

스테레오 영상을 이용한 3차원 표면좌표 추출 알고리즘

원성혁*, 김민기*, 김병우*, 이기식*, 김현배*

*고려대학교 전자 및 정보공학과

전화 : 0415-860-1352 / 팩스 : 0415-862-1527

Extraction of the three-dimensional surface coordinate from a stereo image

Sung Hyuk Won, Min Gi Kim, Byung Woo Kim, Gee Sik Lee and Hyun Bae Kim

Department of Electric and Information Engineering, Korea University

E-mail : techwon@kebi.com, mgkim@hard.korea.ac.kr, supermankim@hotmail.com,

gslee@hard.korea.ac.kr, hbkim@hard.korea.ac.kr

Abstract

In the closed range space, the parallel two CCD cameras are used to acquire a pair of stereo image. The acquired stereo image are computed with Wavelet Transform repeatedly and including the low frequency component, the image size of those are reduced. It is the pyramid structure. The optimum matching point is searched to the pixel. Then applying the optimum matching point to DLT, it extract the three - dimensional surface coordinate from a stereo image.

The direct linear transformation(DLT) method is used to calibrate the stereo camera compute the coordinate on a three dimensional space. To find the parameters for the DLT method, 30 control points which marked on the cylinder type object are used. To improve the matching algorithm, the paper select the pyramid structure for Wavelet Transform . The acquired disparity information is used to represent the really three-dimensional surface coordinate.

3차원적인 위치 계산이 가능하지만 정확한 영상의 3차원 정보 획득은 어려운 실정이다. 이를 보완하기 위해 카메라 보정(Camera Calibration)이란 방법을 사용하는데 이는 실제 3차원 정보를 구할 수 있을 뿐만 아니라 카메라의 위치, 렌즈의 굴곡, 초점, 외부변수 등에 따라 생기는 에러들을 보정해 줄 수 있다. 그래서 더 정확한 정보를 구하기 위해 필요한 작업이기도 하다. 이 방법은 이런 에러 요소를 보정해 주면서 3차원 정보를 구할 수 있기 때문이다.

기본 아이디어는 공간상의 어떤 한 점에서 카메라 CCD 소자에 맺혀지는 직선은 영상평면의 좌표를 포함하면서 일직선상에 존재하고 이 직선은 공간좌표와 영상좌표를 변화해주는 matrix로 표현할 수 있으며 이것을 스테레오로 확장하면 공간상의 좌표를 구할 수 있다.

I. 서론

스테레오 카메라 시스템에서 극상선(Epipolar line)과 + disparity를 만족하는 한 쌍의 영상을 얻고 카메라 변수들이 주어진다면, 특징이 뚜렷한 물체에서는 대략적인

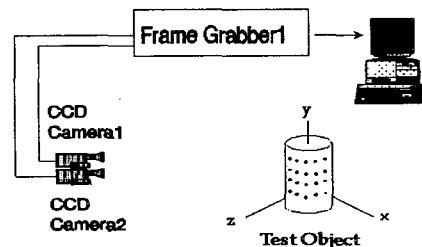


그림 1. 시스템 구성도

Fig. 1 System configuration

II. 본 론

1. 스테레오 정합(Stereo matching)방법

1.1 NCC Algorithm (Normalized Cross-Correlation)

우영상의 한 화소를 중심으로 window를 씌워 그 주변 값을 감안하여 대응되는 좌영상의 화소들 간의 상호 상관도 값은 극상선을 따라 위치하며, 각 화소에서의 정합 가능성은 식(1)과 같이 정합 창틀 내에 위치하는 평균과 분산으로 정규화한 NCC (Normalized Cross-Correlation)함수로서 구해 낸다.

$$P_{k,i,d} = \frac{\sum (g_L(k, i+d) - \mu_L)(g_R(k, i) - \mu_R)}{\sqrt{\sum (g_L(k, i+d) - \mu_L)^2 \sum (g_R(k, i) - \mu_R)^2}} \quad \text{식(1)}$$

(μ_L 과 μ_R 은 좌·우영상의 상관관계 창틀 내 화소들의 평균값)

1.2 MAD(Mean of absolute difference)

한 화소를 중심으로 window를 씌워 그 주변 값을 감안하여 수평방향으로 window를 이동해 가며 MAD(Mean of absolute difference)를 구하여 최소가 되는 값의 위치를 disparity로 판단한다.

$$MAD = \frac{1}{Block\ Size} \sum_{(k,i) \in B} |g_R(k, i) - g_L(k, i+d)| \quad \text{식(2)}$$

1.3 Dynamic Programming

1.3.1 극상선(Epipolar line) 상의 정합

$a_0, a_1, a_2, \dots, a_m$ 과 $b_0, b_1, b_2, \dots, b_n$ 을 왼쪽과 오른쪽 영상 화소들의 열이라 하고 이를 중심으로 각각의 화소주위에 값의 차이, 식(2)를 통해서 구한다. 이를 $MAD(g_L(i), g_R(j))$ 라 하자.

만약 한 화소가 다른 이미지에 정합되는 화소가 없다면 임의의 값 null에 해당된다. 이를 Occlusion되었다고 한다. Occlusion이 된 pixel에 대해선 $\frac{M}{2}$ 의 값을 할당한다. 여기서 M은 pixel들이 matching이 되는 최대 거리이다. $g_L(i)$ 과 $g_R(j)$ 두 pixel 사이에 거리가 M보다 크다면 이들은 정합된 것으로 간주하지 않고 null값을 할당한다.

1.3.2 일대일 정합

일대일 정합은 식(3)과 같다.

$$C(i, j) = \min \left\{ C(i-1, j-1) + MAD(g_L(i), g_R(j)), \right. \\ \left. C(i-1, j) + \frac{M}{2}, \right. \\ \left. C(i, j-1) + \frac{M}{2} \right\} \quad \text{식(3)}$$

일대일 정합은 Occlusion과 정합이 되는 pixel을 구분하면서 이들 값 중에서 최소가 되는 값을 계속 찾아가면서 정합을 한다.

1.3.3 일대다, 다대일의 정합

한 이미지에 투영된 영역은 다른 이미지에서 투영된 영역보다 더 좁거나 더 넓을 수 있다. 기울어진 표면은 한 이미지 안의 극상선의 m개의 화소들의 부분과 다른 이미지 안의 극상선의 n개의 화소 부분으로 투영될 수 있다. 이 경우에 좁은 영역이 넓은 영역에 정합될 수 있고 그 반대로 넓은 영역이 좁은 영역에 정합될 수 있다. 이러한 일대다(one to many) 또는 다대일(many to one)의 정합을 위해 식(4) 같은 식을 정의한다.

$$C(i, j) = \min \left\{ C(i-1, j-1) + MAD(g_L(i), g_R(j)), \right. \\ \left. C(i-1, j) + MAD(g_L(i), g_R(j)), \right. \\ \left. C(i, j-1) + MAD(g_L(i), g_R(j)), \right. \\ \left. C(i-1, j) + \frac{M}{2}, \right. \\ \left. C(i, j-1) + \frac{M}{2} \right\} \quad \text{식(4)}$$

2. Wavelet transform의 계층적 특성을 이용한 정합

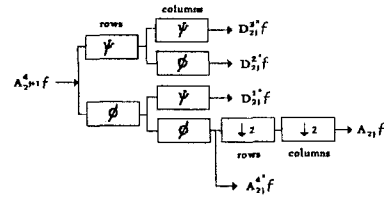


그림 2. Stereo matching에서 제안된 각 채널들의 구조
Fig. 2 The construction of the channels for the proposed stereo matching

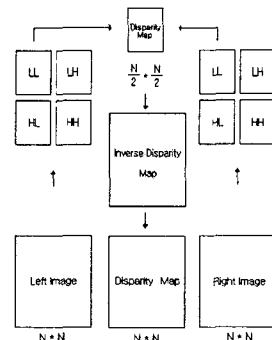


그림 3. Stereo matching을 위한 제안된 계층 구조
Fig. 3 The proposed pyramid structure

Wavelet transform을 한 영상은 각각의 채널 Blurred image(LL), Horizontal image(LH), Vertical image(HL), Diagonal image(HH)로 나타나게 되나 여기서는 LL 성분만을 가지고 정합을 하여 첫 번째 disparity map을 구성하고 inverse transform하여 원래 영상크기의 두 번째 disparity map을 구하게 된다. 두 번째 disparity map을 근거로 다시 주위의 값을 정합하여 더욱 정확한 disparity map을 구성하게 된다.

$\frac{N}{2} * \frac{N}{2}$ 크기의 disparity map을 구성하기 위해 DP(Dynamic programming)을 이용하였고 정확한 disparity map을 구성하기 위해 두 번째 map을 근거로 NCC 방법을 적용하였다.

3. DLT(Direct Linear Transformation) 이론

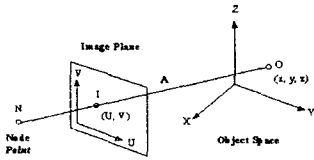


그림 4. 물체 좌표, 영상 좌표, 교점
Fig. 4. Object coordinate, Image coordinate and Node point

그림 4.은 실제 물체(O)와 투영된 영상(I)과 교점(Node Point : x_0, y_0, z_0)의 관계를 나타낸 그림이다. 여기서 (x, y, z)는 물체공간 좌표(object-space coordinate)이며 (U,V)는 영상평면 좌표(image-plane coordinate)이다. N, I, O의 점이 한 직선 상에 위치하는 데 이것을 동일선상 조건 (collinearity condition) 이라고 한다. 벡터A와 B가 동일한 공간에 존재한다고 가정하면 $\vec{B} = c\vec{A}_1$ 란 식을 유출해낼 수 있고 c는 상수이다. 그러나 주의해야할 것은 벡터A는 물체공간참조구조(object-space reference frame)인 반면에 벡터B는 영상평면참조구조(image-plane reference frame)이기 때문에 변화해줄 필요가 있다.

식1을 벡터 A와 행렬 곱하면

$$\vec{A}_1 = R\vec{A} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \vec{A} \quad \text{식(5)}$$

수식을 구할 수 있다.

$$\vec{B} = c \cdot \vec{A}_1 = c \cdot (R \cdot \vec{A}) \Rightarrow \begin{pmatrix} U - U_0 \\ V - V_0 \\ -d \end{pmatrix} = c \cdot \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix} \quad \text{식(6)}$$

DLT 변수들로 다시 표현하면 다음과 같다.

$$\begin{aligned} u + \Delta u &= \frac{L_1 \cdot x + L_2 \cdot y + L_3 \cdot z + L_4}{L_9 \cdot x + L_{10} \cdot y + L_{11} \cdot z + 1} \\ v + \Delta v &= \frac{L_5 \cdot x + L_6 \cdot y + L_7 \cdot z + L_8}{L_9 \cdot x + L_{10} \cdot y + L_{11} \cdot z + 1} \end{aligned} \quad \text{식(7)}$$

여기서 ($\Delta u, \Delta v$)는 광학왜곡으로 생기는 에러이다.

$$\begin{bmatrix} -x_i - y_i - z_i - 1 & 0 & 0 & 0 & 0 & 0 & 0 & u_i x_i & u_i y_i & u_i z_i \\ 0 & 0 & 0 & 0 & -x_i - y_i - z_i - 1 & v_i x_i & v_i y_i & v_i z_i \\ \alpha r_1^2 A_i & \alpha r_1^4 A_i & \alpha r_1^6 A_i & \alpha \beta A_i & (r_1^2 + 2\alpha^2) A_i \\ \beta r_1^2 A_i & \beta r_1^4 A_i & \beta r_1^6 A_i & \alpha \beta A_i & (r_1^2 + 2\beta^2) A_i \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \\ L_8 \\ L_9 \\ L_{10} \\ L_{11} \\ L_{12} \end{bmatrix} + \begin{bmatrix} u_i \\ v_i \end{bmatrix} = 0 \quad \text{식(8)}$$

$$\begin{aligned} \alpha &= u_i - u_0, & \beta &= v_i - v_0 \\ r_1^2 &= \alpha^2 + \beta^2 \\ A_i &= L_9 x_i + L_{10} y_i + L_{11} z_i + 1 \end{aligned}$$

$$\begin{bmatrix} B_{x_{11}} & B_{x_{21}} & \dots & B_{x_{161}} \\ B_{y_{11}} & B_{y_{21}} & \dots & B_{y_{161}} \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ L_3 \\ \vdots \\ L_{11} \\ \vdots \\ L_{16} \end{bmatrix} + \begin{bmatrix} D_{x_i} \\ D_{y_i} \end{bmatrix} = 0 \quad \text{식(9)}$$

식(9)는 $BL + D = 0$ 이므로 해를 구하면,

$$\begin{aligned} L &= -(B^T W B)^{-1} B^T W D \\ &= -N^{-1} D^* \end{aligned} \quad \text{식(10)}$$

이 된다. 기본적으로 11개의 parameter가 사용되지만 좀 더 정확한 camera calibration을 위해서는 5개가 더 사용된다.

$$\begin{bmatrix} (\hat{u}L_9 - L_1) & (\hat{u}L_{10} - L_2) & (\hat{u}L_{11} - L_3) \\ (\hat{v}L_9 - L_5) & (\hat{v}L_{10} - L_6) & (\hat{v}L_{11} - L_7) \end{bmatrix}_i \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + \begin{bmatrix} \hat{u} - L_4 \\ \hat{v} - L_8 \end{bmatrix}_i = 0 \quad \text{식(11)}$$

(여기서 i는 camera 수, j는 point 수이다.)

이 해를 풀면,

$$\begin{aligned} B\Delta_j + C &= 0 \\ \Delta_j &= -(B^T W B)^{-1} (B^T W C) \end{aligned} \quad \text{식(12)}$$

이 된다. 이 해를 구함으로써 모르는 3차원 공간의 좌표를 얻을 수 있다.

4. 실험 결과

4.1 DLT 변수 결정과 제안된 스테레오 정합 알고리즘 결과

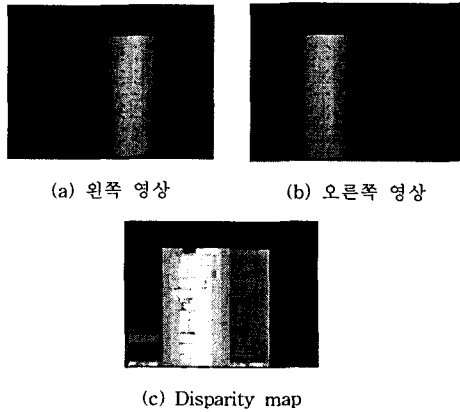


그림 5. DLT 변수들을 구하기 위한 스테레오 영상과 정합 결과

Fig. 5 The stereo image to acquire DLT parameters and the matching result

4.2 DLT를 이용한 3차원 표면 재구성

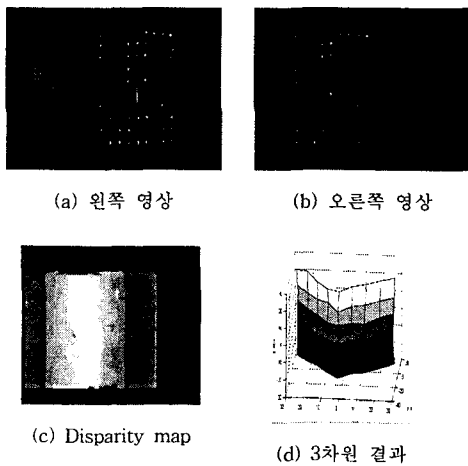


그림 6. 3차원 실측 좌표를 구하기 위한 실제 스테레오 영상과 결과

Fig. 6 The real stereo image to acquire the 3D measurement coordinate and the result

III. 결 론

본 논문에서는 카메라에서 71cm 떨어진 물체를 찍은

영상(340*240)을 Wavelet transform 하여 크기를 $\frac{1}{4}$ 로 만들어서 정합에 이용하였는데 부드럽게 변하는 곡면에서는 아직도 많은 오차가 있고, 렌즈의 왜곡과 테스트 물체의 초기 좌표정보 오차가 있어 3차원 실측치로 표현하였을 때는 아직 많은 오차가 보였다.

그래서 다음 실험에서는 좀더 정확한 3차원 정보 추출을 위하여 640*480 크기의 영상을 가지고, 스테레오 정합부분에서는 현재에는 사용하지 않은 영상의 high frequency 부분을 적용하려고 하며, 3차원구성 부분에서는 DLT parameters 16개를 모두 사용하여 렌즈의 왜곡을 보정하여 더욱 정확한 3차원 실측치를 추출할 수 있는 시스템을 구성하고자 한다. 그러면 산업분야 뿐만 아니라 의학분야에서도 손쉽게 3차원 정보를 추출, 사용할 수 있는 알고리즘이 될 것으로 기대된다.

[참고 문헌]

[1] 안혜정, "Stereo Image Matching을 이용한 Skin Furrows에 대한 연구", 정보공학과, 고려대학교, 1997년 2월

[2] W.G.Lee, Y.S.Kim, K.H.Do, and Y.H.Ha, "정합일치성을 이용한 반복 이완 정합 알고리즘" 전자공학회 논문지 제 32권 B편 제 1호, 1995년

[3] Yuichi Ohta and Takeo Kanade, "Stereo by Intra-and Inter Scanline Search Using Dynamic Programming", IEEE Tran. on PAMI, vol.7 No.2 March, pp.139 ~ 154, 1985

[4] 강진만, 문종섭, 김민기, "역 가중치를 이용한 스테레오 매칭 알고리즘", 고려대학교 과학기술 논집 제 16권, pp.37 ~ 53, 1998년

[5] Y.S. Kim, J.J. Lee and Y.H. Ha, "Stereo Matching Algorithm based on Modified Wavelet Decomposition Process", Pattern Recognition, vol.30, No.6, pp.929 ~ 952, 1997

[6] Abdel-Aziz, Y.I., & Karara, H.M.(1971). "Direct linear Transformation from comparator coordinates into object space coordinates in close-range photogrammetry", Proceeding of the Symposium on Close-Range Photogrammetry(pp. 1-18). Falls church, VA: American Society of Phtogrammetry.

[7] Kwon, Y.-H.(1989). "The effects of different control point conditions on the DLT calibration accuracy", Unpublished class projected report, Pennsylvania State University.