

## 다시점 스테레오 영상 기반 3차원 깊이정보 획득 기술 연구

\*박순용, \*이민재, \*\*배툼, \*\*\*엄기문, \*\*\*정원식

\*경북대학교 IT대학 전자공학부

\*\*경북대학교 IT대학 컴퓨터학부

\*\*\*한국전자통신연구원 테라미디어연구그룹

## 3D Depth Reconstruction Technique based on Multi-view Stereo Images

\*Park, Soon-Yong, \*Lee, Min-jae, \*\*Pathum Bandara, \*\*\*Um, Gi-Mun \*\*\*Cheong, Won-Sik

\*Electronics Engineering, Kyungpook National University

\*\*Computer Science and Engineering, Kyungpook National University

\*\*\*Tera-Media Research Group, Electronics and Telecommunications Research Institute

## 요약

본 논문에서 개발하고자하는 다시점 스테레오 영상 기반의 3차원 깊이 정보 획득 기술은 스테레오 비전, light field, 가상시점, 방송 콘텐츠, 등 다양한 분야의 기술이 융합된 기술로 연구의 중요성이 매우 높다. 본 논문에서는 SGM 기반의 멀티베이스라인 스테레오 정합 기술을 개발하고 다시점 스테레오 영상에 적용하여 깊이 정보를 획득하였다. 두 시점 간의 스테레오 정합에 있어서 다방향의 에너지 최소화 기술을 적용하고 시점 간의 정합비용함수를 누적하여 마지막으로 S공간 누적방법으로 최적의 깊이영상을 획득하였다. 기존의 스테레오 정합에 비하여 멀티베이스라인 스테레오 정합의 성능 향상을 확인하고 Middlebury 스테레오 영상을 이용하여 성능을 분석하였다.

## 1. 서론

스테레오 정합은 두 카메라에서 획득한 영상으로부터 동일한 물체의 영상좌표를 추출하여 물체의 3차원 거리 정보를 획득하는 기술이다. 스테레오 정합을 정합 픽셀의 밀도에 따라 구분하면 영상의 모든 픽셀에 대한 시차를 계산하는 밀집(dense) 스테레오 정합이 있고, 또는 영상의 특징점(feature point)에 대한 시차를 계산하는 성긴(coarse) 스테레오 정합이 있다. 한편 스테레오 정합의 기술적 측면에서 구분하면 지역(local) 정합과 전역(global)정합으로 나눌 수 있다. 지역적 스테레오 정합은 영상의 일부영역의 값들을 비교하여 그 차이가 최소가 되는 점을 정합점으로 간주하고, 반면 전역 방법은 영상의 스캔 라인 전체의 정합 에너지를 최소화 시키는 방식이다. 전역기반 방식은 일반적으로 속도와 정확도의 측면에서 지역 방식보다 우수하다고 알려져 있다.

광역기반 방식으로는 그래프컷(Graph cut), 신뢰확산방식(Belief propagation)과 SGM(Semi-global matching), 동적 프로그래밍(Dynamic programming) 등이 있다. 한편 최근 스테레오 정합 연구에서는 딥러닝(Deep Learning)을 기반으로 정합을 구현하는 연구가 보고되고 있고 그 성능도 우수한 것으로 나오고 있다. 기존의 일반적인 방법 중에서 그래프컷과 신뢰확산방식은 확률모델을 바탕으로 하여 영상의 각 픽셀에 해당하는 모든 깊이값들에 대해서 확률을 구한 다음 가장 확률이 높은 깊이값을 선택하는 방법이다. SGM의 경우는 한 점에 대해서 8방향 또는 16방향에서 에너지를 누적한 후, 모든 방향의 누적된 값들을 더하여 더해진 에너지가 최소가 되는 깊이값을 해당 점의

깊이값으로 정하는 방법을 사용한다. 이 점은 다이나믹 프로그래밍이 정합선(scan-line)내에서만 광역에너지가 최소가 되도록 하는 것에 비하여 상하좌우와 대각선까지 포함한 여러 방향의 광역에너지를 고려하기 때문에 더욱 정확도가 높은 결과를 낸다.

스테레오 정합은 두 개의 카메라에서만 적용되는 것이 아니라 카메라가 3대 이상인 경우에도 적용할 수 있다. 이러한 경우를 멀티베이스라인(Multi-baseline) 스테레오 정합이라 한다. 스테레오 정합에 사용하는 카메라의 수가 많을수록 영상을 저장하는 메모리의 증가나 계산시간이 길어지는 단점이 있지만 시점이 다른 영상정보를 동시에 사용하는 장점이 있어 시차 계산의 정밀도가 높아지는 장점이 있다.

본 연구에서는 멀티베이스라인 스테레오 정합 기술을 다루고자 한다. 특히 SGM 스테레오 정합을 멀티베이스라인 스테레오 정합으로 확장하여 시차 계산의 성능을 높이는 연구를 진행하였다. 특히 3가지의 멀티베이스라인 스테레오 정합 기술을 제안하고 그 성능을 비교하였다.

## 2. 다시점 스테레오 정합 기술

다시점 스테레오 영상의 정합을 이용한 거리정보의 획득은 M. Okutomi[1]가 발표한 "A multiple-baseline stereo"로 베이스라인 거리가 다른 다시점의 스테레오 카메라를 이용하여 기존의 두 개의 카메라를 이용한 스테레오 정합보다 정확한 거리영상을 획득하였다.

그림 1과 같이 N개의 다시점 카메라를 사용하여 스테레오 정합을

수행하는 조건을 가정해보자. 그림에서  $B_{ij}$ 는 인접한  $i, j$  시점 사이의 베이스라인 거리이고,  $d_{ij}$ 는  $i, j$  시점 사이의 시차이다. 구하고자 하는 물체의 한 점  $P$ 는 3차원 벡터로 표현된다.  $N$ 개의 스테레오 카메라가 있다면 두 시점 사이에 스테레오 정합을 수행할 수 있고 따라서 총  $N C_2$  조합으로 스테레오 정합을 실행할 수 있다. 따라서  $N C_2$  개의 시차 영상을 획득할 수 있다.

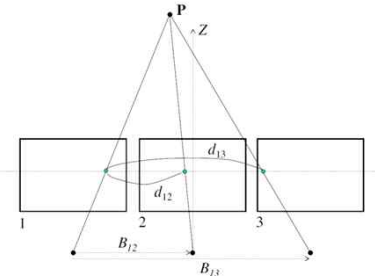


그림 1. 다시점 스테레오 카메라의 시차와 베이스라인 관계

그림 1과 같이  $N=3$  인 경우를 고려해보자. 세 개의 시점을 가진 다시점 스테레오 카메라에서 물체의 한 점  $P$ 까지의 거리  $Z$ 는 시점 1과 2사이의 시차  $d_{12}$ , 시점 1과 3 사이의 시차  $d_{13}$ 에 의하여 다음과 같이 표현할 수 있다.

$$d_{12} = \frac{B_{12}}{Z}, d_{13} = \frac{B_{13}}{Z} \quad (1)$$

### 3. 멀티베이스라인 정합비용의 합성

SGM[2] 기반의 멀티베이스라인 스테레오 정합 기술의 마지막 방법으로 SGM의 S-비용공간을 합성하는 방법을 제안하였다. 두 장의 스테레오 영상의 정합에서 SGM은 정합비용  $C_1$ 을 8방향 또는 16방향의 누적비용합수  $L_i(x,d)$ 를 생성한다. 여기는  $i$ 는 누적방향에 대한 인덱스이다. 그리고 모든 누적방향에 대한  $L_i(x,d)$ 의 합(sum)을 구하여  $S(x,d)$  공간을 생성하고 여기서 가장 작은 비용을 가진 시차를 결정한다.

$$L_r(P, d) = C_1(P, d) + C_2(P, d) + \min(a, b, c, d) - E \quad (2)$$

$$S(P, d) = \sum_r L_r(P, d) \quad (3)$$

여기서  $C_1$ 은 기존의 두 카메라 사이의 비용함수이고  $C_2$ 가 멀티베이스라인의 컬러일관성 함수이다. 멀티베이스라인 스테레오 정합에서도 동일한 알고리즘을 적용할 수 있고 단 5대의 카메라에서 획득한 스테레오 영상에서 아래 그림과 같이 4개의 서로 다른 S-비용공간을 생성할 수 있다. 예를 들어 image1과 image2사이의 SGM 스테레오 정합으로  $S_{12}(x,d)$ , image1과 image3사이의 SGM 스테레오 정합으로  $S_{13}(x,d)$ , image1과 image4사이의 SGM 스테레오 정합으로  $S_{14}(x,d)$ , image1과 image5사이의 SGM 스테레오 정합으로  $S_{15}(x,d)$  비용공간을 생성할 수 있다. 이들 4개의 비용공간의 합(sum)을 구하기위해서는 시차 공간을 정규화해야한다. 따라서 4개의 S-비용공간을 정규화한 후  $S_{12}(x,d), S_{13}(x,d), S_{14}(x,d), S_{15}(x,d)$ 를 합하여 최종적으로 합성된 비용공간  $S_1(x,d)$ 을 생성할 수 있다. 최종 S-비용함수 공간에서 최소의 비용을 가진 시차를 결정하게 된다.

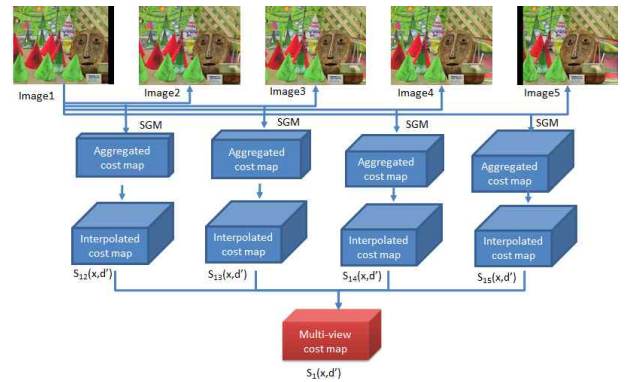


그림 2. 멀티베이스라인 카메라 영상의 S-비용공간의 합성 방법

### 4. 멀티베이스라인 스테레오 정합 실험

본 연구에서는 제안한 멀티베이스 라인 스테레오 정합 기술에 대한 실험을 진행하였다. 실험은 주로 스테레오 정합에 많이 사용되는 미들버리 (Middlebury) 스테레오 벤치마크 영상 중 Teddy, Cones, Venus, 영상을 사용하였다. 또한 정량적인 비교를 위하여 미들버리 영상의 스테레오 정합의 평가에 사용되는 MSE 오차 분석을 통하여 시차 영상의 정확도를 분석하였다. 시차 영상의 오차 분석은 두 가지의 영상 마스크 (mask)를 사용하였다. 첫째는 모든 픽셀의 오차를 분석하는 All 마스크를 사용하였고, 두 번째는 폐색영역을 제외한 픽셀의 시차의 오차를 계산하기위하여 nonocc 마스크를 사용하였다.

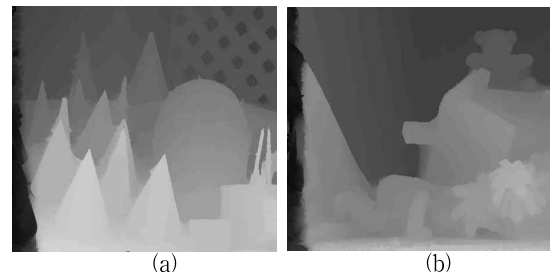


그림 3. 다시점 영상 기반 스테레오 정합 결과

Data Set	Non Occlusion Mask	All Mask
Cones	5.911708	33.692605
Teddy	3.905063	27.050343
Venus	3.111348	3.873346

### 논문사사

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천기술 개발)

### 참고문헌

[1] M. Okutomi and T. Kanade, "A multiple-baseline stereo," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 4, pp. 353-363, (1993).  
 [2] H. Hirschmüller, "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information," Proceedings of IEEE conference, CPVR, 2005.  
 [3] K. He, J. Sun, and X. Tang, "Guided image filtering." ECCV 2010.