i)$ 와의 히스토그램 확산 거리를 나타낸다. 또한 프레임간의 시간적 유사도를 적용하기 위하여 $|n-p|/\sigma_T$ 를 식에 추가하였다. 이는 $|n-p|$ 가 커질수록 즉, 비교하는 프레임들의 시간 차가 커질수록 $w_{n,p}$ 가 작아지게 한다. σ_T 는 σ_S 와 같은 가중 상수를 나타내며, 시간적 유사도에 대한 가중치의 민감도를 결정한다.

또한 제안된 알고리즘에서는 장면전환 시점을 고려하고, 가중치를 변이 값 히스토그램에 가하여 사용하는 대신, 정규화 된 변이 값 히스토그램에 가하여 아래와 같이 사용하였다.

$$h_p^w(i) = \sum_{n \in N'(p)} w_{n,p} h_n^{nor}(i)$$

(7)

$N'(p)$ 는 p의 주변 프레임들을 가리키지만, $N(p)$ 와는 다르게 p의 앞의 장면전환 시점 전의 프레임들과 p의 뒤의 장면전환 시점 후



(b) 제안한 방법

그림 3. 기존의 방법과 제안한 방법의 성능 비교 1

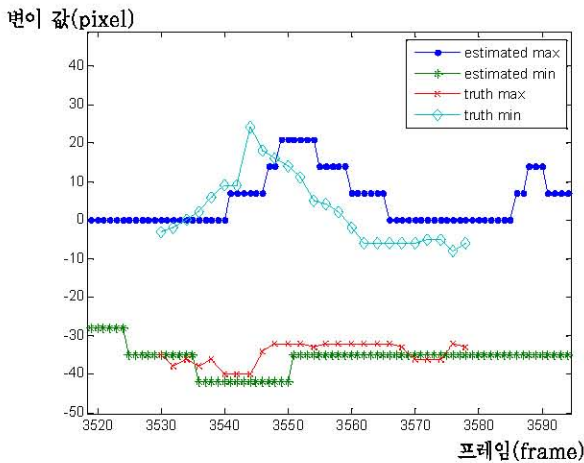
의 프레임들이 있을 경우 제외하게 된다. 장면전환 시점은 Canny Edge 알고리즘[7]을 기반으로 Edge의 변화량에 따라 장면 전환을 찾아내는 특징 기반의 알고리즘[8]을 사용하였다. 또한 $h_n(i)$ 대신 $h_n^{nor}(i)$ 를 사용함으로써 대응점이 적게 추출되는 프레임과 많이 추출되는 프레임의 기여도가 균형을 이루도록 하였다.

가중치가 적용된 히스토그램이 바뀌었기 때문에 이에 따라 한계치도 새롭게 적용해 주어야 한다. 사용하는 히스토그램이 정규화 되어 0에서 1사이의 값을 가지기 때문에 사용할 한계치 또한 0에서 1사이의 값을 가져야 한다. 이 논문에서는 한계치를 다음과 같이 결정 하였다.

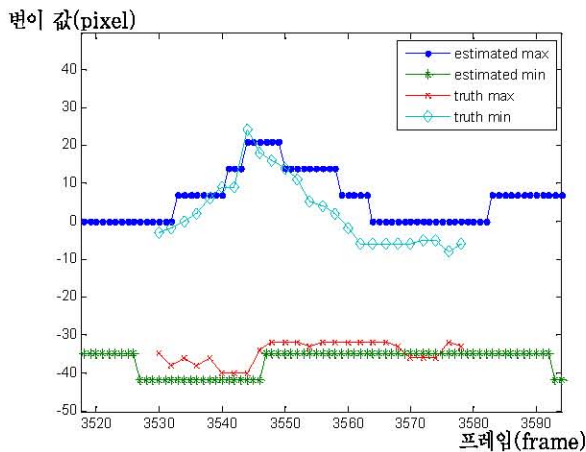
$$T_h = \begin{cases} 0.2 & \text{if } d < 0 \\ 0.1 & \text{otherwise} \end{cases}$$

(8)

기존의 논문에서와 같이 음의 값을 갖는 변이 값(영상 표면보다 들어가 보이게 하는 변이 값)이 양의 값을 갖는 변이 값(영상 표면보다 나와 보이게 하는 변이 값)보다 덜 중요하기 때문에, 보다 높은 한계치를 적용하도록 하였다. 개선된 식 (6), (7), (8)을 사용하여 최종적으로 탐색 범위를 구하는 식은 (5)와 같게 하였다.



(a) 기존의 방법



(b) 제안한 방법

그림 4. 기존의 방법과 제안한 방법의 성능 비교 2

4. 실험 결과

기존의 알고리즘을 사용하여 탐색 범위를 구한 결과와 제안한 알고리즘을 이용하여 탐색 범위를 구한 결과를 비교하였다. 실험에 사용한 영상은 'Skydiving' 영상을 사용하였으며, 그림 2와 같은 장면에서 해당하는 실험 결과를 그림 3, 4에 나타내었다. 결과 그래프를 비교해 보면 기존의 알고리즘에 비해 제안한 알고리즘의 성능이 상당히 좋은 것을 알 수 있다. 그림 4에서 기존 방법의 결과를 보면 장면전환 시점을 고려하지 않아, 장면 전환 후의 갑작스러운 깊이 변화가 제대로 인지되지 못했으며, 프레임간의 시간적 유사도도 고려하지 않아 성능에 부정적인 영향을 미쳤지만, 제안한 방법은 이를 고려하여 성능이 개선된 것을 확인할 수 있다. 그림 5의 경우 장면전환 시점이 주변에 없고, 깊이 변화만 있는 부분에 대해서도 성능이 개선되었음을 확인할 수 있다.

표 1은 그림 4와 그림 5에 해당하는 부분에서 실제 값과 추정 값 사이의 상관도(cross correlation)을 구한 결과이다. 결과를 보면 제안한 알고리즘을 사용했을 때 실제 변이 탐색 범위와의 상관도가 더 높음을 알 수 있으며, 따라서 제안한 방법이 기존의 방법보다 더 좋은 성능을 나타낼 수 있다.

	최대 범위		최소 범위	
	기존 방법	제안 방법	기존 방법	제안 방법
그림 2	0.2238	0.8807	0.0915	0.9086
그림 3	0.5288	0.9034	0.4751	0.7723

표 1. 기존 방법과 제안 방법의 상관도 비교

5. 결론

본 논문에서는 스테레오 비디오의 변이 탐색 범위를 추정하는 방법을 제안하였다. 기존의 방법에서 시간의 흐름에 따라 깊이 값이 변하는 경우 성능적인 면에서 취약했었던 문제점을 개선하였다. 개선한 방법은 알고리즘이 장면 전환 시점과 프레임 간의 시간 유사도를 고려하도록 하였으며, 보다 세밀한 히스토그램 유사도를 추정하기 위해 히스토그램 확산 거리를 사용하였다. 그리고 히스토그램에 가중치를 적용하고 한계치를 사용할 때, 정규화 된 히스토그램을 사용함으로써 추출되는 대응점 수의 영향 없이 프레임 간에 균일한 중요도를 갖도록 하였다. 실험의 결과에서 제안한 방법이 효과적으로 기존 방법의 성능을 개선하였음을 보여주었다.

6. 참고 문헌

- [1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int'J. Computer Vision*, vol. 47, no. 1-3, pp. 7-42, Apr. 2002.
- [2] D. Min, S. Yea, Z. Arican, and A. Vetro, "Disparity search range estimation: enforcing temporal consistency," *Proc. IEEE ICASSP*, 2010.
- [3] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *International Joint Conference on Artificial Intelligence*, pp. 674-679, 1981.
- [4] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," *Carnegie Mellon University Technical Report CMUCS-91-132*, Apr. 1991.
- [5] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robust Features," *Computer Vision and Image Understanding(CVIU)*, vol. 110, no. 3, pp. 346-359, 2008.
- [6] H. Ling and K. Okada. Diffusion distance for histogram comparison. *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 246 - 253, 2006.
- [7] John Canny, "A Computational Approach to Edge Detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 679-698, 1986.
- [8] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classifying scene breaks," *proc. ACM Multimedia*, pp.189-200, 1995.