

Feature 인식을 위한 2D 가우시안 파라미터 기반 스테레오 매칭 방법

박용희

(주)리주비놀

paul@rjvnor.com

Feature-Aware Stereo Matching Using 2D Gaussian Parameter

Yong Hee Park

Rejuvenol Corp.

요약

본 논문은 stereo matching 시 disparity를 더욱 정밀하게 계산하기 위한 가우시안 파라미터 벡터 방식을 제안한다. 일반적으로 stereo matching은 이미지의 raw intensity를 사용한다. 이 경우 노이즈 발생 시 매칭 오류가 발생하고, 주변에 비슷한 패턴이 있을 경우 해당 픽셀을 잘못 선택하는 상황이 발생한다. 따라서 본 논문은 feature를 더욱 강조하기 위해 raw 데이터 대신 feature 벡터를 생성해 매칭하는 방법을 제안한다. 벡터 요소로는 주변 픽셀의 패턴을 파악하기 위해 가우시안 파라미터를 사용한다. 매칭 비용 계산은 Birchfield - Tomasi(BT) 알고리즘을 사용하며 해당 feature 벡터에 적합하도록 확장해서 사용한다. 결과적으로 feature가 약한 평면 영역에서 mismatch가 줄어듦을 확인했고 전체적인 정확도 또한 향상하였다.

I. 서론

스테레오 매칭은 두 이미지를 통해 깊이 데이터를 추정하는 중요한 기술로, 3차원 복원, 자율주행, 로보틱스, 공간 인식 등 다양한 분야에서 활용된다. 이를 위해 두 이미지의 픽셀 매칭 관계인 disparity를 계산한다. 정확한 disparity 계산을 위해서는 두 픽셀을 정확히 매칭하는 것이 중요하다 [1].

기존 스테레오 매칭은 raw intensity 기반 비용 알고리즘을 사용해왔다. 그러나 이러한 방식은 노이즈 취약 문제, 저텍스처 영역에서 명확한 매칭 점이 존재하지 않는 문제, 반복 패턴 영역에서 오정합 발생 가능성과 같은 고질적인 문제점이 있다. 이를 해결하기 위해서는 raw intensity 값이 아닌, 픽셀 주변의 local structure를 반영한 feature 벡터가 필요하다.

본 논문에서는 픽셀 주변 패턴을 가우시안 분포 형태로 모델링하여 2차원 가우시안 파라미터 벡터를 구성하고, Birchfield - Tomasi (BT) 알고리즘 계산을 가우시안 특징 벡터의 각 채널에도 동일하게 적용하는 새로운 매칭 구조를 제안한다.

II. 관련연구

스테레오 매칭은 두 이미지의 disparity를 안정적이고 정확하게 계산하는 것을 목적으로 한다. 정합 비용을 얼마나 개인하게 계산하느냐에 따라 disparity의 품질이 크게 달라진다. 본 논문에서는 비용을 계산하기 위해 Birchfield - Tomasi (BT) [2] 알고리즘을 사용한다. BT 알고리즘은 스테레오 매칭 시 발생하는 오류를 줄이기 위해 픽셀의 intensity 값뿐 아니라 인접한 서브픽셀 intensity 범위를 함께 고려하는 알고리즘이다.

III. 본론

본 논문에서는 스테레오 매칭의 정합 비용을 보다 정밀하고 강건하게 계산하기 위해 가우시안 기반 파라미터 벡터 생성 방식과 이를 BT 알고리즘에 확장 적용하는 방법을 제안한다. 제안 방법은 기존 intensity 기반

BT 비용의 장점을 유지하면서, 픽셀 주변의 지역 구조적 정보를 반영할 수 있는 새로운 feature 생성 방식을 도입한다.

정확한 비용 계산을 위해서는 픽셀 주변의 지역 패턴과 구조를 적절히 표현하는 feature가 필요하다. 기존 raw intensity는 단일 스칼라 값으로 이러한 정보를 충분히 반영하지 못하고, 저텍스처 영역이나 노이즈 환경에서는 매우 취약하다. 이를 개선하기 위해 본 논문에서는 각 픽셀 (x, y) 주변의 (n, n) 크기의 윈도우를 적용해 가우시안 기반 파라미터를 산출한다. 이를 통해 해당 영역에서의 질량 중심을 구하고 이를 intensity를 대신해 사용한다. 파라미터는 2D 가우시안 파라미터인 $\mu_x, \mu_y, \sigma_x, \sigma_y$ 를 사용한다. 각 파라미터는 아래와 같이 계산한다.

$$S_0(x, y) = \sum_{(i, j) \in W} I(x + i, y + j) \quad (1)$$

$$\mu_x(x, y) = \frac{1}{S_0(x, y)} \sum_{(i, j) \in W} I(x + i, y + j)i \quad (2)$$

$$\mu_y(x, y) = \frac{1}{S_0(x, y)} \sum_{(i, j) \in W} I(x + i, y + j)j \quad (3)$$

$$\sigma_x^2(x, y) = \frac{1}{S_0(x, y)} \sum_{(i, j) \in W} (I(x + i, y + j))i^2 - \mu_x^2(x, y) \quad (4)$$

$$\sigma_y^2(x, y) = \frac{1}{S_0(x, y)} \sum_{(i, j) \in W} (I(x + i, y + j))j^2 - \mu_y^2(x, y) \quad (5)$$

$$F(x, y) = [\mu_x(x, y), \mu_y(x, y), \sigma_x^2(x, y), \sigma_y^2(x, y)] \quad (6)$$

(1)은 정규화를 위한 식이다. (2)~(5)는 2D 가우시안에 관한 파라미터이다. (6)는 기존 intensity를 대신 사용할 수 있는 feature 벡터이다.

본 논문에서는 기존 intensity를 사용해 계산하는 BT 알고리즘을 확장해서 사용한다. 각 벡터의 단일 채널 비용을 계산 후 weight를 적용하는 방식을 사용한다.

$$C_{BT}(p, q) = \min(\max(0, p - q_{\max}, q_{\min} - p), \max(0, q - p_{\max}, p_{\min} - q)) \quad (7)$$

$$C_k(x, y, d) = C_{BT}(F_{L,k}(x, y), F_{R,k}(x - d, y)) \quad (8)$$

$$k \in \{\mu_x, \mu_y, \sigma_x^2, \sigma_y^2\}$$

$$C(x, y, d) = w_\mu C_{\mu_x}(x, y, d) + w_\mu C_{\mu_y}(x, y, d) + w_\sigma C_{\sigma_x^2}(x, y, d) + w_\sigma C_{\sigma_y^2}(x, y, d) \quad (9)$$

$$C_{sum}(x, y, d) = \sum_{(i, j) \in N} C(i, j, d) \quad (10)$$

$$d(x, y) = \arg \min_{d \in D} C_{sum}(x, y, d) \quad (11)$$

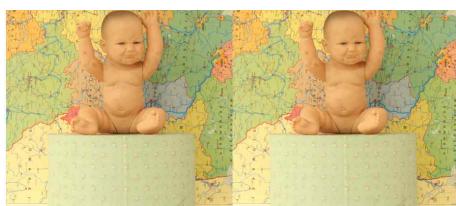
단일 채널 BT 비용은 (7)과 같이 계산된다. (8)에서 Gaussian 파라미터 벡터는 네 개의 채널을 가지며 각 채널에 대한 동일한 BT 비용을 독립적으로 계산한다. (9)에서 최종적으로 각 채널의 비용에 weight를 적용하여 disparity를 구하기 위한 비용을 계산한다.

(10)에서 비용의 합계는 block sum으로 계산된다. 해당 윈도우 크기는 가우시안 파라미터를 구하기 위한 윈도우와 구분된다. (11)에서 disparity 선택은 winner-takes-all (WTA) 방식을 사용한다.

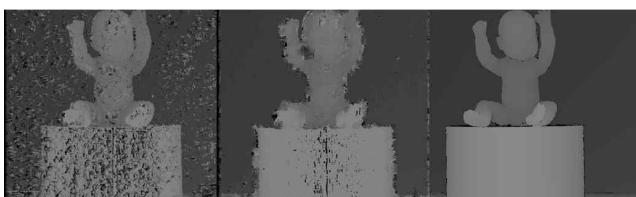
IV. 성능평가

본 논문에서 제안한 방법의 성능을 평가하기 위해 기존 intensity 기반 BT 알고리즘과 본 논문의 파라미터 벡터 기반 BT 알고리즘의 결과를 비교한다. 좌우 이미지 데이터셋과 ground truth disparity는 vision.middlebury.edu [3]에서 가져와 사용했다. 좌우 이미지 샘플은 그림 1과 같다.

본 논문에서 파라미터는 다음과 같이 설정하였다. 가우시안 파라미터를 생성하기 위한 윈도우 크기는 30, 비용 block sum을 위한 윈도우 크기는 5, disparity 탐색 거리는 32, 평균 weight는 1, 분산 weight는 0.8로 설정했다. 스테레오 매칭 결과는 그림 2와 같다.



<그림 1. 이미지 샘플>



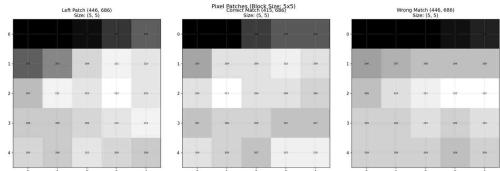
(a) Intensity 기반 BT, (b) 제안한 방법 (c) Ground truth
<그림 2. 스테레오 매칭 결과>

```
1. Normal BT (raw intensity)...  
MAE: 5.800, Accuracy: 52.85%, Time: 5.0s  
2. BT on Gaussian Dist ( $\alpha * BT(\text{mean}_x) + \beta * BT(\text{mean}_y) + \gamma * BT(\text{var}_x) + \delta * BT(\text{var}_y)$ )...  
MAE: 4.074, Accuracy: 68.29%, Time: 13.9s
```

<그림 3. 성능 평가>

검은 색상으로 보이는 부분은 mismatch 된 영역으로 기존 방법 대비 줄어든 것을 확인할 수 있다. 특히 (a)에서는 feature가 적은 평면 부근에서 mismatch가 심한데, 이를 크게 보완한 것으로 확인된다. 다만 얼굴과 팔 영역이 분리가 되지 않는 문제가 있다. 이는 가우시안 윈도우 크기로 인한 문제로 보이며 기존 intensity의 편차가 큰 영역은 오히려 필터처럼 blur된 것이 원인으로 파악된다. 정량적 결과는 그림 3과 같다.

그림 3은 정확도 관련 수치이다. Ground truth와 비교했을 때 정확도는 raw intensity가 52.85%, 가우시안 벡터가 68.29%로 약 15.44% 향상했다. 연산시간은 5.0초에서 13.9초로 2.78배 상승했다. 이는 파라미터를 구하기 위한 preprocess와 복수 채널로 인한 계산량 증가가 원인이다.



(a) 왼쪽 이미지 픽셀, (b) 잘 매칭된 예시 (c) 잘못 매칭된 예시
<그림 4. 픽셀 매칭 예시>

```
==== Gaussian Parameters ====  
Left: mean=(15.03, 16.35), var=(80.01, 76.98)  
Correct: mean=(15.03, 16.36), var=(80.03, 77.04)  
Wrong: mean=(14.88, 16.39), var=(79.98, 76.53)
```

<그림 5. 가우시안 파라미터 예시>

그림 4는 intensity 기반 BT에서 잘못 매칭된 픽셀 영역과 제안한 방법에서 올바르게 매칭된 픽셀 영역 예시이다. 그림 5는 해당 픽셀에서의 가우시안 파라미터의 수치이다. 실제 환경에서 카메라가 동일한 대상을 촬영하여도 항상 같은 값을 가지지 않는다. 따라서 intensity에만 의존할 경우 위와 같이 매칭이 실패하는 경우가 발생한다. 실제로 제안한 방법에서 선택한 픽셀이 intensity 기반 BT 대비 가우시안 파라미터 값이 더 가까운 것을 확인할 수 있다. 이를 통해 2D 가우시안 파라미터가 해당 픽셀 영역에서 feature를 표현할 수 있음을 알 수 있다.

IV. 결론

본 논문은 raw intensity를 대신해 특정 영역의 가우시안 파라미터를 스테레오 매칭에 사용하는 방법을 제안했다. 성능평가에서 시각화를 했을 때 해당 파라미터를 통해 저텍스트 영역에 feature를 표현할 수 있음을 확인하였다. 수치상의 정확도 또한 전체적으로 향상되었음을 확인했다.

하지만 intensity의 편차가 큰 영역에서는 오히려 정확도가 낮아졌다. 추후 벡터 요소로 가우시안 이외 feature를 잘 표현할 수 있는 다른 파라미터를 테스트할 예정이며 cost aggregation 등 확장 적용할 예정이다.

참고 문헌

- [1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision (IJCV), vol. 47, no. 1 - 3, pp. 7 - 42, 2002.
- [2] Birchfield, S., & Tomasi, C. "A pixel dissimilarity measure that is insensitive to image sampling," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 1998.
- [3] <https://vision.middlebury.edu/stereo/data/>