

# 3D 콘텐츠 생성에서의 스테레오 매칭 알고리즘에 대한 매칭 비용 함수 성능 분석

홍광수\* · 정연규\* · 김병규\*

## 요 약

스테레오 매칭 과정에 있어서 매칭 비용을 구하는 것은 매우 중요한 과정이다. 이러한 스테레오 매칭 과정의 성능을 살펴보기 위하여 본 논문에서는 기존에 제안된 매칭 비용 함수들에 대한 기본 개념들을 소개하고 각각의 성능 및 장점을 분석하고자 한다. 가장 간단한 매칭 비용 함수는 매칭 되는 영상의 일관된 밝기를 이용하여 좌, 우 영상 간 서로 대응하는 대응점을 추정하는 과정으로, 본 논문에서 다루는 매칭 비용함수는 화소 기반과 윈도우 기반의 매칭 비용 방법으로 크게 두 가지로 나눌 수 있다. 화소 기반의 방법으로는 절대 밝기차(the absolute intensity differences: AD)와 sampling-intensitive absolute differences of Birchfield and Tomasi (BT) 방법이 있고, 윈도우 기반의 방법으로는 차이 절대 값의 합(sum of the absolute differences: SAD), 차이 제곱 값의 합(sum of squared differences: SSD), 표준화 상호 상관성(normalized cross-correlation: NCC), 제로 평균 표준화 상호 상관성(zero-mean normalized cross-correlation: ZNCC), census transform, the absolute differences census transform (AD-Census) 이 있다. 본 논문에서는 앞서 언급한 기존에 제안된 매칭 비용 함수들을 정확도와 시간 복잡도를 측정했다. 정확도 측면에서 AD-Census 방법이 평균적으로 가장 낮은 매칭 오차를 보여줬고, 제로 평균 표준화 상호 상관성 방법은 non-occlusion과 all 평가 항목에서 가장 낮은 매칭 오차를 보여 주지만, discontinuities 평가 항목에서는 블러 효과 때문에 높은 매칭 오차를 보여 주었다. 시간 복잡도 측면에서는 화소 기반인 절대 밝기차 방법이 낮은 복잡도를 보여 주었다.

## Performance Analysis of Matching Cost Functions of Stereo Matching Algorithm for Making 3D Contents

Gwang-Soo Hong\* · Yeon-Kyu Jeong\* · Byung-Gyu Kim\*

## ABSTRACT

Calculating of matching cost is an important for efficient stereo matching. To investigate the performance of matching process, the concepts of the existing methods are introduced. Also we analyze the performance and merits of them. The simplest matching costs assume constant intensities at matching image locations. We consider matching cost functions which can be distinguished between pixel-based and window-based approaches. The Pixel-based approach includes absolute differences (AD) and sampling-intensitive absolute differences (BT). The window-based approach includes the sum of the absolute differences, the sum of squared differences, the normalized cross-correlation, zero-mean normalized cross-correlation, census transform, and the absolute differences census transform (AD-Census). We evaluate matching cost functions in terms of accuracy and time complexity. In terms of the accuracy, AD-Census method shows the lowest matching error ratio (the best solution). The ZNCC method shows the lowest matching error ratio in non-occlusion and all evaluation part. But it performs high matching error ratio at the discontinuities evaluation part due to blurring effect in the boundary. The pixel-based AD method shows a low complexity in terms of time complexity.

**Key words :** Depth map, Matching cost, Matching Cost function, Stereo matching

## 1. 서 론

최근 3D영화의 흥행과 더불어 3D 영상 제작 기술들이 높은 관심을 받고 있는데, 스테레오 매칭 분야는 변이 정보를 획득하기 위해서 꾸준히 연구 되고 있는 분야이다. 스테레오 매칭에 있어서 매칭 비용을 결정하는 것은 매우 중요한 과정이다. 대부분의 스테레오 알고리즘은 좌, 우 영상간의 유사성을 측정하는 방법을 가지고 있다. 이런 유사성을 나타내는 척도를 매칭 비용이라 하는데 매칭 비용은 각각의 화소에 모든 변이에서 계산되어진다. 가장 간단한 매칭 비용 함수는 매칭되는 이미지의 일관된 밝기를 이용하여 좌, 우 영상 간 서로 대응하는 대응점을 추정한다. 그러나 더 견고한 매칭 비용 함수 모델은 노이즈의 변화에도 대응점을 추정한다. 그레이스케일 또는 색상 이미지에서의 일반적인 화소 기반 매칭 비용 방법 (pixel-based matching cost method)에는 절대 밝기차(absolute intensity difference: AD) [1] 와 밝기 차 제곱(squared intensity differences: SD) [2], sampling-insensitive absolute differences [3]나 truncated version이 있다. 일반적인 윈도우 기반 매칭 비용 방법(window-based matching cost method)은 차이 절대 값의 합(sum of absolute the differences: SAD), 차이 제곱의 합(Sum of squared differences: SSD), 표준화 상관성(normalized correlation: NC), 표준화 상호상관성(normalized cross-correlation: NCC), 제로 평균 표준화 상호상관성(zero-mean normalized cross correlation: ZNCC)이 있다. 그리고 Census transform [4]과 Census transform을 개선한 AD-Census [5] 방법이 있다.

본 논문은 다음의 구성은 다음과 같다. 2장에서는 매칭 비용 함수에 대해서 소개하고, 3장에서는 언급된 매칭 비용 함수를 이용한 실험 결과에 대해 살펴본다. 마지막으로 4장에서는 결론에 대한 내용을 끝으로 본 논문을 마친다.

## 2. 매칭 비용 (Matching cost)

매칭 비용은 좌, 우 영상 간의 화소나 그 주변의 값들을 비교하여 일치하는 지점을 찾는 방법이다. 좌,

우 영상 간 매칭 정도를 판단하기 위한 방법으로 화소기반 방법, 영역 기반 방법으로 나눌 수 있다. 화소기반의 방법으로 절대 밝기 차는 상응하는 화소에서 각 화소의 밝기의 일관성을 이용한 매칭 비용 추정 방법으로 식은 다음과 같다.

$$C_{AD}(p, d) = |I_L(p) - I_R(p-d)| \quad (1)$$

여기서  $I_L$ 은 좌 영상,  $I_R$ 은 우 영상,  $p$ 는 매칭 비용을 구하는 화소를 나타내고  $d$ 는 변이를 나타낸다. 일반적으로 변이  $d$ 는  $d_{\min}$  에서  $d_{\max}$  까지의 범위를 가지고 모든 변이 범위의 매칭 비용을 구한다.

추가적으로, 본 논문에서는 Birchfield et al. [3]이 제안한 sampling-insensitive absolute difference 방법을 사용했다. 이 방법은 상응하는 화소와 선형 보간된 주변 화소의 절대 거리를 계산하는 방법으로 식은 다음과 같다.

$$\begin{aligned} C_{BT}(p, d) &= \min(A, B) \\ A &= \max(0, I_L(p) - I_R^{\max}(p-d), I_R^{\min}(p-d) - I_L(p)) \\ B &= \max(0, I_R(p-d) - I_L^{\max}(p), I_L^{\min}(p) - I_R(p-d)) \\ I^{\min}(p) &= \min(I^-(p), I(p), I^+(p)) \\ I^{\max}(p) &= \max(I^-(p), I(p), I^+(p)) \\ I^-(p) &= (I(p) - [1 \ 0]^T) / 2 \\ I^+(p) &= (I(p) + [1 \ 0]^T) / 2 \end{aligned} \quad (2)$$

윈도우 기반의 방법으로는 차이 절대 값의 합이 있다. 이 방법은 매칭 비용을 구하고자하는 화소의 주변 영역의 화소를 이용한 것으로, 일반적으로 사각형의 윈도우를 사용한다.

$$C_{SAD}(p, d) = \sum_{q \in N_p} |I_L(q) - I_R(q-d)| \quad (3)$$

다음에 소개할 윈도우 기반 매칭 비용 방법은 확률적으로 가우시안 노이즈 처리에 최적화된 방법으로 표준화 상호 상관성 방법이다 [6]. 그러나 표준화 상호 상관성 방법은 깊이 불연속 지점에서의 블러 효과

를 이끄는 경향이 있다. 식은 다음과 같다.

$$C_{NCC}(p, d) = \frac{\sum_{q \in N_p} I_L(q) \cdot I_R(p-d)}{\sqrt{\sum_{q \in N_p} I_L(q)^2 \cdot \sum_{q \in N_p} I_R(q-d)^2}} \quad (4)$$

표준화 상호 상관성 방법에서 추가된 제로 평균 표준화 상호 상관성을 소개한다. 제로 평균 표준화 상호 상관성 매칭 비용 방법은 밝기의 차이와 독립적으로 평균과 표준편차의 정규화에 의한 차이로 구해질 수 있다. 식은 다음과 같다.

$$C_{ZVCC}(p, d) = \frac{\sum_{q \in N_p} (I_L(q) - \bar{I}_L(p))(I_R(q-d) - \bar{I}_R(p-d))}{\sqrt{\sum_{q \in N_p} (I_L(q) - \bar{I}_L(p))^2 \sum_{q \in N_p} (I_R(q-d) - \bar{I}_R(p-d))^2}} \quad (5)$$

여기서  $\bar{I}_L$ 은 좌 영상의 평균 픽셀 값을 나타내고,  $\bar{I}_R$ 은 우 영상의 평균 픽셀 값을 나타낸다. 다음의 윈도우 기반 방법으로는 census transform이 있다 [4]. 식은 다음과 같다.

$$C_{census} = \otimes_{i,j} \zeta(p, p + [i, j]) \quad (6)$$

여기서

$$\zeta(p, p + [i, j]) = \begin{cases} 1, & \text{where } p > p + [i, j] \\ 0, & \text{otherwise.} \end{cases}$$

여기서 화소  $p$ 와  $p$  근접 이웃 화소에 대한 윈도우를 1차원 배열로 재배열하고 중심 화소보다 작으면 "0", 크거나 같으면 "1"로 변환하여 census vector를 형성한다. 좌, 우 영상 간 상응하는 census vector를 구하고 좌, 우 영상에서 구해진 census vector간의 해밍 거리를 구한다. 여기에 해밍 거리의 값이 매칭 비용으로 저장된다.

AD-Census transform [5]방법은 Census transform을 개선한 방법으로, 반복되는 영역이나 유사한 지역적 구조를 가지는 영역에 대해서 더 견고한 방법이다. 식은 다음과 같다.

$$C_{ADcensus}(p, d) = \rho(C_{census}(p, d), \lambda_{census}) + \rho(C_{AD}(p, d), \lambda_{AD}) \quad (7)$$

식 (7)에서  $C_{census}(p, d)$ 는 census transform의 매칭 비용이고 RGB 채널의 절대 밝기 차의 평균의 값을 나타내는  $C_{AD}(p, d)$ 는 다음과 같이 정의된다.

$$C_{AD}(p, d) = \frac{1}{3} \sum_{i=R, G, B} |I_i^L(p) - I_i^R(p-d)| \quad (8)$$

식(7)에서의  $\rho(c, \lambda)$ 는 다음과 같이 정의된다.

$$\rho(c, \lambda) = 1 - e^{-\frac{c}{\lambda}} \quad (9)$$

여기서  $c$ 는  $\lambda$  값에 의해 조절되는 입력 값으로, 본 논문에서는 census transform을 위한 임계값,  $\lambda_{census} = 30$ , 절대 밝기의 평균의 값을 위한 임계값,  $\lambda_{AD} = 10$  으로 구현하였다.

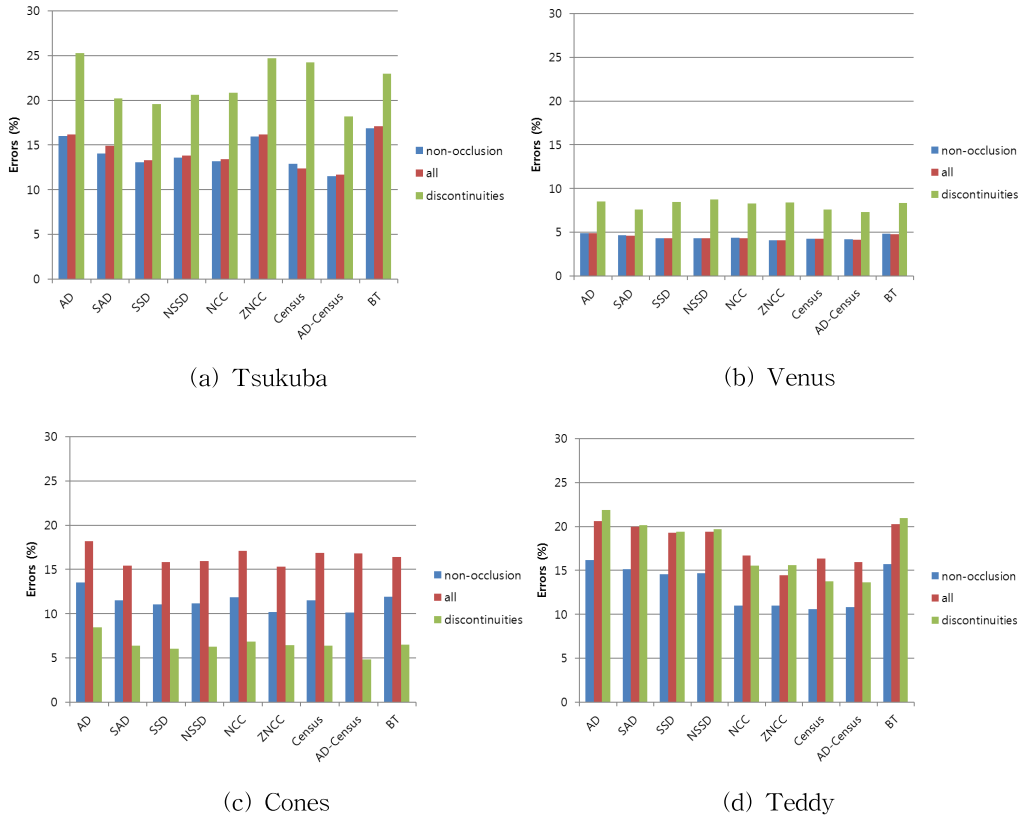
본 논문에서는 실험을 위한 매칭 비용 함수로 절대 밝기 차, 차이 절대 값의 합, 차이 제곱 값의 합, 표준화 절대 제곱 값의 합, 표준화 상호 상관성, 제로 평균 표준화 상호 상관성 AD-Census, BT를 사용하였다.

깊이 영상 추정을 위해 매칭 비용 계산 과정 후 매칭 비용 통합 (aggregation cost) 과정에서  $9 \times 9$  윈도우를 사용하였고, 최적화 단계에서 WTA (winner-take-all) 방법을 사용하였다 [8]. WTA 방법의 식은 아래와 같이 표현된다.

$$d_p = \operatorname{argmin}_{d \in S_d} C(d). \quad (10)$$

### 3. 실험 결과 및 고찰

본 논문에서 사용된 이미지는 평행 카메라 모델에



(그림 1) 참조 영상에서의 매칭 오차

의해서 획득된 영상이다. 이러한 영상들은 대부분 양의 시차가 발생하는 시점에 의해 얻은 것이므로 좌영상을 기준으로 우 영상이 얼마만큼 왼쪽으로 이동했느냐를 변이로 선택하였다. 실험의 결과는 middlebury 웹사이트 [7]에서 제공하는 평가 방법을 통해 성능을 평가 하였으며, 성능 분석을 위해 Tsukuba, Venus, Teddy, Cones 영상을 사용했다. 깊이 영상 추정 정확도는 앞서 언급한 구조광(structured light) 기술 [9]에 의해 추정된 참조 영상의 groundtruth 영상을 이용하여 측정한다. 본 논문에서는 깊이 영상의 추정의 정확도와 속도, 두 가지 측면에서 매칭 함수를 분석하였다. 실험한 결과 영상의 성능 평가는 Scharstein et al. [8] 가 제안한 방법을 사용하였는데, 평가

항목은 non-occlusion, all, discontinuities의 3가지에 의해서 평가된다.

Non-occlusion은 패색 영역이 아닌 부분에 대한 깊이 값 에러 화소의 개수를 나타내고, all은 전체 영역에 대한 깊이 값 에러 화소의 개수, discontinuities는 깊이 값이 변하는 영역에 대한 깊이 값 에러 화소의 개수를 나타낸다.

그림 1은 참조 영상인 Tsukuba, Venus, Cones, Teddy에서의 각 매칭 비용 방법에 대한 세 가지 평가 항목에서의 매칭 에러율을 나타낸 것이다. Tsukuba 참조 영상의 평가 항목인 그림 1.(a)에서 Tsukuba 영상의 윤곽선이 많은 특징을 가지고 있어 discontinuities 평가 항목이 다른 항목들 보다 대체로 높게 나오

는 것을 볼 수 있다. 화소 기반의 방법 중에 절대 밝기차 방법이 BT 방법보다 전체적으로 약 1%의 낮은 에러율을 보였다. 윈도우 기반의 방법들 중 non-occlusion 항목에서 가장 낮은 매칭 에러율을 보인 방법은 AD-Census 방법으로 11.5%의 에러율을 보였고, all 항목과 discontinuities 항목에서도 각각 11.6%와 18.1%의 에러율로 가장 낮은 매칭 에러율을 보였다.

그림 1(b)에서는 화소 기반인 방법들 중 BT 방법이 절대 밝기차 방법보다 0.2% 낮은 매칭 에러율을 보였고, 윈도우 기반의 방법들 중에서는 제로 방법인 non-occlusion과 all 항목에서 다른 방법들 보다 우수하나 discontinuities 항목에서 높은 매칭 에러율을 보였다. Census transform 방법은 평균적으로 제로 평균 표준화 상호 상관성 방법보다 0.2% 낮은 매칭 에러율을 보였다.

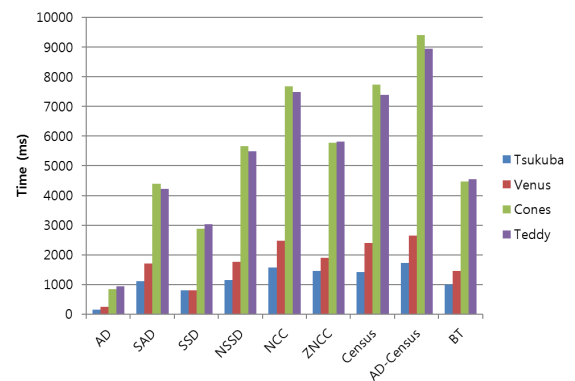
그림 1(c)에서는 대체로 all 항목의 에러율이 높는데, cones 영상이 패색 영역이 많은 영상이므로 이 같은 결과가 나왔다. 화소 기반의 방법들 중 BT 방법이 절대 밝기차 방법보다 전체적으로 약 1.2%의 낮은 매칭 에러율을 보였고, 윈도우 기반 방법들 중 제로 평균 표준화 상호 상관성 방법이 평균적으로 가장 낮은 에러율을 보였다. AD-Census 방법은 제로 평균 표준화 상호 상관성 방법보다 all 항목에서 2%의 높은 매칭 에러율을 보였고, discontinuities 항목에서 2% 낮은 매칭 에러율을 보였다.

그림 1(d)에서는 화소 기반의 방법들 중에 절대 밝기차 방법이 BT 방법보다 1% 낮은 에러율을 보였고, 윈도우 기반 방법들 중에서는 제로 평균 표준화 상호 상관성 방법과 AD-Census 방법이 가장 낮은 매칭 에러율을 보였다.

제로 평균 표준화 상호 상관성 방법과 Census transform 방법, 그리고 AD-Census 방법이 다른 매칭 함수 낮은 매칭 에러율을 가지는 것을 볼 수 있다. 제로 평균 표준화 상호 상관성 방법 같은 경우는 윤곽선이 많은 영상에서 Census transform 방법과 AD-Census 방법보다 높은 매칭 에러율을 보였고, Census transform 방법은 평균적으로 낮은 매칭 에러율을 보였지만, 비 텍스처 영역에서 높은 매칭 에러율을 보였다. AD-Census 방법도 Census transform 방법과 마찬가지로 평균적으로 낮은 매칭 에러율을 보였

고, 비 텍스처 영역과 깊이 불연속 지점에서 낮은 매칭 에러율을 보였다.

그림 2는 매칭 함수들의 시간 복잡도를 나타낸 그래프이다. 매칭 함수들 중에 절대 밝기차 방법이 가장 낮은 복잡도를 가지고 있고, AD-Census 방법이 가장 높은 복잡도를 가지고 있는 것을 볼 수 있다. 윈도우 기반 방법들 중에서는 차이 제곱 값의 합 방법이 가장 낮은 복잡도를 가지고 있다.



(그림 2) 시간 복잡도 비교 그래프

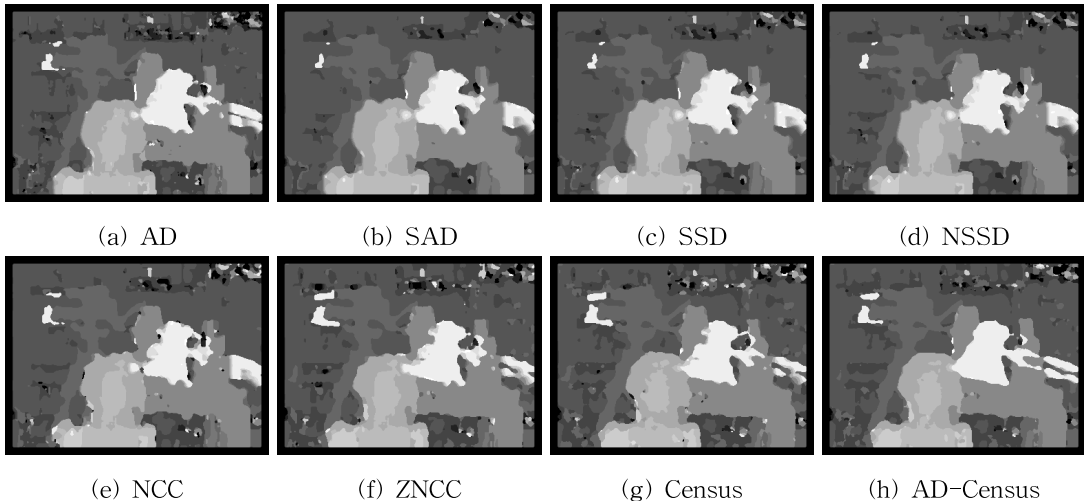
그림 3~4는 Middlebury 웹사이트 [7]에서 제공하는 각 참조 영상의 매칭 함수들을 이용한 추정된 깊이 영상을 나타낸 것이다. 개별 방법에 따라서 주관적 품질이 달라짐을 알 수 있고 또한 실험 영상에 따라서 적용 방법의 성능이 달라짐을 확인할 수 있다.

## 4. 결 론

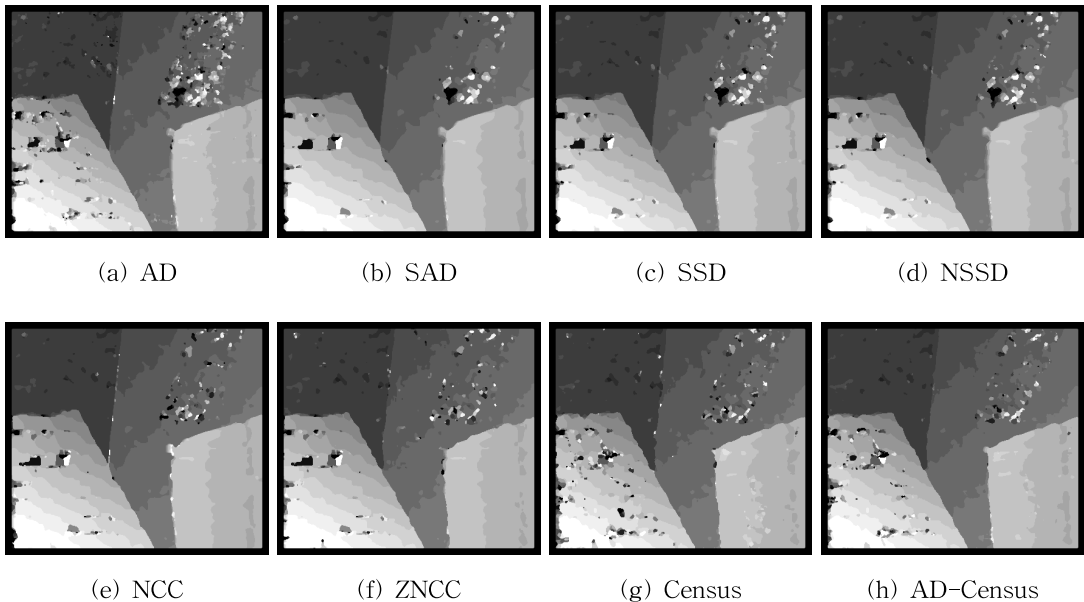
본 논문에서는 현재 스테레오 매칭을 위해 많이 사용되는 8개의 매칭 비용 함수를 깊이 영상 추정의 정확도 측면과 시간 복잡도 측면에서 비교를 하였다. 정확도 측면에서는 매칭 비용 함수들 중 평균적으로 AD-Census 방법이 평균적으로 가장 낮은 매칭 에러율을 보였고, 제로 평균 표준화 상호 상관성 방법은 AD-Census 방법보다 non-occlusion 평가 항목과 all 평가 항목에서 1% 낮은 매칭 에러율을 보여줬지만, 블러링 효과 때문에 깊이 불연속 지점에서 2% 높은

매칭 에러 율을 보여줬다.

낮은 매칭 시간 복잡도 측면에서는 화소 기반인 절대 밝기차 방법이 가장 낮은 시간 복잡도를 보여주었고, 윈도우 기반 방식에서는 차이 제곱 값의 합 방법



(그림 3) 추정된 깊이 영상 (Tsukuba 영상)



(그림 4) 추정된 깊이 영상 (Venus 영상)

이 가장 낮은 시간 복잡도를 보여줬다.

## 참고문헌

- [1] T. Kanade, "Development of a video-rate stereo-machine," In Image Understanding Workshop, pp. 549-557, 1994.
- [2] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," International Journal Computer Vision, vol. 2, no. 3, pp. 283-310, 1989.
- [3] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 4, pp. 401-406, April 1998.
- [4] R. Zabib and J. Woodfill, "Non-parametric Local transforms for computing visual correspondence," Third Europe an conference on Computer Vision, May 1994.
- [5] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," GPUCV 2011.
- [6] H. Hirschmuller, P. R. Innocent, and J M. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors," International Journal of Computer Vision, vol. 47, no. 1/2/3, pp. 229-246, April-June 2002.
- [7] <http://vision.middlebury.edu/stereo>
- [8] D. Scharstein and R. Szeliski, "A Taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision, vol. 47, no. 1, pp. 7-42, May 2002.
- [9] D. Scharstein and R. Szeliski, "High-Accuracy Stereo depth maps using structured light," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 195-202, 2003.

## [저자 소개]



홍 광 수 (Gwang-Soo Hong)

2011년 02월 선문대학교 컴퓨터공학  
이학사  
2012년 03월 선문대학교  
컴퓨터공학 석사  
2013년 06월 선문대학교  
컴퓨터공학과 박사과정  
(현재)

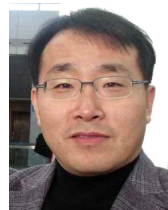
email : hongzolv@mpcl.sunmoon.ac.kr



정 연 규 (Yeon-Kyu Jeong)

2012년 2월 선문대학교 컴퓨터공학과  
이학사  
2012년 ~ 선문대학교 컴퓨터공학과  
석사과정

email : JeongYK@mpcl.sunmoon.ac.kr



김 병 규 (Byung-Gyu Kim)

1995년 02월 부산대학교 전기공학과  
공학사  
1998년 02월 한국과학기술원  
전기및전자공학 석사  
2004년 02월 한국과학기술원  
전기및전자공학 박사  
2004년~2008년 한국전자통신연구원  
선임연구원  
2009년~2012년 선문대학교  
컴퓨터공학과 조교수  
2013년~ 현재 선문대학교  
컴퓨터공학과 부교수

email : bg.kim@ieee.org