

---

# 영상의 깊이정보 추출을 위한 weighted cost aggregation 기반의 스테레오 정합 기법

윤희주\* · 차의영\*

Weighted cost aggregation approach for depth extraction of stereo images

Hee-Joo Yoon\* · Eui-Young Cha\*

## 요 약

스테레오 비전 시스템(stereo vision system)은 2차원 영상정보를 이용하여 3차원 깊이 정보를 획득하는 데 유용한 방법으로, 그동안 많은 연구가 진행되었다. 3차원 깊이 정보를 획득하기 위해서는 영상의 대응점을 찾아야 하는데, 속도와 정확성을 동시에 만족시키기가 어렵다. 이러한 문제를 해결하기 위해, 본 논문에서는 적용적 가중치(weight)를 적용한 cost aggregation 기반의 스테레오 정합 기법을 제안한다. 이 방법은 스테레오 영상의 특징을 이용하여 가중치를 획득하고, 색상정보, 밝기정보에 거리정보를 이용하여 가중치를 적용한 후, 이를 이용하여 대응점을 찾아 깊이 정보를 추출한다. 제안된 방법의 성능을 평가하기 위하여 ground truth가 존재하는 다양한 스테레오 영상을 이용하여 실험하였으며, 실험 결과 다양한 영상에서도 향상된 결과를 보였다.

## ABSTRACT

Stereo vision system is useful method for inferring 3D depth information from two or more images. So it has been the focus of attention in this field for a long time. Stereo matching is the process of finding correspondence points in two or more images. A central problem in a stereo matching is that it is difficult to satisfy both the computation time problem and accuracy at the same time. To resolve this problem, we proposed a new stereo matching technique using weighted cost aggregation. To begin with, we extract the weight in given stereo images based on features. We compute the costs of the pixels in a given window using correlation of weighted color and brightness information. Then, we match pixels in a given window between the reference and target images of a stereo pair. To demonstrate the effectiveness of the algorithm, we provide experimental data from several synthetic and real scenes. The experimental results show the improved accuracy of the proposed method.

## 키워드

Disparity map, depth extraction, stereo matching, weighted matching, cost aggregation

## I. 서론

스테레오 비전 시스템(stereo vision system)은 양안을 이용해 원근감을 느끼는 인간의 시각체계를 모방한 것으로, 두 대 이상의 카메라를 이용하여 획득한 2차원 영상 정보로부터 3차원 정보를 획득하는 것이다[1]. 스테레오 비전 시스템에서 가장 중요하면서도 시간을 요하는 일은 스테레오 정합(stereo matching) 문제를 해결하는 것이다.

스테레오 정합은 스테레오 영상 간의 대응점(correspondence point)을 찾아서 이 점들 간의 시차(disparity)를 구하는 과정으로 3차원 깊이 정보를 획득하는데 있어서 가장 중요한 과정이라 할 수 있다. 스테레오 정합 문제를 해결하기 위한 방법으로는 크게 특징 기반 방법(feature based approach)과 영역 기반 방법(area based approach)이 있다.

특징 기반의 스테레오 정합은 스테레오 영상의 특징을 추출하여 각 영상의 탐색영역 내에서 정합 후보 점들을 비교하는 정합 방법으로, 특징 추출에 사용되는 특징으로는 에지(edge), 분할된 경계선(line segment), 선소(edge segment), 윤곽선(contour), 모서리(corner) 등이 있다. 특징 기반 방법은 처리속도를 향상시킬 수 있는 반면, 영상 전체 영역에 대한 정확한 시차 획득이 불가능하다는 단점이 있어 깊이 정보를 획득하기 어렵다.

영역 기반의 스테레오 정합은 상관관계 기반 스테레오 정합(correlation based stereo matching)이라고도 하며, 정합을 결정하기 위해서 스테레오 영상의 정합 윈도우의 밝기 정보 변화를 이용하므로 잡영(noise)에 약한 면이 있으나, 전체 시차 정보의 획득이 가능하다[2,3].

본 논문에서는 스테레오 영상의 대응점을 찾아 3차원 깊이 정보를 획득하기 위해 적응적 가중치를 적용한 cost aggregation 기반의 스테레오 정합 기법을 제안한다.

본 논문은 다음과 같이 구성되어 있다. II장에서는 전체 시스템의 동기와 개괄을 서술하고, III, IV장에서는 본 논문에서 제안하는 가중치 획득 방법과 cost aggregation 기반의 스테레오 정합 방식에 대해 설명한다. V장에서 실험 결과를 보여주고, VI장에서 결론을 맺는다.

## II. 전체 시스템 개요

그림 1은 본 논문에서 제안하는 전체 시스템의 구조를 도식화한 것이다. 평행하게 설치된 두 대의 카메라를 통해 획득한 스테레오 영상의 특징 정보를 이용하여 가중치를 획득하고, 각 영상의 색상 정보, 밝기 정보, 거리 정보에 가중치를 적용한 후, 기준영상에 대한 대응점을 찾아 정합하여 3차원 깊이 정보를 추출한다. 그림 1의 세부적인 내용은 다음 장에서 설명된다.

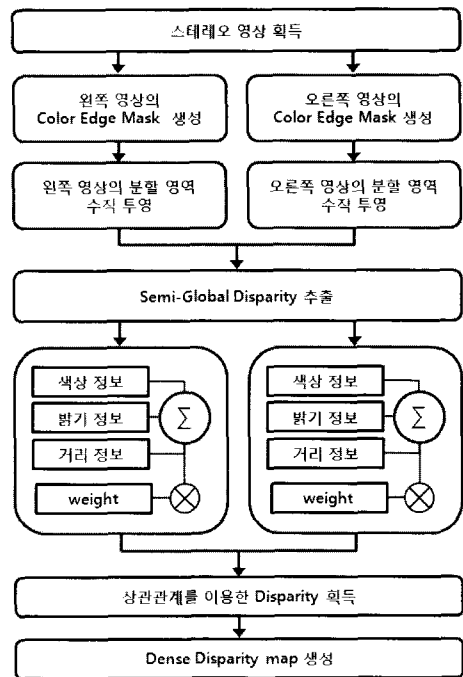


그림 1 제안된 알고리즘 구조도  
Fig 1 The flowchart of the algorithm

## III. Semi-global disparity 획득

본 논문에서 스테레오 영상의 깊이정보를 추출하기 위해 사용될 상관관계 기반의 정합 방법은 윈도우 크기에 민감하므로 이를 보완하기 위하여 정합에 사용될 특징정보에 적응적 가중치(weight)를 적용한 후 정합을 수행한다. 가중치를 이용한 스테레오 정합 방법에 대해 많은 연구가 진행되고 있으며, 반복적으로 가중치를 생성

하는 방법을 많이 채택하고 있다[4]. 이러한 반복적 가중치 생성 방법의 경우, 초기 가중치에 민감하며 연산 시간이 오래 걸리는 단점을 가지고 있다[5,6]. 이를 보완하기 위하여 본 논문에서는 각 영상의 특징 정보를 이용하여 semi-global disparity를 획득하고, 이를 가중치로 사용하는 방법을 제안하였다.

### 3.1 Color edge mask 생성

Semi-global disparity를 추출하기 위해 각 영상의 특징 정보를 이용한다. 에지(edge) 정보를 영상의 특징 정보로 사용할 경우, 대부분의 에지 추출 방법이 영상의 밝기 정보만을 사용하므로 정확한 특징을 추출하기 어렵다. 본 논문에서는 영상의 색상 정보를 포함하고, 잡영(noise)에 강인한 특징을 추출하기 위해 아래 식 (1)과 같은 방법을 사용하여 color edge mask,  $M$ 을 생성하였다.

$$M_C = (D_C - I_C), \quad C \in \{R, G, B\} \quad (1)$$

식(1)의  $D$ 는 팽창(dilation) 연산을 수행한 영상,  $I$ 는 원본 영상을 나타내며,  $R, G, B$  채널 각각에 대해 모두 수행한다.

### 3.2 Semi-global disparity 추출

좌·우 영상에서 유사도가 높은 대응점을 찾기 위해 비용 함수(cost function)를 사용한다. 비용함수로는 SAD(sum of absolute differences), SSD(sum of squared differences), MAD(mean of absolute differences), NCC(normalized cross correlation) 등이 있으며, 본 논문에서는 가중치 생성을 위한 유사도 측정에 SSD 비용함수를 사용한다. 비용함수는 일반적으로 정방형의 윈도우(window)를 사용하므로 윈도우의 크기에 따라 다른 결과를 초래하게 되며, 많은 연산을 필요로 한다. 본 논문에서는 식 (1)을 이용하여 획득한 color edge mask를 다음 식 (2), (3)과 같이 분할 영역(sub-block)으로 나누어 수직 투영 성분을 획득하여 상관관계를 계산한다.

$$P_{S_i}(x) = \sum_{y \in S_i} M_C(x, y) \quad (2)$$

$$F_{S_i}(x, g) = \{P_{R_{S_i}}(x) - P_{L_{S_i}}(x + g)\}^2 \quad (3)$$

여기서,  $P_{S_i}$ 는 분할 영역  $S_i$ 의 수직 투영 결과이며,  $i$ 는 분할 영역의 개수를 나타내며,  $F_{S_i}$ 는 분할 영역별 투영 결과의 SSD 수행과정을 식으로 표현한 것으로 식 (3)의 결과가 최소값을 가지게 되는  $g$ 가 semi-global disparity,  $g^*$ 가 된다.

## IV. 상관관계를 이용한 비용(cost) 생성

상관관계 기반의 스테레오 정합은 윈도우 내의 밝기 정보 변화를 이용하여 정합을 결정하므로, 윈도우의 크기에 따라 다른 결과를 초래한다. 윈도우의 크기가 작으면 비교적 정확한 물체의 경계를 찾을 수 있으나 잡영이 많이 생기게 되고, 윈도우의 크기가 크면 잡영은 줄일 수 있으나 물체의 경계가 확장되는 문제점이 있어 상관관계 기반의 스테레오 정합은 3차원 깊이 정보를 추출하기에 문제점이 많이 존재한다[7]. 이를 보완하기 위해 본 논문에서는 각 영상의 밝기 정보 뿐 아니라, 색상, 거리 정보의 상관관계를 이용하여 비용(cost) 값을 구한 뒤, 이들을 합하여 정합을 수행한다.

### 4.1 밝기 정보를 이용한 비용 계산

스테레오 영상은 평행하게 설치된 두 대 이상의 카메라를 통해 획득한 영상이므로 동일한 위치의 물체라 하더라도 외부 광원에 의해 밝기 차가 생길 수 있다. 이러한 밝기 차가 오정합을 유발할 수 있으므로 입력 영상에 히스토그램 스트레칭(histogram stretching)기법을 적용하여 영상을 개선한 뒤, 밝기(brightness) 정보를 이용하여 비용을 계산한다.

$$C_B(B_R^*, B_L^*) = \frac{\sum_{x, y \in W} \{B_R^*(x, y) - B_L^*(x + d, y)\}^2}{(255 \times w)} \quad (4)$$

$$B^*(x, y) = \frac{B(x, y) - B_{\min}}{B_{\max} - B_{\min}} \times 255 \quad (5)$$

식 (5)는 영상의 히스토그램 스트레칭 과정을 수식화한 것으로, 식 (4)의  $B_L^*$ 과  $B_R^*$ 은 히스토그램 스트레칭을 통해 개선된 밝기 정보 영상을 나타내며,  $W$ 는 윈

도우를 나타내며  $w$ 는 윈도우의 크기를 나타낸다. 식 (4)는 상관관계 함수를 비용 집합(cost aggregation) 정합 기법에 적용하기 위하여 픽셀의 최대 밝기 차(255)에 윈도우 크기를 곱한 ( $255 \times w$ )로 나누어 0과 1사이의 값으로 정규화(normalize)한 것이다.

#### 4.2 색상 정보를 이용한 비용 계산

스테레오 비전은 인간의 시각적 인지 능력을 착안하여 만들어진 것으로 인간의 색인지는 중요한 정보이다. HSI모델은 인간의 색인지에 기반을 둔 사용자 지향성의 색상모형으로 Hue는 어떤 구체적인 색상을 만들기 위해 색을 조합할 필요가 없이 Hue의 좌표축 자체가 색상을 나타낸다. Hue는 0과 360 사이의 각도 값으로 표현되므로, 일반적인 형태의 유사도(similarity)를 구하는 상관관계 함수를 사용할 경우 정확한 상관관계를 구할 수 없다. 그러므로 sine과 cosine 함수를 이용하여 180도 차이를 나타내는 보색과의 거리를 가장 먼 거리(비유사도)로 설정하도록, 다음과 같이 식 (6)을 사용한다.

$$SM(h_1, h_2) = 1 - \frac{\sqrt{S_1 + S_2}}{2} \quad (6)$$

$$S_1(x, y) = \{\sin(h_1(x, y)) - \sin(h_2(x, y))\}^2$$

$$S_2(x, y) = \{\cos(h_1(x, y)) - \cos(h_2(x, y))\}^2$$

식 (6)의 경우, 두 색상  $h_1$ 과  $h_2$ 의 유사도가 높을수록 큰 값을 가지게 되므로, 본 논문에서 제안한 비용을 집합하여 사용하기 위하여, 식 (6)을 식 (7)과 같이 비유사도(dis-similarity)를 구하는 식으로 수정하여 사용한다.

$$C_{SM}(h_R, h_L) = \frac{\sqrt{S_1 + S_2}}{2} \quad (7)$$

#### 4.3 거리정보를 이용한 가중치 적용

스테레오 영상에서 최적의 비용을 가지는 정합점을 찾기 위해 semi-global disparity를 이용하여 가중치 적용 후, 비용을 집합하여 정합한다.

Semi-global disparity,  $g^*$ 와의 거리를 가중치로 사용하기 위하여 다음 식 (8)을 이용하여 각 픽셀의 가중치를 생성한다.

$$W(x, d) = \frac{(x + d) - g^*}{\max((width - g^*), g^*)} \quad (8)$$

$$f(x, y, d) = \operatorname{argmin}_{x, y \in W} \{(C_B + C_{SM})W(x, d)\} \quad (9)$$

식 (9)는 식 (4)와 식 (7)의 비용 계산 함수에 식 (8)의 가중치 함수를 적용한 비용함수로  $f$ 가 최저값을 가지는  $d$ 가 기준영상의 깊이가 된다.

### V. 실험 결과

본 논문에서 제안된 방법은 Pentium4 3.0GHz, 1GB RAM의 PC환경에서 Visual C++.NET으로 구현하였으며, 성능 평가를 위해 ground truth가 존재하는 다양한 스테레오 표준 영상을 이용하였다.

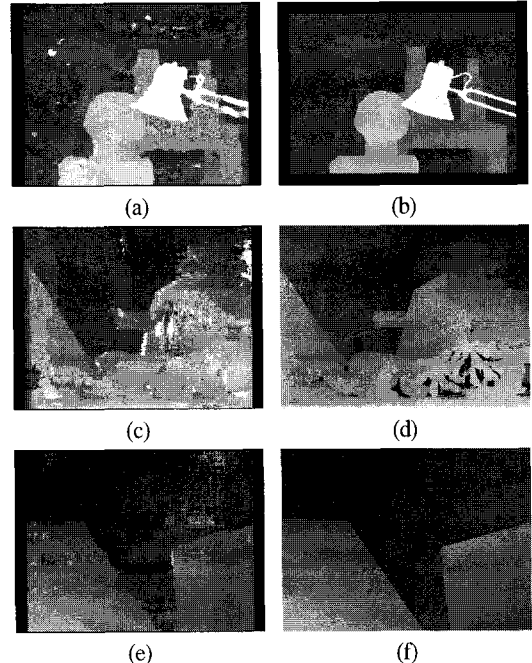


그림 2. 시차지도 (a) tsukuba (384×288) (b) ground truth (c) teddy (450×375) (d) ground truth (e) venus (434×383) (f) ground truth

Fig 2. The disparity map (a) tsukuba (384×288) (b) ground truth (c) teddy (450×375) (d) ground truth (e) venus (434×383) (f) ground truth

그림 2는 본 논문에서 제안한 방법을 통해 획득한 다양한 스테레오 영상의 시차지도를 나타낸 것이다. 그림 2의 스테레오 영상은 모두  $11 \times 5$  크기의 동일 윈도우를 사용했으며,  $g^*$ 의 생성에 사용되는 분할영역의 개수는 3으로 설정하였다. 그림 2에서 보는 바와 같이, 다양한 환경에서의 서로 다른 크기의 영상에서 동일한 크기의 윈도우를 통해 비교적 정확한 깊이 정보를 획득한 것을 알 수 있다.

## VI. 결론 및 향후 과제

본 논문에서는 스테레오 영상으로부터 정합점을 찾아 3차원 깊이 정보를 획득하기 위하여 색상, 밝기 정보를 이용하여 비용(cost)을 계산한 뒤, 거리 정보를 이용하여 가중치를 적용하는 방법을 제안하였다. 제안한 방법에서는 색상정보와 에지(edge)정보를 포함하는 color edge mask를 생성하여 semi-global disparity를 획득하고, 이를 가중치로 사용함으로써 반복적으로 가중치를 생성하는 방법에 비해 보다 빠르고 정확하게 가중치를 생성할 수 있었으며, 윈도우 크기에 민감한 단점을 보완할 수 있었다. 또한 제안한 방법을 통해, 다양한 영상에서 비교적 깊이 정보가 잘 표현된 시차지도 생성함을 알 수 있었다. 그러나 더욱 정확한 시차지도 생성하기 위하여 반복적 무늬(texture)에 따른 폐색영역(occlusion)에 대한 정합 오류 문제에 대한 연구가 요구된다.

## 참고문헌

- [1] Clark F. Olson, "Maximum-Likelihood Image Matching," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, Issue 6, Page(s) : 853-857, June 2002
- [2] YongSheng Chen, YiPing Hung, "Fast Block Matching Algorithm Based on the Winner- Update Strategy," IEEE Transactions on Image Processing, Vol. 10, Issue 8, Page(s) : 1212-1222, August 2001
- [3] Federico Tomabari, Stefano Mattocchia, Luigi Di Stefano, "Segmentation-based adaptive support for accurate stereo correspondence," PSIVT, pp.427-438, 2007
- [4] Kuk-Jin Yoon, In So Kweon, "Adaptive support-weight approach for correspondence search", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.28, Issue 4, Page(s):650-656, April 2006
- [5] Yihua Xu, Dongsheung Wang, "Stereo Computation using Radial Adaptive Windows", IEEE Pattern Recognition Proceedings 16th International Conference, Vol. 3, Page(s) : 595-598, August 2002
- [6] Yilei Zhang, Minglun Gong, Yee-Hong Yang, "Local Stereo Matching with 3D Adaptive Cost Aggregation for Slanted Surface Modeling and Sub-Pixel Accuracy," IEEE, Pattern Recognition 2008 ICPR 19th International Conference, Vol.10, Page(s):1-4, Dec 2008
- [7] T. Kanade, M. Okutomi, "A stereo matching algorithm with an adaptive window : theory and experiment", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.16, Issue 9, Page(s) : 920-932, September 1999

## 저자소개



윤 희 주 (Hee-Joo Yoon)

2004년 : 동서대학교 공과대학  
정보통신공학 공학사

2006년 : 부산대학교 공과대학  
멀티미디어공학 공학석사

2008년 : 부산대학교 공과대학 컴퓨터공학과 박사  
수료

※ 관심분야 : 스테레오비전, 영상처리, 로봇비전



차 의 영 (Eui-Young Cha)

1979년 : 경북대학교 공과대학  
전자공학 공학사  
1982년 : 서울대학교 자연대학  
전자계산학 이학석사

1998년 : 서울대학교 공과대학 컴퓨터공학 공학박사

1981년~1985년 : 한국전자기술연구소 연구원

1995년~1996년 : University of London 방문교수

1985년~현재 : 부산대학교 공과대학

정보컴퓨터공학부 교수

※관심분야 : 로봇비전, 신경회로망, 웨이블릿