

스테레오 연속 영상을 이용한 구조 복원의 정제

A New Refinement Method for Structure from Stereo Motion

박성기, 권인소
(Sung-Kee Park and In So Kweon)

Abstract : For robot navigation and visual reconstruction, structure from motion (SFM) is an active issue in computer vision community and its properties are also becoming well understood. In this paper, when using stereo image sequence and direct method as a tool for SFM, we present a new method for overcoming bas-relief ambiguity. We first show that the direct methods, based on optical flow constraint equation, are also intrinsically exposed to such ambiguity although they introduce robust methods. Therefore, regarding the motion and depth estimation by the robust and direct method as approximated ones, we suggest a method that refines both stereo displacement and motion displacement with sub-pixel accuracy, which is the central process for improving its ambiguity. Experiments with real image sequences have been executed and we show that the proposed algorithm has improved the estimation accuracy.

Keywords : computer vision, structure from motion, bas-relief ambiguity, small motion, robust and direct method.

I. 서론

로봇의 자율 주행과 visual reconstruction과 관련하여 카메라의 운동에 의한 구조복원의 문제는 컴퓨터 비전에 있어서 현재 활발히 연구되어지는 분야이며, 그에 대한 기하학적인 특징들은 잘 규명되어져 왔다. 그 중에서 미세 운동(small motion)을 전체로 하는 유동시(optical flow)와 직접 방법[1, 2, 7]에 있어서, 카메라의 운동에 의한 구조 복원의 문제는 회전 운동과 병진 운동을 잘 구별할 수 없는 운동 변수 추정의 모호성이 존재하는데 이를 'bas-relief 모호성(ambiguity)'이라고 한다. 본 연구에서는 보정된 스테레오 카메라를 통하여 연속인 스테레오 영상이 주어질 때, 강한 직접 방법(robust and direct method)에 기초하여 이러한 운동 모호성을 개선하는 방법을 제시하고자 한다.

Bas-relief 모호성과 관련한 기존의 연구로서는 Weng [2]이 미세 운동이 갖는 한계점을 분석하였으며, 또한 추정 방법과 관련하여 비선형 최적화 해법을 제시하였다. Szeliski와 Kang [3]은 information matrix의 Hessian의 특이점을 통하여 bas-relief 모호성에 대한 분석을 시도하였다. 최근에 Soatto와 Brockett [4]는 대응점의 추정 오차가 큰 경우에도 강인하게 수립하는 SFM 알고리즘을 제시하였으며, bas-relief 모호성은 카메라의 운동에 의한 구조 복원의 문제에 있어서 고유한 특성이라는 점을 수학적으로 증명하였다. 또한, 이러한 운동의 모호성을 극복하는 방법으로서 planar-parallax를 이용하는 방법 [5], 선형 변환을 이용하여 부공간(subspace)에서 운동 매개 변수를 나누어서 추정하는 방법 [6] 등이 제시되어지기도 했다.

그러나, 카메라의 운동에 의한 구조복원의 문제는 다른 컴퓨터 비전의 문제와 마찬가지로 대응점을 얼마나 정확히 찾을 수 있는지가 가장 핵심적인 관건이다. 이는 기존의 연구[1, 4, 7]에서 규명되었던 바와 같이 유동시와 코너 점을 찾는 문제는 그 오차의 정도가 클 뿐 아니라 그 오차의 분포도 정규분포를 이루지 아니하는 것으로 알려져 있다. 따라서 기존의 선형적인 방법에

기초한 해법으로서는 정확한 해를 기대할 수 없을 뿐 아니라 그러한 운동의 모호성을 해결할 수 없다고 하겠다. 그렇다고 비선형 최적화 기법을 적용하고 강한 통계학의 기법들을 이용한다 하더라도 그것이 bas-relief 모호성을 해결하는 충분 조건이 될 수는 없다. 왜냐하면, bas-relief 모호성을 이루는 경우의 운동 매개변수를 추정하기 위한 평가함수의 형태가 global minimum을 구별하기 어려운 long valley 형상 [7]을 이룬다는 점 때문인데, 이는 어떤 수치해법을 적용하느냐의 문제로서 해결될 사항이 아니라 영상의 대응점을 픽셀 이하의 단위에서 얼마나 정확히 찾아 주느냐가 더 중요한 점이라고 하겠다. 그러나 앞서 언급한 것처럼 운동 대응점을 이미지 밝기값에만 의존하여 찾는 방법으로는 그 정확성을 보증할 수 없을 뿐 아니라 그 오차의 확률적 모델도 단순하게 정규분포를 이룬다고 할 수 없다. 이러한 상황에서 기존의 모든 단선적 연구 방법, 즉 대응점을 먼저 이미지 밝기값에만 의존하여 찾고 그러한 대응점에 수치해법을 적용하는 알고리즘은 기본적으로 bas-relief 모호성 문제를 해결할 수 없다고 하겠다.

따라서, 본 연구에서는 이러한 bas-relief 모호성을 해결하기 위하여 대응점과 카메라 운동 매개변수를 상호 재귀적으로 추정하는 방법을 제시하고자 한다. 기본적으로 기존의 직접 강인 추정법에 의하여 얻은 운동 매개변수와 깊이 정보를 초기 추정값으로 활용하여 다시 이를 바탕으로 스테레오 매칭(stereo matching)과 운동 대응점(motion correspondence)을 픽셀 단위 이하로 계산하여 이를 바탕으로 운동 매개변수와 3차원 대상 깊이를 정제하는 방법이 본 연구의 기본 구조이다.

제 2장에서는 구체적으로 정제 방법에 대한 상세한 내용과 알고리즘을 다루고자 하며, 제 3장에서는 실제 실험을 통하여 본 연구의 타당성을 보이고자 하며, 그리고 제 4장에서는 본 연구에 대한 결론을 제시하고자 한다.

II. 새로운 정제 방법 (A new refinement method)

그림 1은 본 연구와 관련한 스테레오 연속 영상을 나타내며 전체적인 표기법을 나타내고 있다. 그림에서 image-1과 image-2

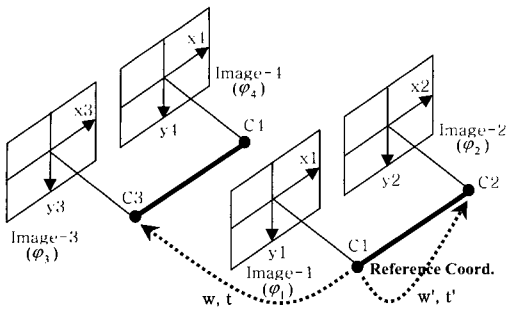


그림 1. 연속적인 스테레오 카메라 운동.
Fig. 1. Consecutive motion of stereo camera.

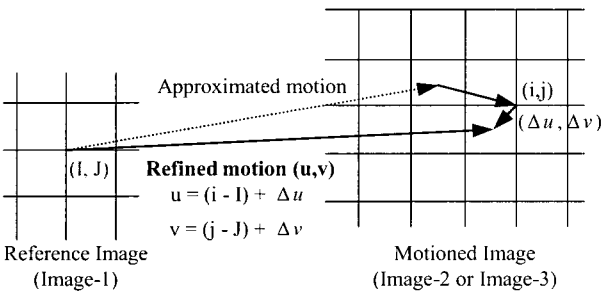


그림 2. 정제 과정의 도식적 표현.
Fig. 2. Graphical representation about the refinement process.

는 스테레오 영상의 쌍을 나타내며, 이때 카메라의 병진 및 회전 운동 \mathbf{t} , \mathbf{w} 에 의해 새로운 스테레오 영상 image-3과 image-4를 얻을 수 있다. 스테레오 카메라가 보정된 경우에는 그림에서 \mathbf{t}' 과 \mathbf{w}' 은 알 수 있으므로, 스테레오 카메라의 운동 추정은 병진운동 \mathbf{t} 와 회전운동 매개변수 \mathbf{w} 를 추정하는 문제로 볼 수 있으며, 이를 해결하는 강인 및 직접 추정법 [1]에서는 그 추정이 (1)과 같이 주어진다.

$$\min_{\mathbf{a}} \sum_i \rho(r_i, \sigma), \quad r_i = \mathbf{C}_i \mathbf{a} + d_i \quad (1)$$

$$\mathbf{a} = [t_x \ t_y \ t_z \ w_x \ w_y \ w_z]^T,$$

$$\mathbf{C}_i = [(\nabla I_i^T \mathbf{B}_i \mathbf{w}' + I_{ii}') \nabla I_i^T \mathbf{A}_i \quad -(\nabla I_i^T \mathbf{A}_i \mathbf{t}') \nabla I_i^T \mathbf{B}_i],$$

$$d_i = -(\nabla I_i^T \mathbf{A}_i \mathbf{t}') I_{ii}$$

여기서, i 는 기준 이미지(image-1)에서의 각 픽셀을 나타내며 \mathbf{a} 는 추정 하고자 하는 운동 매개 변수, 그리고 \mathbf{C}_i, d_i 는 기준 영상의 밝기값의 평면 구배 및 시간 구배와 좌표값 등이 결합되어진 벡터 및 스칼라 양이다 [1]. 그러나, 모든 이미지 픽셀의 밝기 값 분포는 유동시 제한조건식 (optical flow constraint equation) [13, pp.282]을 만족할 수 없기 때문에 운동 매개변수의 추정을 모든 픽셀에 대하여 수행하는 것으로 마무리하는 것에는 문제점이 있다. 그래서, 본 연구에서는 (1)에 의한 추정 과정을 초기 추정으로 간주하고자 하며, 이를 바탕으로 다시 더 정확한 대응점들을 계산한 후에 이를 바탕으로 운동 매개변수를 다시 정제 추정하고자 한다. 그림 2는 정확한 대응점을 다시 계산하는 과정을 도식적으로 보여주고 있다. 그림에서 볼 수 있는 바와 같이

근사 운동 정보는 (1)에 의해서 주어지며, 이를 바탕으로 먼저 픽셀 단위의 대응점을 찾고 다시 그 위치를 기준으로 픽셀 이하 단위로 대응점을 찾게 된다. 이 과정은 image-1과 image-2사이의 스테레오 매칭에 대하여 수행되어지고, 다시 image-1과 image-3사이의 운동 대응점에 대하여도 동일하게 수행되어진다. 이때, 스테레오 매칭에 의하여 3차원 깊이 z 가 정제되어지면, 이를 바탕으로 운동 대응점은 다음의 (2)에 의하여 운동 매개변수를 다시 정제하는데 이용되어 진다.

$$u = \frac{dx}{dt} = \frac{1}{z}(t_x - xt_z) - w_z y + w_y(1 + x^2) - w_x xy,$$

$$v = \frac{dy}{dt} = \frac{1}{z}(t_y - yt_z) + w_z x - w_x(1 + y^2) + w_y xy \quad (2)$$

위에서 u 와 v 는 대상 물체의 영상면상의 운동 속도를 나타내며, t_x, t_y, t_z 와 w_x, w_y, w_z 는 각각 카메라의 병진 및 회전 속도, x 와 y 는 정규 영상좌표, 그리고 z 는 카메라 좌표계를 기준으로 하는 대상 물체의 z 좌표값을 의미한다 [8, pp.179].

1. 스테레오 매칭의 정제 방법

근사화된 스테레오 매칭이 기존의 강인 및 직접 추정법에 의하여 주어질 때, 그에 대한 정제된 스테레오 매칭은 다음과 같은 픽셀 단위의 매칭과 픽셀 이하 단위의 매칭으로 얻을 수 있다. 여기서, 두 단계로 나누어서 매칭을 수행하는 이유는 어떠한 방법으로도 픽셀 단위의 매칭이 없이 근사화된 추정위치로부터 바로 픽셀 이하 단위로 추정하는 것이 정확성을 보장하기 힘들며 계산 시간의 관점에서도 불리하기 때문이다.

픽셀 단위의 매칭은 먼저 기준 이미지의 임의의 픽셀 (I,J)에 대한 스테레오 운동 (u,v)가 주어질 때, image-2에서의 임시의 픽셀 단위의 매칭 (i,j)는 다음과 같이 주어질 수 있다.

$$(i, j) = \text{nearest int} (I + u, J + v) \quad (3)$$

그러나, 픽셀점 (i,j)는 정확한 픽셀 단위의 매칭점이 아닐 수 있으므로 그 주변을 기준으로 정확한 픽셀 매칭점을 찾아야 한다. 보정 정보를 미리 알고 있는 보정된 스테레오 이미지에서는 영역 상관관계(area correlation)기법[9]이 유용하다고 알려져 있으므로, 본 연구에서는 SAD (sum of absolute difference)를 유사도 측도로 사용한다. 따라서, 정확한 픽셀 단위의 스테레오 매칭은 임시의 픽셀 단위의 매칭점- (3) -을 중심으로 SAD를 계산하고 그 값이 (4)를 만족하면 그 픽셀점은 정확한 픽셀 단위의 스테레오 매칭점이 된다.

$$SAD(i, j : I, J) = \sum_{(m,n) \in D} |I_{\phi_1}(I, J) - I_{\phi_2}(i - m, j - n)| \quad (4)$$

여기서, D 는 image-2에서 검색하고자 하는 주변 픽셀 영역을 나타낸다. 만약, image-1의 픽셀 (I,J)가 위의 식을 만족하지 못하면, 그 픽셀은 정제 과정에서 제외되는 것으로 한다. SAD를 이용하여 픽셀 단위의 스테레오 매칭 점을 얻으면 그 측도를 바탕으로 보간법을 수행하여 픽셀 이하 단위의 매칭을 수행할 수도 있으나 이는 이미지 밝기값 자체에 기반한 보간법이 아니므로 오차의 가능성이 존재한다[8, pp.80]. 따라서, 본 연구에서는 다

유의 과 같이 유동시 제한 조건식에 기초하여 픽셀 이하 단위의 스테레오 매칭을 수행한다.

$$E_{D1}(\Delta u) = \sum_{x \in D1} \mathbf{W}^2(x) \mathbf{W}_p^2(x) [I_x(x) \cdot \Delta u + I_t(x)]^2 \quad (5)$$

여기서, \mathbf{W} 는 영역 중심에 가중치를 부여하는 창함수로서 본 연구에서는 가우시안 커널 [13, pp. 158]을 사용하며, \mathbf{W}_p 는 픽셀 밝기값의 분포가 선형특성을 이루는지를 반영하는 창함수인데, 이는 유동시 제한 조건식 자체가 밝기값의 선형적 분포라는 가정을 내포하고 있기 때문이다. 본 연구에서는 이를 고려하여 \mathbf{W}_p 를 다음과 같이 정의하고자 한다.

$$\mathbf{W}_p(i, j) = \frac{1}{c_1 + c_2 \times \text{planeness}(i, j)} \quad (6)$$

$$\text{planeness}(i, j) = \frac{\mathbf{a}(i, j) \times \mathbf{b}(i, j) \cdot \mathbf{c}(i, j)}{|\mathbf{a}(i, j) \times \mathbf{b}(i, j)| |\mathbf{c}(i, j)|}$$

$$\mathbf{a}(i, j) = [1 \quad 0 \quad (I_{\varphi_1}(i+1, j) - I_{\varphi_1}(i, j))]^T,$$

$$\mathbf{b}(i, j) = [0 \quad 1 \quad (I_{\varphi_1}(i, j+1) - I_{\varphi_1}(i, j))]^T,$$

$$\mathbf{c}(i, j) = [1 \quad 1 \quad (I_{\varphi_1}(i+1, j+1) - I_{\varphi_1}(i, j))]^T$$

그러므로, (4)와 (5)를 이용하여 정제된 스테레오 매칭점을 얻을 수 있으며, 이때의 스테레오 이미지 사이의 정제된 운동은 다음과 같이 정의할 수 있다.

$$u = (i - I) + \Delta u \quad (7)$$

이때, 정제된 3차원 깊이 z 는 (7)을 (2)에 대입함으로써 얻을 수 있으며, 보정된 스테레오 영상에서는 $v=0$ 으로 나타난다.

2. 운동 대응점의 정제방법

픽셀 이하 단위의 스테레오 매칭이 얻어진 기준 이미지의 픽셀들에 대하여, 기준 이미지 *image-1*과 운동 이미지 *image-3* 사이에 운동 대응점(motion correspondence)을 다음과 같이 정제할 수 있다. 먼저 스테레오 매칭의 (3)을 이용하여 임시의 대응점 픽셀을 찾을 수 있다. 그러나, 픽셀 단위의 정확한 대응점은 스테레오 매칭에서 사용한 영역 상관관계법을 운동 대응점에서는 그대로 사용할 수가 없는데, 이는 회전 운동 성분을 포함하는 운동 대응점에서는 영역 상관 관계법이 정확한 측도가 될 수 없기 때문이다 [8]. 따라서, 본 연구에서는 이미지 밝기값을 국소 영역에서 이차다항 곡면식으로 근사화시켜 그 곡면식의 계수를 비교하는 방식으로 픽셀 단위의 운동 대응점을 얻고자 하였다. 이는 회전 운동에 의한 곡면식의 변화가 영역 상관 관계법보다는 더 둔감하다는데 바탕을 두고 있다 [11]. 이를 구현하기 위하여 기준 이미지의 임의의 픽셀 (I, J) 에 대하여 그 주변 픽셀을 이용하여 그 밝기값을 다음과 같이 이차다항식으로 나타낼 수 있다.

$$\min_{p(I, J)(m, n) \in D2} \sum \left(\begin{array}{l} p_0 \cdot (I+m)^2 + p_1 \cdot (I+m)(J+n) + p_2 \cdot (J+n)^2 \\ + p_3 \cdot (I+m) + p_4 \cdot (J+n) + p_5 - I_{\varphi_1}(I+m, J+n) \end{array} \right)^2 \quad (8)$$

여기서, $p = (I, J) = [p_0 \quad p_1 \quad p_2 \quad p_3 \quad p_4 \quad p_5]^T$ 는 다항식의 계수 벡터이다. 이와 동일한 방식으로 운동 이미지 *image-3*의 임시 대응점도 그 주변 픽셀들을 이용하여 다음과 같이 근사화가 가능하다.

$$\min_{p(i, j)(m, n) \in D2} \sum \left(\begin{array}{l} p_0 \cdot (i+m)^2 + p_1 \cdot (i+m)(j+n) + p_2 \cdot (j+n)^2 \\ + p_3 \cdot (i+m) + p_4 \cdot (j+n) + p_5 - I_{\varphi_3}(i+m, j+n) \end{array} \right)^2 \quad (9)$$

따라서, 이미지 밝기값의 근사화된 다항식의 계수를 같은 차수별로 비교하는 방식으로 (10)과 같이 유사 측도를 구성할 수 있다.

$$\begin{aligned} PMC_1(i, j; I, J) &= \sum_{k=0}^2 |p_{(I, J)}(k) - p_{(i, j)}(k)|, \\ PMC_2(i, j; I, J) &= \sum_{k=3}^4 |p_{(I, J)}(k) - p_{(i, j)}(k)|, \\ PMC_3(i, j; I, J) &= |p_{(I, J)}(5) - p_{(i, j)}(5)| \end{aligned} \quad (10)$$

이때, 임시의 운동 대응점 (i, j) 가 정확한 픽셀 단위의 대응점이 아닐 수 있으므로 그 주변 픽셀에 대하여 식 (11)을 적용하여 정확한 픽셀 단위의 대응점을 찾을 수 있다.

$$\begin{aligned} \min_{(i, j) \in R} PMC_1(i, j; I, J), \\ \min_{(i, j) \in R} PMC_2(i, j; I, J), \\ PMC_3(i, j; I, J) \leq M_{Threshold} \end{aligned} \quad (11)$$

여기서, R 은 *image-3*에서의 검색 영역을 나타낸다. 위 식을 이용하여 픽셀 단위의 운동 대응점을 얻으면 그 위치를 기준으로 다시 정확하게 픽셀 단위 이하의 대응점을 찾는 과정이 필요하다. 본 연구에서는 유동시 제한 조건식에 정확한 운동 모델식을 적용하여 픽셀 단위 이하의 운동 대응점을 구하고자 한다. 이때, 유동시 제한 조건식에 미소운동모델[10]의 (2)를 결합하여 다음과 같은 식을 얻는다 [11].

$$I_x \Delta u + I_y \Delta v + c_0 t_z + c_0_2 w_x + c_0_3 w_y + c_0_4 w_z + I_t = 0 \quad (12)$$

$$c_0_1 \equiv (\Delta x \cdot I_x + \Delta y \cdot I_y) / z,$$

$$c_0_2 \equiv (x \Delta y + y \Delta x + \Delta x \Delta y) \cdot I_x + (2y \Delta y + \Delta y \Delta y) \cdot I_y,$$

$$c_0_3 \equiv -(x \Delta y + y \Delta x + \Delta x \Delta y) \cdot I_x - (2x \Delta x + \Delta x \Delta x) \cdot I_y,$$

$$\text{여기서, } c_0_4 \equiv \Delta y \cdot I_x - \Delta x \cdot I_y$$

위에서는 I_z, w_x, w_y, w_z 는 이미지 전체에 작용하는 운동 매개 변수이며, $c_0_1, c_0_2, c_0_3, c_0_4$ 는 기준 이미지의 픽셀 좌표값

(I,J) 위치에서의 밝기값 및 좌표 값으로부터 계산되어지는 값이다. 물론 위 식도 한 픽셀에 대하여 6개의 미지수에 대한 하나의 방정식을 나타내므로 under-constraint 조건식이다. 따라서, image-3의 중심 픽셀 (i,j)에서의 픽셀 이하의 운동 ($\Delta u, \Delta v$)는 (13)과 같이 block matching을 수행하여 얻을 수 있다.

$$E_{D2}(\mathbf{mp}) = \sum_{\mathbf{x} \in D2} \mathbf{W}^2(\mathbf{x}) \mathbf{W}_p^2(\mathbf{x}) \left(\begin{matrix} I_x(\mathbf{x})\Delta u + I_x(\mathbf{x})\Delta v + c_{01}(\mathbf{x})t_z + c_{02}(\mathbf{x})w_x \\ + c_{03}(\mathbf{x})w_y + c_{04}(\mathbf{x})w_z + I_t(\mathbf{x}) \end{matrix} \right)^2 \quad (13)$$

여기서, $\mathbf{mp} = [\Delta u \ \Delta v \ t_z \ w_x \ w_y \ w_z]^T$ 를 나타낸다. 그러나, 위 식의 추정은 다음과 같은 어려운 점이 있다. 중심 픽셀 주위의 넓은 영역을 추정에 이용하면 그에 따른 3차원 깊이(z)가 일정하다는 가정을 위반하기 쉬우므로 추정의 오차를 가지며, 또한 중심 픽셀의 작은 영역을 추정에 이용하면 매개변수의 수가 많은 가운데 적은 수의 측정값으로부터의 추정이 이루어지므로 그 결과는 노이즈에 민감하게 된다. 따라서 본 연구에서는 매개변수의 수를 줄이는 방법으로 유동시 제한 조건식[13, pp.282]의 이미지 밝기값의 시간 미분값에 대한 등가치를 다음과 같이 새롭게 정의하고자 한다.

$$I_{t,E}(\mathbf{x}) \equiv I_t(\mathbf{x}) + c_{01}(\mathbf{x})t_z + c_{02}(\mathbf{x})w_x + c_{03}(\mathbf{x})w_y + c_{04}(\mathbf{x})w_z \quad (14)$$

여기서, 4개의 운동 t_z, w_x, w_y, w_z 는 초기의 강인 추정법에 의하여 얻어진 값을 이용한다는 전제 아래서 미리 알 수 있는 매개 변수로 가정한다. 따라서, 픽셀 단위 이하의 운동 대응점의 추정은 미소운동모델을 이용한 block matching 방법에 의하여 다음과 같이 정리할 수 있다.

$$E_{D2}(\Delta u, \Delta v) = \sum_{\mathbf{x} \in D2} \mathbf{W}^2(\mathbf{x}) \mathbf{W}_p^2(\mathbf{x}) \left(I_x(\mathbf{x}) \cdot \Delta u + I_x(\mathbf{x}) \cdot \Delta v + I_{t,E}(\mathbf{x}) \right)^2 \quad (15)$$

그러므로, (11)과 (15)를 이용하여 기준 이미지의 임의의 픽셀 (I,J)의 정제된 운동 대응점은 다음과 같이 얻을 수 있다.

$$\begin{aligned} u &= (i - I) + \Delta u, \\ v &= (j - J) + \Delta v \end{aligned} \quad (16)$$

이때, 정제된 운동 대응점과 스테레오 매칭에서 얻은 정제된 깊이 정보를 이용하여 운동 매개변수의 정제된 추정은 다음의 식과 같이 강인 추정법을 이용하여 다시 계산할 수 있다.

$$\min_{\mathbf{t}, \mathbf{w}} \sum_i \rho \left(\frac{1}{z} \mathbf{A} \cdot \mathbf{t} + \mathbf{B} \cdot \mathbf{w} - \mathbf{u}, \sigma \right) \quad (17)$$

$$\mathbf{A} = \begin{bmatrix} -1 & 0 & x \\ 0 & -1 & y \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} xy & -(1+x^2) & y \\ (1+y^2) & -xy & -x \end{bmatrix}$$

여기서, 운동 매개변수는 다시 반복적으로 추정을 해야 하는데, 이는 (14)에서의 운동 매개 변수값이 근사 초기치를 이용했기 때문이다. 따라서, (14)의 등가 이미지 밝기값의 시간 미분값

은 다음과 같이 재귀적 형태로 표현할 수 있다.

$$I_{t,E,n+1}(\mathbf{x}) \equiv I_t(\mathbf{x}) + c_{01}(\mathbf{x})t_{z,n} + c_{02}(\mathbf{x})w_{x,n} + c_{03}(\mathbf{x})w_{y,n} + c_{04}(\mathbf{x})w_{z,n} \quad (18)$$

여기서, n 은 반복 스텝을 나타낸다.

3. 제안 알고리즘

스테레오 연속 영상으로부터 카메라의 운동과 3차원 깊이를 정제하는 방법으로써의 본 연구의 알고리즘을 정리하면 다음과 같다.

- 단계 1: 기존의 강인 직접 추정법을 수행하여 운동 매개변수와 깊이의 초기치를 가장 낮은 피라미드 레벨에서 추정한다.
- 단계 2: (4)를 만족하는 스테레오 매칭 점들을 찾는다.
- 단계 3: (5)와 (7)을 이용하여, 단계-2에서 얻은 점들에 대하여 깊이를 정제 추정한다.
- 단계 4: 단계-2를 만족하는 점들에 대하여, (10)을 만족하는 운동 대응점들을 찾아낸다.
- 단계 5: (15)와 (16)을 이용하여 픽셀 단위 이하의 운동 대응점을 계산한다.
- 단계 6: (17)을 이용하여 운동 매개변수를 추정한다.
- 단계 7: (18)을 이용하여 이미지 밝기값의 등가 미분값을 갱신한다.
- 단계 8: 단계-5에서 단계-7의 과정을 (17)의 오차의 합이 작아질 때까지 수행하며, 오차의 합이 작아지지 아니하는 시점에서 정제 과정을 종료한다.

III. 실험

연속적인 실제 스테레오 이미지를 바탕으로 하여 본 알고리즘에 대한 검증 실험을 수행하였다. 그림 3은 실험 이미지의 한 예를 보여주고 있는데 이때의 스테레오 카메라의 외부 보정변수는 $t_x = 8mm, t_y = t_z = 0, w_x = w_y = w_z = 0$ 이며, 보정은 Tsai 알고리즘 [12]을 이용하였다. 실제 실험에 의해서도 bas-relief ambiguity가 잘 나타날 수 있도록 카메라의 시야각을 작게 구성하고 또한 3차원 깊이의 변화도 작게 구성하였다. 이 경우의 focal length는 16mm, 최대 시야각은 23°, 그리고 카메라는 SONY-XC75를 사용하였다.

표 1에서는 그림 3의 스테레오 연속 이미지에 대한 기존의 강인 추정법에 의한 운동 매개변수의 추정의 결과와 본 연구에 의한 정제 방법이 적용된 경우의 추정 결과를 비교하여 보여주고 있다. 표 1에서 볼 수 있는 바와 같이 기존의 강인 추정법에 의해서도 운동 매개변수의 추정의 모호성이 발생함을 알 수 있는데 회전 w성분이 직진 운동 매개변수 t로 전이되어 나타남을 알 수 있다. 그러나 본 연구에서 제안한 알고리즘은 기존의 강인 추정법의 추정 결과를 바탕으로 그러한 모호성을 개선할 수 있음을 보여주고 있으며, 또한 제안 알고리즘의 단계-8을 통하여 재귀적으로 더 운동 매개 변수를 정확하게 추정해 나갈 수 있음을 보여주고 있다. 그림 4는 알고리즘의 단계-8의 재귀적 과정을 통하여 운동 매개변수 추정의 오차를 줄여나갈 수 있는지를 보여주고 있다. 그림 4에서 Terr은 병진 운동매개 변수의 오차의 합계

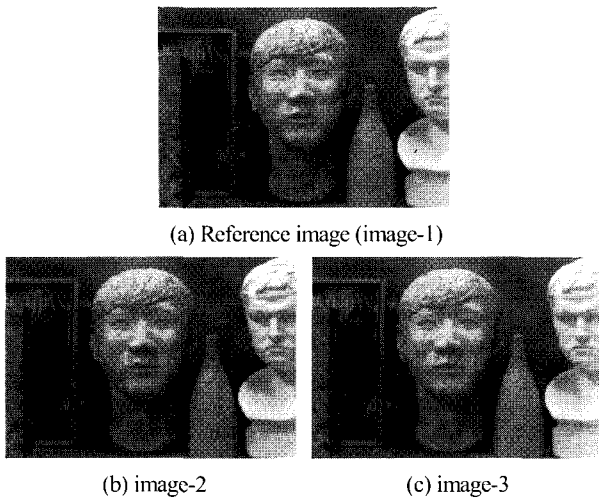


그림 3. 실제 실험을 위한 스테레오 연속 영상.
Fig. 3. Stereo sequence image for experiment.

표 1. 그림 3에 대한 운동 변수 추정.

Table 1. Motion parameter estimation about Fig. 3.

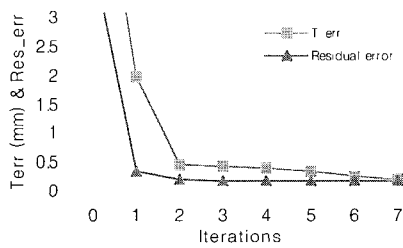
운동변수	*A	B	C	D
t_x (mm)	0.0	6.66	1.96	0.20
t_y (mm)	0.0	-0.10	-0.40	0.09
t_z (mm)	0.0	1.24	0.55	0.27
w_x (mm)	0.0	0.00	-0.02	0.00
w_y (mm)	0.5	0.08	0.34	0.44
w_z (mm)	0.0	-0.03	-0.02	-0.02

*A : Exact motion parameters

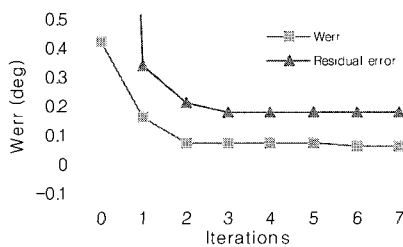
B: Robust method by only global optimization

C : Proposed method (at 1st iteration)

D : Proposed method (at 7th iteration)



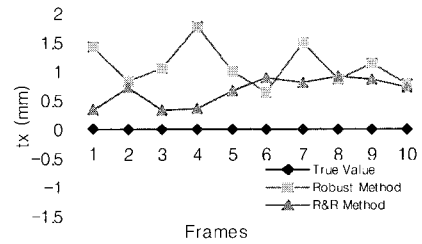
(a) Error of translational parameters



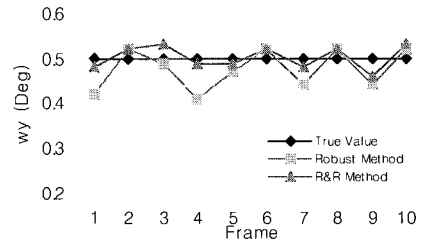
(b) Errors of rotational parameters

그림 4. 재귀적 과정에 따른 추정 오차.

Fig. 4. Error transition according to iterative refinement.



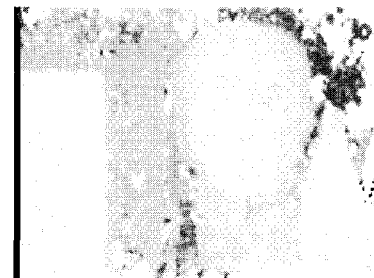
(a) Estimation of t_x



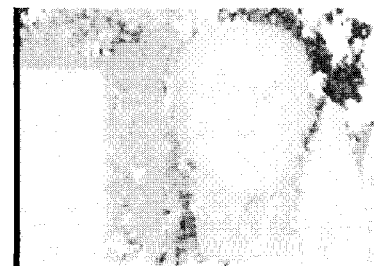
(b) Estimation of w_y

그림 5. 다수의 연속 영상에 대한 운동 추정 결과.

Fig. 5. Motion estimation with multiple sequence images.



(a) Depth estimation by previous method[1]



(b) Depth estimation by proposed method

그림 6. 그림 5의 4번째 영상에서의 깊이 추정 결과.

Fig. 6. Depth estimation with 4-th sequence in Fig. 5.

를 나타내며, Werr은 회전 운동 매개변수의 추정 오차의 합을 나타낸다.

그림 5는 그림 3의 실험 구성에 대하여 연속적인 스테레오 영상을 획득하여 그 연속 추정의 예를 보여주는 그림이다. 그림에서 볼 수 있듯이 1, 4, 7, 및 9번째 스테레오 영상에서는 bas-relief 모호성이 심하게 나타남을 알 수 있으며, 본 연구에서 제안한 알고리즘에 의하여 그러한 추정 모호성을 현저히 개선할 수 있음을 보여주고 있다. 그림 6은 그림 5의 4번째 영상에 대한 깊이의 추정의 예를 보여주고 있다. 그림 6의 (a)는 기존의 강인 추정법

에 의한 깊이 추정의 결과이며, 그림 6의 (b)는 본 알고리즘에 의하여 정제된 운동 매개변수 정보를 다시 이용하여 깊이를 다시 추정한 예를 보여주고 있다. 그림에서 아주 희거나 검은 부분은 추정의 오차가 있거나 계산이 불가능한 영역인데 본 알고리즘에 의한 방법이 깊이 추정의 정확성도 더 높일 수 있음을 보여주고 있다.

IV. 결론

본 연구에서는 보정된 스테레오 연속 영상으로부터 카메라의 운동과 3차원 깊이를 복원하는 경우에 발생하는 bas-relief 모호성 문제를 다루었다. 본 연구의 가장 큰 특징은 운동 모호성을 개선할 수 있는 가장 중요한 요소가 운동 대응점들을 픽셀 단위 이하로 얼마나 정확하게 얻을 수 있는냐에 초점을 맞추어 기존 의단선적인 연구와는 달리 재귀적 방법을 제시하였다. 이를 위해 기존의 강인 추정법을 대응점의 초기치를 구하는 방법으로 사용하여 다시 스테레오 매칭과 운동 대응점들을 픽셀 단위 이하로 추정하는 방법을 제시하였다. 또한 실제 스테레오 이미지를 이용한 실험을 통하여 본 알고리즘이 bas-relief 모호성을 개선할 수 있는 점과 3차원 깊이(depth)도 운동 매개 변수의 추정의 정확성에 따라 개선될 수 있음을 보였다.

참고문헌

- [1] S. K. Park and I. S. Kweon, "Robust and direct estimation of 3-D motion and scene depth from stereo image sequence," *Pattern Recognition*, pp. 1713-1728, Vol. 34, No. 9, 2001.
- [2] J. Weng, N. Ahuja and T. S. Huang, "Optimal motion and structure estimation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 9, pp. 864-884, 1993.
- [3] R. Szeliski and S. B. Kang, "Shape ambiguities in structure from motion," *IEEE Trans. on Pattern Analysis and Machine*

Intelligence, Vol. 19, No. 5, pp. 506-512, 1997.

- [4] S. Soatto and R. Brockett, "Optimal structure from motion: Local Ambiguities and Global Estimates," *CVPR*, pp. 282-288, 1998.
- [5] M. Irani, B. Rousso and S. Peleg, "Recovery of Ego-Motion using region alignment," *IEEE Tr. On Pattern Analysis and Machine Intelligence*, Vol. 19, No. 3, pp. 268-272, 1997.
- [6] A. Jepson and D. Heeger, "Linear space method for recovering rigid motion," *Spatial Vision in Humans and Robots*, Cambridge University Press, 1992.
- [7] G. P. Stein and A. Shashua, "Model-Based brightness constraints: On direct estimation of structure and motion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 9, pp. 992-1015, 2000.
- [8] A. M. Tekalp, *Digital Video Processing*, Prentice Hall PTR, Upper Saddle River, NJ 07458, 1995.
- [9] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window : Theory and Experiment," *IEEE Tr. On Pattern Analysis and Machine Intelligence*, Vol. 16, No. 9, pp. 920-932, 1994.
- [10] H. Longuet-Higgins and P. Prazdny, "The interpretation of a moving retinal image," *Proceedings of the Royal Society of London B*, Vol. 208, pp. 385-397, 1980.
- [11] S. Park, Robust and Direct Estimation of Camera Motion and 3-D Structure from Stereo Image Sequence, PhD Thesis, KAIST, Korea, 2000.
- [12] Y. Tsai, "A versatile camera calibration technique for High-Accuracy 3D machine vision metrology using Off-the-Shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, Vol. 3, No. 4, pp. 323-344, 1987.
- [13] B. K. P. Horn, *Robot Vision*, The MIT Press, USA, 1986.

박성기

제어 · 자동화 · 시스템공학 논문지 제 8 권 제 5 호 참조.

권인소

제어 · 자동화 · 시스템공학 논문지 제 4 권 제 3 호 참조.