

Relaxation Technique을 이용한 3차원 정보의 추출

○ 김 영 구 조 동 옥 최 병 옥  
한 양 대 학 교

Extraction of the 3-Dimensional Information Using Relaxation Technique

Yeong Gu Kim Dong Uk Cho Byeong Uk Choi

HANYANG UNIVERSITY

ABSTRACT

Images are 2-dimensional projection of 3-dimensional scenes and many problems of scene analysis arise due to inherent depth ambiguities in a monocular 2-D image.

Therefore, depth recovery is a crucial problem in image understading.

This paper proposes modified algorithm which is focused on accurate correspondance in stereo vision.

The feature we use is zero-crossing points and the similarity measure with two property evaluation function is used to estimate initial probability. And we introduce relaxation technique for accurate and global correspondance.

I. 서론

인간의 눈이 깊이정보를 추출하는 것과 같이 2-D image로부터 3-D 정보를 얻고자 하는 human vision system에 관한 연구는 B.Julesz의 random dot stereogram 실험(1971)이후 많이 진행되어 왔었다.

그러한 연구중에서 stereo vision은 laser나 초음파 등을 필요<sup>(1)</sup>로 하지 않고, 기타의 shape from methods<sup>(2)</sup>처럼 실험 환경의 특별한 제약이 없기 때문에 일반적인 scene을 처리 대상으로 할 수 있다는 장점이 있다.

따라서 최근 계산기의 비약적 성능 향상과 sensor 기술의 발전에 따라 중요한 3차원 정보 추출의 도구로 실용화 되고 있으며, 이 stereo vision 방법으로 추출된 정보는 robot의 passive navigation, 지형도 제작, 항공측량, 그리고, 3-D 물체 인식 등을 위하여 중요한 data로서 이용되고 있다.

stereo vision에서는 두 camera의 상대적 변이로 인한 두 화상의 시각차(disparity)를 결정하는 과정이 가장 중요하며, 본 논문에서는 보다 정확한 correspondance를 위하여 zero-crossing feature point에 대해 두 가지의 similarity를 측정하고 이들의 contextual information으로부터 relaxation labeling algorithm을 적용하였다.

그리고, camera geometry에 constraint를 가해 scanline search와 triangulation시 단순화를 기했다.<sup>(3)(4)(5)</sup>

II. 본론

1. Feature Extraction

correspondence ambiguity를 줄이기 위하여 이웃점들과 뚜렷이 구별되는 특별한 점을 선택 이용하는 것이 좋으며, 구석점이나 경계선<sup>(6)</sup>, zero-crossing<sup>(7)</sup>등을 feature point로 이용한 연구가 있었다.

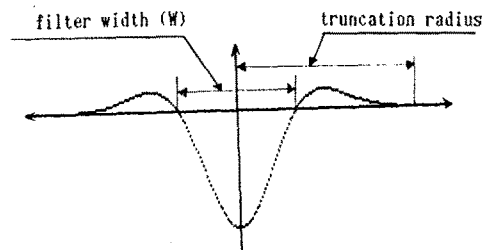
본 논문에서는 Marr와 Poggio가 인간의 시각 mechanism으로 부터 고안한 sombrero filter를 image에 씌워 처리한 후 그로부터 zero-crossing을 추출하여 feature point로 이용하였다. 이 filter는 Gaussian 분포함수에 Laplacian ( $\nabla^2$ )을 취한 것이다.<sup>(8)</sup>

$$\nabla^2 G(x,y) = \left( \frac{x^2+y^2}{\sigma^2} - 2 \right) \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

$$G(x,y) = \sigma^2 \cdot \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

이 filter의 형태는 그림(1)의 모양을 2축 중심으로 회전 시킨 꼴이며 filter의 폭(W)은 filter의 negative portion의 폭(  $2\sqrt{2}\sigma$  ) 과 같다. 본 논문에서는 implementation시에 negative maximum의 0.1%이하 되는 부분은 잘라내었다.

$$\text{truncation radius} \doteq 1.5W \doteq 3\sqrt{2}\sigma$$



$$\sigma=1, W=2.83, \text{ truncation radius}=4.24$$

그림1 sombrero filter의 모양

그리고 zero-crossing은 filter로 처리한 화상을 horizontal scanning 하면서 sign이 변화하거나 zero의 값을 갖는 점을 탐색함으로써 추출하였으며, filtered image의 sign 변화를 zero-crossing의 sign으로 기록하였다.

이 zero-crossing을 feature point로 이용하면 다음과 같은 몇 가지 장점이 있다. gray level이 의미있게 변화하는 변곡점으로서 object의 boundary를 포함하며, 단 하나의 pixel로 나타나기 때문에 두렵지 않아 triangulation시에 정확하고, threshold없이 filtered image의 sign 변화만으로 추출 가능하다. 그러나 filter의 폭이 작으면 macro structure에 너무 민감하며, 폭이 너무 크면 blurring 효과가 생기는 단점이 있다.

2. Candidate의 결정

camera의 축을 평행하게 유지하면서 base line 상에 일정한 간격(b)을 둔 단순한 camera geometry를 설정하였으므로 두 화상에 대한 각 feature point의 disparity는 그림(2)와 같이 화상의 epipolar line과 같은 방향인 X축 방향으로 일어난다.

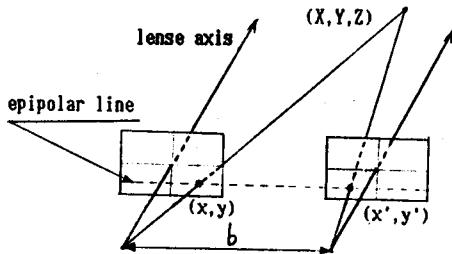


그림2 candidate 점의 탐색 영역

그러므로 (x, y)의 좌표를 갖는 왼쪽 화상의 zero-crossing 점에 대한 오른쪽 화상의 correspondence candidate (x', y')는 다음 두 조건을 만족하는 점들의 집합이다.

- i)  $(R(x', y') - x - DMAX \leq x' \leq x + \epsilon; y - \epsilon \leq y' \leq y + \epsilon)$
- ii) L(x, y)와 R(x', y')의 zero-crossing signal이 같다.

여기서,  $\epsilon$ 은 잡음이나 작은 왜곡에 대한 영향을 고려한 수직 변이이며,  $DMAX (= b \cdot f / Z_{min})$ 는 물체와 camera 사이의 최소 거리에 constraint를 가함으로써 구할 수 있다.

(b: base line의 길이, f: lense의 중심과 화상과의 거리) 따라서 오른쪽의 화상에서 나타날 correspondence point의 위치 범위를 i)와 같이 예측할 수 있다.

본 논문에서는 candidate를 disparity label로 표시하며, 따라서 어떤 특정한 feature point(x<sub>i</sub>, y<sub>i</sub>)에 대한 candidate point들은 그 점에 대한 disparity label set L<sub>i</sub>의 element로 나타낼 수 있다. 그리고 label set L<sub>i</sub>에는 occluding으로 인해 correspondence가 이루어지지 않을 경우에 대응될 no-match label I\*을 둔다.

3. Similarity Measure

적절한 property의 선정과 그에 대한 타당한 similarity 척도의 결정이 올바른 corespondence를 행하기 위하여 중요한 요소이다.

본 논문에서는 Euclidean distance와 gradient direction이라는 두가지 property를 similarity 측정을 위한 정보로 이용하여 어떤 한 property에 대한 의존성을 최소화 하였다.

i) Euclidean distance는 5x5의 사각 window를 왼쪽 화상의 feature point와 candidate point에 씌워 구할 수 있으며, 이로 부터 측정된 similarity S<sub>1</sub>(l)은 Euclidean distance의 재곱인 E(l)에 반비례한다.

$$E(l) = \sum_{x,y} \sum_{x',y'} (L(x,y) - R(x'+l_x, y+l_y))^2$$

$$S_1(l) = \frac{1}{1 + C_1 \cdot E(l)}$$

여기서 C<sub>1</sub>은 E(l)의 weight로서 양의 값이다.

ii) gradient direction은 3x3 방향 미분 operator로 부터 구한 gradient의 vector 표현으로 나타내었으며, gradient direction의 similarity S<sub>2</sub>(l)은 두 gradient direction의 차 G(l)에 반비례한다.

$$G(l) = |GL(x,y) - GR(x+1, y+1)|$$

여기서의  $GL(x,y) = GL(x)I + GL(y)J$ 이다.

$$S_2(l) = \frac{1}{1 + C_2 \cdot G(l)}$$

위에서, C<sub>2</sub>는 G(l)에 대한 weight로서 양의 값이다.

4. Initial Probability 결정

Euclidean distance와 gradient direction을 정보로 하여 측정된 similarity의 합은 각 candidate의 label에 대한 초기 확률값의 결정을 위해 이용된다.

$$S(l) = \sum_{k=1}^n w_k \cdot S_k(l)$$

$$(0 \leq S(l) \leq 1)$$

이며 w<sub>1</sub>, w<sub>2</sub>는 각 similarity 척도에 대한 weight로서 w<sub>1</sub> + w<sub>2</sub> = 1 이다.

일반적으로 maximum similarity를 갖는 candidate의 label이 올바른 correspondence를 이룰 확률이 높지만, maximum similarity의 값 자체가 작다면 no-match label의 확률값도 높아진다. 따라서 어떤 feature point A<sub>i</sub>에 대한 no-match label의 초기 확률은 P<sub>i</sub><sup>(0)</sup>(I\*) = 1 - MAX(S(l))로 추정할 수 있으며 matchable하다는 조건하에 추정해본 각 label의 초기 확률은 아래의 식과 같다.

$$P_i(l/matchable) = \frac{S(l)}{\sum_{k=1}^n S(l^k)}$$

따라서 Bayes' rule에 의해, I\*을 제외한 모든 label에 대한 초기 확률을 다음과 같이 추정 할 수 있다.

$$P_i^{(0)}(l) = P_i(l/matchable) \cdot (1 - P_i^{(0)}(I*))$$

$$= \frac{S(l)}{\sum_{k=1}^n S(l^k)} \cdot MAX(S(l))$$

이 초기 확률은 이웃하고 있는 candidate point들의 확률값과 상호 의존하고 있으며, consistency property를 이용하여 updating할 수 있다.

5. Disparity Labeling을 위한 Relaxation Process

relaxation labeling process는 local ambiguity를 감소시키기 위해 contextual information을 이용하는 병렬적, 반복적인 technique이다.

relaxation technique은 몇 가지 방법이 있으며 computer vision의 여러 분야에 적용되어 왔다(SOCCO) 본 논문에서는 S.W. Zucker의 relaxation technique를 도입하여, scene의

disparity labeling을 확률론적으로 접근하며 행하였다.

각 label에 대한 확률은 다음과 같은 함수에 의해 updating된다.

$$P_i^{(t+1)} = F(P_i^{(t)}; Q_i^{(t)}(1))$$

여기서, t는 반복한 숫자이고  $Q_i^{(t)}(1)$ 는 neighborhood contribution이다.

neighborhood contribution은 correspondence를 행하고자 하는 feature point  $A_i$ 의 각 label에 대해 이웃에 존재하는 feature point  $A_j$ 들이 consistency 조건을 만족하면서 상호 지지해 줄 수 있는 정도를 나타내는 값의 합이다.

이때  $A_i$ 의 이웃 점들의 집합  $N(i)$ 는

$$N(i) = \{j \mid \text{MAX}(|x_i - x_j|, |y_i - y_j|) \leq R, i \neq j\}$$

를 만족하는 한정된 영역 내의 점들이며, compatibility  $r_{ij}$ 은  $A_i$ 의 disparity label l과  $A_j$ 의 disparity label l'의 disparity 차가 허용치  $\Theta$  이하이면 compatible하여 1의 값을 갖고, 허용치를 넘으면 independent하여 0의 값을 갖는다.

$$\text{즉, if } |l-l'| \leq \Theta, \quad r_{ij}(l, l') = 1 \\ \text{else} \quad r_{ij}(l, l') = 0$$

neighborhood contribution;

$$Q_i^{(t)}(1) = \sum_{j \in N(i)} C_{ij} \sum_{l' \in L} r_{ij}(l, l') P_j^{(t)}(l'), \quad 1 \in L^*$$

이며,  $C_{ij} = (1 + c \cdot d(A_i, A_j))^{-1}$ 는 neighborhood contribution 계산시에 각 neighborhood와의 거리에 반비례하도록 가해진 weight이다. 따라서 neighborhood contribution  $Q_i^{(t)}(1)$ 은  $A_i$  주위의 feature point가 l과 유사한 disparity label을 가지고 있지 않을때 0에 가깝고, l과 유사한 disparity label을 가지면서 그에 대한 확률이 높다면 큰 값을 가진다.

이 값은 각 label에 대한 확률을 updating하게 된다. 이때 updated probability는

$$P_i^{(t+1)}(1) = P_i^{(t)}(1) \cdot [k_1 + k_2 \cdot Q_i^{(t)}(1)], \quad 1 \in L^* \\ P_i^{(t+1)}(l^*) = P_i^{(t)}(l^*)$$

로 나타낼 수 있으며, 여기서의 상수  $k_1, k_2$ 는 disparity labeling process의 수렴 속도를 결정하는 역할을 한다.

그런데  $Q_i^{(t)}(1) \geq 0$  이므로  $\sum_{l \in L} P_i^{(t+1)}(l) \neq 1$  이다. 그러므로 t+1번 반복한 후 어떤 feature point  $A_i$ 의 각 disparity label에 대한 normalized probability는 다음과 같다.

$$P_i^{(t+1)}(1) = \frac{P_i^{(t+1)}(1)}{\sum_{l \in L} P_i^{(t+1)}(l)}$$

매 반복 수행시  $P_i^{(t+1)}(1)$ 의 값을 탐색하여 0.8 이상의 값을 가지면 그 disparity label이 correspondence된 것으로 간주하며, 0.8 이하의 값을 가지면 candidate label set에서 제외시킨다. 이와 같은 방법으로 각 점에 대해 correspondence되는 disparity label이 결정되면 각 점의 쌍에 대한 화상의 좌표를 알 수 있으므로 triangulation을 통해 3차원 정보를 추출할 수 있다.

### 6. Triangulation

camera geometry는 그림(3)과 같이 두 camera의 lense축이 평행하면서 base line에 수직을 이루도록 설정하였다.

따라서 왼쪽화상의 점 L(x,y)와 오른쪽 화상의 점 R(x',y')가 correspondence를 이룰 때, 그 점에 대한 real world상의 좌표 (X,Y,Z)는 삼각 비례법에 의해 유도된 다음과 같은 수식

으로부터 계산할 수 있다.

$$X = \frac{b \cdot (x + x') / 2}{x - x'} \quad Y = \frac{b \cdot (y + y') / 2}{x - x'} \quad Z = \frac{b \cdot f}{x - x'}$$

여기서 좌표계의 원점 0은 두 camera의 base line의 중점이며 b는 base line의 길이이고, f는 lense의 중심과 image사이의 거리를 나타낸다. 따라서 depth Z는 disparity (x-x')에 반 비례 하고 base line의 길이에 비례함을 알 수 있다.

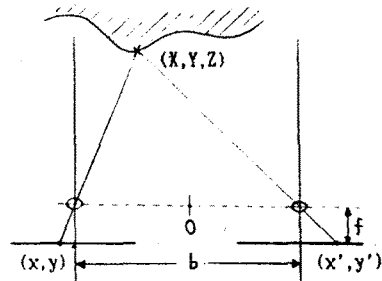


그림3 camera의 geometry와 triangulation

이때 lense의 중심과 화상 사이의 거리는, 실제의 길이(L)를 알고 있는 '+' 모양의 단순한 object를 scene상에 두고 이를 촬영했을 때 투영된 길이(l)와 focal length(f)로부터 구한다.

$$f = f_0 \cdot (L + 1) / L$$

그런데, 실질적으로  $L \gg 1$  이므로 f는 focal length와 거의 같다.

### 7. 실험 및 고찰

3-D 정보를 추출하고자 하는 물체가 약 1.5-2m의 거리에 위치하고 있는 scene을 base line의 길이가 15cm인 두 대의 CCD camera(f=16mm)로 촬영하여, 240X256 image를 취한 후 8 bit로 양자화 하여 image data로 사용하였다.

또한 본 실험은 IBM-PC-AT상에서 C language로 구현하였으며, 그 개괄적인 처리 과정은 그림(4)에 flow-chart로 나타내었다.

relaxation technique을 이용하여 5회의 반복시 98% 이상의 feature point가 특정한 disparity label에 대응되었고, 그중 similarity가 높고 또한 consistency 조건을 만족하는 disparity label이 없을 경우나 시각자에 의해 occluding이 일어났을 경우에는 no-match label l\*에 대응 되었다.

그리고 추출한 3 차원 정보와 real world의 좌표값은 scene과 object의 유형에 따라 1-2% 내의 error bound를 가졌다.

### III. 결론

본 연구는 computer vision system에서 보다 정확한 correspondence를 이루고자 하는데 중점을 두었으며, 이를 위해 sombrero filter 처리를 하여 edge에 해당하는 zero-crossing을 추출하고 이에 대한 similarity를 두 가지의 적절한 property로 측정한 후, 초기 확률로부터 contextual information을 고려하면서 반복 수행하는 relaxation technique을 도입하였다.

따라서 global information의 고려가 없이 행하는 correlation 방법에서 일어날 수 있는 mismatch의 위험은 줄일 수 있었을 뿐 아니라 보다 정확한 correspondence를 행할 수 있었다.

그러나 preprocessing시의 convolution과 relaxation 수행시의 반복 때문에 많은 시간이 소요되므로 실시간 처리를 위해서는 적합하지 않지만 정교한 3-D 정보를 요하는 많은 분야에 이용 가능하다.

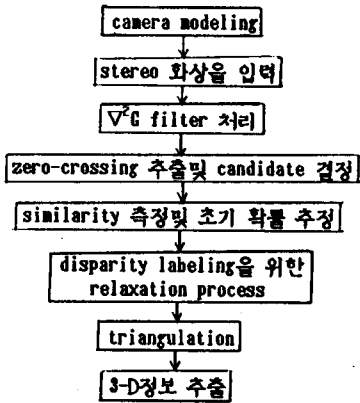


그림4 flow-chart

참고 문헌

- [1] R. A. Jarvis, "A Perspective on Range Finding Techniques for Computer Vision", IEEE Trans. PAMI, 1983
- [2] P. K. Allen "Sensing and Describing 3-D Structure" Proc. of Robotics & Automation, 1986
- [3] G. Medioni & R. Nevatia, "Segment-Based Stereo Matching", CVGIP, 1985
- [4] W. E. L. Grimson, "Computational Experiments with a Feature Based Stereo Algorithm", IEEE Trans. PAMI, 1985
- [5] S. W. Zucker & E. V. Krishnamurthy & R. Harn "Relaxation Process for Scene Labeling : Conversions, Speed and Stability", IEEE Trans. SMC, 1976
- [6] S. T. Barnard & W. B. Thompson, "Disparity Analysis of Images", IEEE Trans. PAMI, 1980
- [7] Y. C. Kim & J. K. Aggarwal, "Finding Range from Stereo Images", Proc. of CVPR, 1985
- [8] Y. Yakimovsky, "A System for Extracting 3-D Measurements from a Stereo Pair TV Cameras", CGIP, 1978
- [9] A. Rosenfeld, "Relaxation: Pixel-Based Methods", Fundamentals in Computer Vision, pp.373-383, Cambridge Univ. press, 1983
- [10] B. K. Paul Horn, "Robot Vision" pp.299-326, MIT Press, 1986

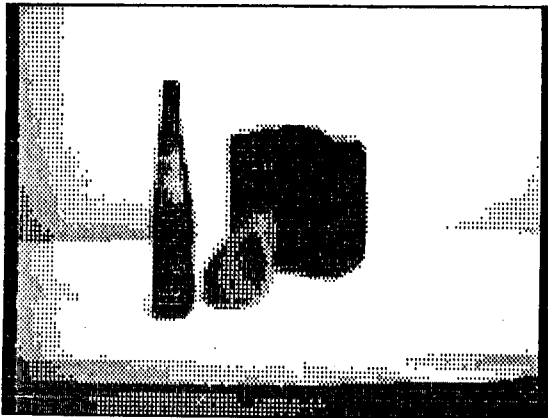


그림5 stereo 화상의 왼쪽

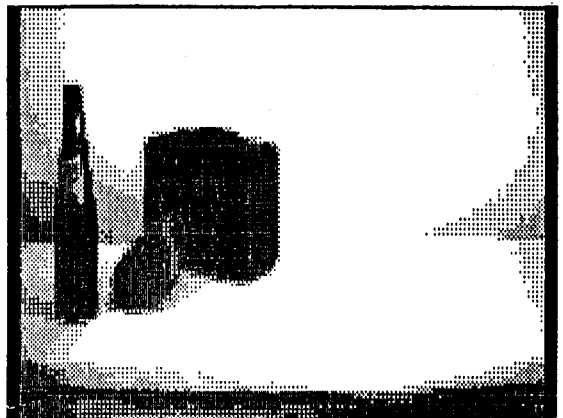


그림6 stereo 화상의 오른쪽

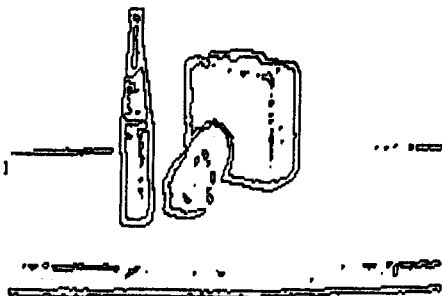


그림7 왼쪽 화상의 zero-crossing (sombrero filter의 폭  $W=2\sqrt{2}$ )

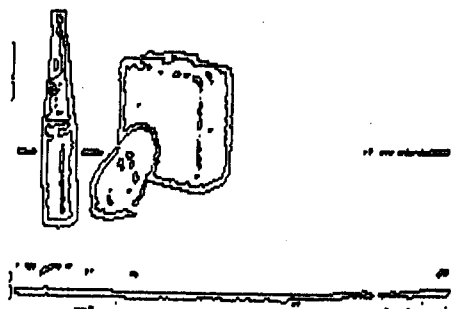


그림8 오른쪽 화상의 zero-crossing