

# 단안 영상 시퀀스에서 움직임 추정 기반의 3차원 깊이 정보 추출 알고리즘

박 준 호<sup>†</sup> · 전 대 성<sup>††</sup> · 윤 영 우<sup>†††</sup>

## 요 약

2차원 영상으로부터 3차원 영상으로 복원하는 일은 일반적으로 카메라의 초점에서 영상 프레임의 각 픽셀까지의 깊이 정보가 필요하고, 3차원 모델의 복원에 관한 일반적인 수작업은 많은 시간과 비용이 소모된다. 본 논문에서는 카메라의 움직임이 포함되어 있는 단안 영상 시퀀스로부터 3차원 영상 제작에 필요한 상대적인 깊이 정보를 실시간으로 추출하는 알고리즘을 제안하고, 하드웨어를 구현하기 위하여 알고리즘을 단순화하였다. 이 알고리즘은 카메라 이동에 의한 영상의 모든 점들의 움직임은 깊이 정보에 종속적이라는 사실에 기반을 두고있다. 블록매칭 알고리즘에 기반을 둔 전역 움직임 탐색에 의한 움직임 벡터를 추출한 후, 카메라 회전과 확대/축소에 관한 카메라 움직임 보상을 실행하고 깊이 정보 추출 과정이 전개된다. 깊이 정보 추출 과정은 단안 영상에서 객체의 이동거리를 분석하여 움직임 벡터를 구하고 프레임내의 모든 픽셀에 대한 평균 깊이를 계산한 후, 평균 깊이에 대한 각 블록의 상대적 깊이를 산출하였다. 모의 실험 결과 전경과 배경에 속하는 영역의 깊이는 인간 시각 체계가 인식하는 상대적인 깊이와 일치한다는 것을 보였다.

## 3D Depth Information Extraction Algorithm Based on Motion Estimation in Monocular Video Sequence

Joon-Ho Park<sup>†</sup> · Dae-Seong Jeoune<sup>††</sup> · Young-Woo Yoon<sup>†††</sup>

## ABSTRACT

The general problems of recovering 3D from 2D imagery require the depth information for each picture element from focus. The manual creation of those 3D models is consuming time and cost expensive. The goal in this paper is to simplify the depth estimation algorithm that extracts the depth information of every region from monocular image sequence with camera translation to implement 3D video in realtime. This paper is based on the property that the motion of every point within image which taken from camera translation depends on the depth information. Full-search motion estimation based on block matching algorithm is exploited at first step and then, motion vectors are compensated for the effect by camera rotation and zooming. We have introduced the algorithm that estimates motion of objects by analysis of monocular motion picture and also calculates the average of frame depth and relative depth of region to the average depth. Simulation results show that the depth of region belongs to a near object or a distant object is in accord with relative depth that human visual system recognizes.

키워드 : vision, 3D, stereo

### 1. 서 론

최근에 멀티미디어 기술의 발달과 더불어 영상 정보의 형태는 인간의 시각 기능에 기초하여 보다 사실적이고 입체적 형태로 표현되고 있는 실정이다. 입체 영상의 표현 기법에 관련된 연구는 3차원 영상의 제작 및 표현 기법이 주된 연구 대상이 되어왔으며, 2차원 영상을 3차원 입체 영상으로 변환

하기 위해서는 영상의 깊이 정보가 필요하다. 깊이 정보 추출 기법은 3차원 모델링 분야에 주로 사용되어 왔다[1, 2]. 객체의 3차원 모델들은 현실적으로 표현이 불가능한 상황을 가상공간에서 표현하거나, 다양한 멀티미디어 응용분야의 구성 요소로 사용되고 있다[3]. 실세계에서 3차원 객체에 대한 카메라와 인간의 영상 인식 시스템은 원근 투영(perspective projection)을 수행하는데, 이것은 많은 양의 정보 손실을 초래한다[7]. 2차원에서 3차원을 추론하는 것은 매우 복잡한 메카니즘을 요구하기 때문에 3차원 영상처리 분야에서 여전히 미해결의 과제로 남아있으며, 여러 해 동안에 이 분야의 주요한 주제로 간주되어 왔다.

\* 본 연구는 한국과학재단 목적기초연구(과제번호 : 104533)지원으로 수행되었음.

† 정 회 원 : 경운대학교 컴퓨터공학과 교수

†† 정 회 원 : 대구미래대학 멀티미디어정보과학과 교수

††† 총신회원 : 영남대학교 전자정보공학부 교수

논문접수 : 2001년 5월 28일, 심사완료 : 2001년 9월 14일

본 논문에서는 2차원 단안 영상을 3차원 영상으로 실시간으로 변환하는 하드웨어를 구현하기 위하여 카메라의 이동이 있는 2차원 단안 영상으로부터 상대적 깊이 정보를 추출하는 알고리즘을 제안한다. 실시간으로 처리하기 위해서 하드웨어 구현이 용이한 전역 움직임 추정 방식을 도입하였고, 카메라 회전에 의한 전역 움직임을 제거한 영상을 입력 영상으로 사용하였으며, 제안한 알고리즘은 하드웨어 구현에 용이하도록 구성하였다.

본 논문의 구성은 다음과 같다. 2장에서는 연구 배경을 소개하고, 3장은 기존의 관련 연구에 대하여, 4장에서는 제안한 알고리즘의 기본 개념을 설명한다. 5장은 움직임 추정과 카메라의 회전각 보상에 대하여 기술하였고, 6장에서는 카메라 이동에 대한 기하를, 7장에서는 상대적 깊이 정보 추출에 대하여 서술하였다. 8장은 제안한 알고리즘을 모의 실험에 적용하여 실험하고 분석하였으며, 9장에서는 결론 및 향후 연구과제를 나타내었다.

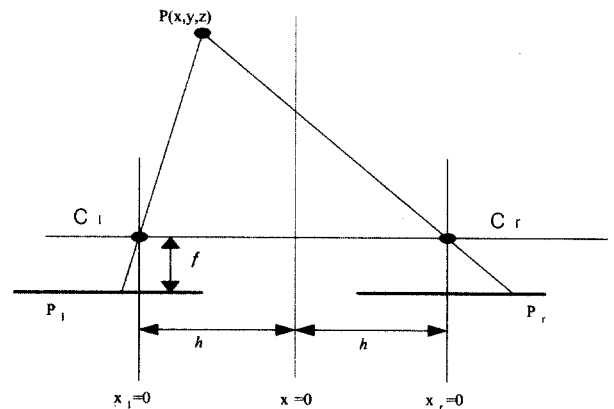
## 2. 관련 연구

3차원(3D) 모델링에 관한 작업을 수작업으로 하는 것은 많은 시간과 노력이 소요되기 때문에 3차원 객체를 자동으로 재구성하는 여러 가지 기법들이 연구되고 있다. 초기의 비전 연구는 생물학과 인공지능분야에서 시작되었으며 스테레오(stereo) 영상으로부터 3차원 영상을 복원하려는 분야에서 많이 연구되었는데, 주로 반복적 공동(iterative cooperative) 알고리즘을 통하여 영상 시퀀스로부터 두 장의 영상을 사용하여 상응(correspondence)을 자동적으로 추출하고자 하였다. 스테레오 비전은 일반적으로 두 대의 카메라에서 획득한 영상 시퀀스에서 동일한 시간에 촬영된 두 장의 프레임만을 가정한다. 이 알고리즘은 두 영상으로부터 유일한 매칭점을 찾아서 3차원 깊이의 중간 형태를 표현하는 평탄 불일치(smooth disparity)를 복원한다[8]. 대부분의 SfM(structure from motion) 기술은 원근 투영(perspective projection) 모델을 가정하고 있는데[3], 최근에는 투영 기하학(projective geometry)이 스테레오 비전에 적용되었고[9], 외극 기하학(epipolar geometry)은 투영 불변식(projective invariants)과 복원분야에서 많은 연구 활동이 이루어지고 있다[10, 11].

인간의 시각은 위치가 다른 두 눈으로부터 다른 관점의 두 영상을 인식한다. 이와 동일한 방법으로 영상을 획득하는 기법이 스테레오 비전분야이다. 위치가 다른 두 곳으로부터 획득된 영상에서 깊이 정보를 유도할 수가 있다. 두 눈이 인식하는 방법과 유사하게 두 영상으로부터 깊이 정보를 추출할 수 있다. 스테레오 비전분야는 카메라 교정(calibration), 좌우 영상의 상응(correspondence) 추출과 3차원 좌표계의 재구성에 응용되고 있다.

(그림 1)은 기본적인 스테레오 기하의 표준 형태를 보여주

고 있는데, 두 대의 카메라에 대한 광축은 서로 평행하고, 상호간의 거리는  $2h$ 라고 가정했을 때, 한 점  $P(x, y, z)$ 를 투영시킨 두 영상을 각각  $P_l$ 과  $P_r$ 이라고 하면,  $z$ 는 카메라로부터의 거리를 나타내고  $x$  축은 수평 거리를 나타낸다.  $x=0$ 은 두 카메라 사이의 정중앙 위치를 말하며, 각각의 영상은 지역(local) 좌표계를 가진다. 즉, 왼쪽에서는  $x_l$ , 오른쪽에서는  $x_r$ 이다. 두 카메라의 위치 차이로 인하여  $x_l$ 과  $x_r$  사이에 불일치(disparity)가 존재하는 것은 분명하다. 다시 말하면,  $|P_l - P_r| > 0$ 이 성립하고, 선분  $PIC_l$ 과 선분  $PIC_r$ 는 직각삼각형의 닮은 변이다. 이러한 사실로부터 다음의 식 (1)이 유도된다.



(그림 1) 기본적인 스테레오 기하의 표준 형태

$$\frac{P_l}{f} = \frac{h+x}{z}, \quad \frac{P_r}{f} = \frac{h-x}{z} \quad (1)$$

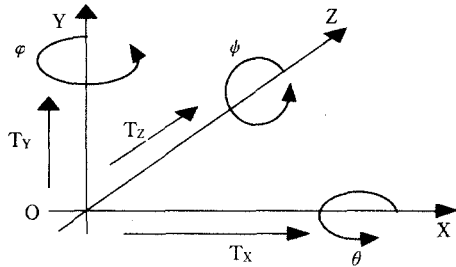
스테레오 비전에서는 식 (2)와 같이 깊이 정보를 구할 수 있다.

$$z = \frac{2hf}{P_r - P_l} \quad (2)$$

## 3. 연구 배경

3차원 영상에서 2차원 원근 영상은 구할 수 있지만, 2차원 영상 시퀀스에서 3차원 영상을 재구성하는 것은 불가능 문제이다. 그래서 초점으로부터 영상의 실제 거리를 구할 수는 없고, 각 영역들간의 상대적인 깊이를 구할 수 있다. 2차원 영상으로부터 3차원 구조를 재구성하는 문제는 카메라 초점으로부터 각 영상 프레임 점까지의 깊이 정보가 필요하다. 여기서, 깊이 정보는 카메라에서부터 객체까지의 직교 거리로 정의될 수 있다. 카메라의 움직임이 포함된 영상에서 움직임 분석을 하면, 대부분의 영역이 움직임을 포함하고 있다. 전경(foreground)에 속하는 영역은 카메라의 움직임에 독립적인 움직임을 나타내지만, 배경(background)에 속하는 영역은 카메라의 이동에 따라 움직임의 변화량이 종속적이다. 이와

같이 카메라의 움직임과 각 영상의 객체들의 움직임과는 긴밀한 관계가 있다. 만일 카메라의 이동에 대한 각 객체들의 움직임을 정확히 알고 있다면, 카메라가 이동하는 중에 영상의 객체가 정지 상태라는 가정 하에서 카메라 초점에서 객체까지의 깊이 정보를 추정할 수 있고, 반대로 영상의 각 영역의 깊이 정보를 정확히 알고 있으면, 객체의 움직임 정보를 쉽게 추출할 수 있다[14]. 카메라의 움직임은 (그림 2)와 같이 회전(rotation)성분과 이동(translation)성분으로 표현될 수 있는데, 회전성분에는 Y축 회전각(panning), X축 회전각(tilting), Z축 회전각(rolling)과 확대/축소(zooming)가 있으며, 이동성분에는 수평이동(tracking), 수직이동(booming), 전후이동(dolling)이 있다. 영상의 깊이 정보는 카메라의 이동성분(translation)에 종속적이지만, 회전과 확대/축소에는 독립적이다.



회전성분 : panning( $\phi$ ), tilting( $\theta$ ), rolling( $\psi$ )  
이동성분 : tracking( $T_x$ ), booming( $T_y$ ), dolling( $T_z$ )

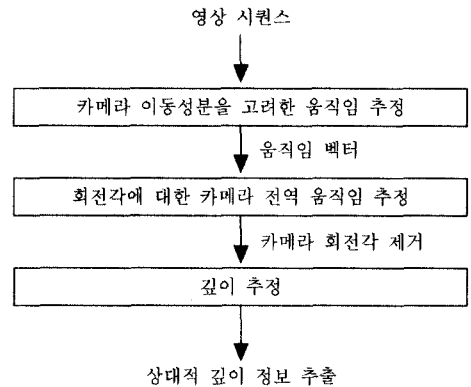
(그림 2) 카메라 이동과 회전의 성분

#### 4. 제안한 깊이 정보 추출 알고리즘

연속된 영상간의 변화차분에 의해서 영상을 배경과 전경으로 분할한다. 카메라의 회전에 의해 배경에 변화차분이 발생하여 전경으로 인식될 수 있으므로 참조 영상은 카메라 회전각에 대한 보상을 한 영상을 사용한다[15]. 움직임 벡터를 분석하여 깊이 정보를 추출하므로 카메라의 회전각에 의하여 발생한 움직임은 보상되어야 한다. 본 논문의 목적은 카메라가 이동하면서 촬영을 한 단안 영상 시퀀스의 각 영역에 대한 깊이 정보를 산출하는 것이다. 제안한 알고리즘은 2차원 단안 영상을 입력하고 (그림 3)에서 기술된 바와 같이 3 단계로 구성된다. 깊이 정보 추출 과정은 카메라로부터 획득된 2차원 단안 영상에 대하여 3 단계로 구성되어 있다. 첫 번째 단계는 움직임 분석 단계이며, MPEG 2-Video에서 적용하고 있는 전역 움직임 탐색 기법을 채택하였다. 전역 움직임 탐색 기법은 시간이 많이 소요되는 단점이 있으나, 국부 극소(local minima)에 빠질 위험이 가장 적고, 보다 정확한 움직임 벡터를 산출하는 장점이 있으며 하드웨어 구현이 용이하고 실시간 처리가 가능하기 때문에 선택하였다[6]. 두 번째 단계가 카메라의 회전각 보상이다. 카메라의 회전에 의한 영상 프레임 내부의 움직임은 객체의 깊이 정보에 관계없이 일정한 움직

임이 나타난다.

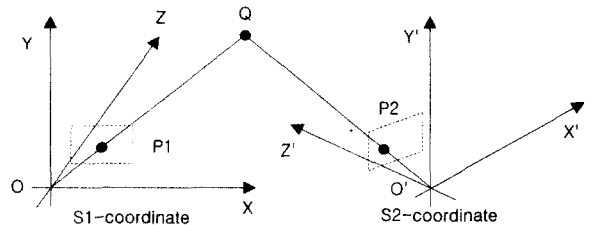
본 논문에서는 객체의 움직임을 기반으로 깊이 정보를 추출하므로 카메라 회전에 의한 객체의 움직임은 보상되어야 하며, 대부분의 영상 시퀀스에는 카메라의 회전에 의한 움직임이 포함되어 있다. 객체의 이동은 국부적(local)인 반면, 카메라의 회전 움직임은 전역적(global)이므로 프레임 전체에 영향을 미친다. 따라서, 카메라의 회전에 의한 움직임이 포함되어 있는 영상에 대해서는 움직임 탐색을 제대로 수행할 수가 없다. 이러한 문제는 영상의 움직임을 분석하여 추정된 카메라의 움직임을 영상에 적용하여 보상함으로써 해결할 수 있다[16]. 마지막 단계에서는 카메라 회전각에 대하여 보상된 움직임 벡터 값을 이용하여 영상 전체 프레임에 대한 각 영역의 움직임 벡터 정보를 기반으로 하여 제안한 알고리즘을 적용하여 프레임내의 각 영역간의 상대적인 깊이 정보를 추출한다.



(그림 3) 깊이 정보 추출 흐름도

#### 5. 움직임 추정과 회전각 보상

강체(rigid object)의 3차원 변위(displacement)는 카메라의 움직임에 의해서 (그림 4)와 같이 S1 좌표계에서 S2 좌표계로 변환되는 모델로 나타낼 수 있다[5].



(그림 4) 카메라 이동과 좌표계의 변환

$$(X', Y', Z') = R \cdot (X, Y, Z) + T \quad (3)$$

여기서,  $R$ 은 3x3 회전 행렬이고,  $T$ 는 3D 이동 벡터, 그리고  $X$ 와  $X'$ 는 시간  $t$ 와  $t'$ 에서 객체의 점들의 좌표를 각각 나타

낸다[6].

원근 투영은 기하학적 광학(geometrical optics) 규칙에 따라 영상 프레임의 정보를 반영한다. 원근 변환의 대수학적인 관계는 다음과 같다.

$$\frac{X}{x} = \frac{Y}{y} = \frac{Z}{f} = \lambda, \quad X = \lambda x, \quad Y = \lambda y, \quad Z = \lambda f \quad (4)$$

3차원 공간의 점 Q (X, Y, Z)가 영상에 맺는 위치를 점 p(x, y)라 하고, S1 좌표계의 초점 거리를 f, 점 Q의 깊이를 λf, S2 좌표계의 초점 거리를 f', 점 Q의 깊이를 μf'이라고 정의하면, 3차원 변위는 식 (5)로 정의된다.

$$[\mu x', \mu y', \mu f']^T = \mathbf{R} \cdot [\lambda x', \lambda y', \lambda f']^T + \mathbf{T} \quad (5)$$

카메라 이동이 (Tx, Ty, 0)인 경우에는, 위 방정식에 R과 T를 대치하여 재정리하면, 다음 방정식 (6)이 성립한다.

$$x' = x - \frac{T_x}{1 - \frac{T_z}{f\lambda}}, \quad y' = y - \frac{T_y}{1 - \frac{T_z}{f\lambda}} \quad (6)$$

여기서, R, T와 λ는 다음과 같다.

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{T} = (-T_x, -T_y, 0), \quad \lambda = \mu$$

확대(zooming)일 경우에는 카메라 이동의 경우처럼 다음 방정식 (7)을 얻는다.

$$x' = \frac{x}{1 - \frac{T_z}{f\lambda}}, \quad y' = \frac{y}{1 - \frac{T_z}{f\lambda}} \quad (7)$$

단,

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{T} = (0, 0, T_z), \quad \mu = \lambda - \frac{T_z}{f}$$

이다.

R이 단위 벡터 성분을 가지는 것은 카메라 회전각에 대하여 보상을 수행하였기 때문이다. 그러므로 X' = X + T로 표현이 가능하다.

x와 y에 대하여 일반화한 새로운 영상 점인 x', y'를 다음 식 (8)과 같이 정규화하였다.

$$x' = \frac{x - \frac{fT_x}{Z}}{1 - \frac{T_z}{Z}}, \quad y' = \frac{y - \frac{fT_y}{Z}}{1 - \frac{T_z}{Z}} \quad (8)$$

단, Z는 점 (x, y)에서의 깊이이다.

카메라 이동이 있는 영상에서 움직임 벡터를 다음과 같이 정의하였다.

$$v_x = x' - x = \frac{-fT_x + xT_z}{Z - T_z} \quad (9)$$

$$v_y = y' - y = \frac{-fT_y + yT_z}{Z - T_z} \quad (10)$$

여기서 (x, y)와 (x', y')는 시간축 t와 t'상에서, Z축에 대한 영상 점(pixel)을 각각 나타낸다. f, Tx, Ty와 Tz는 프레임내에서 유일한 값을 가지는 전역 변수이고, v'x, v'y와 Z는 각 블록의 영역에 따라서 다른 값을 가지는 지역 변수이다. 또한, v'x와 v'y는 블록 정합 기법을 기반으로 하는 전역 탐색에 의해서 구할 수 있다. 블록 정합의 판단 기준은 평균 절대 오차(Mean Absolute Difference : MAD)와 평균 자승 오차(Mean Squared Difference : MSD)가 많이 사용된다[12]. 본 논문에서는 MSD에 비하여 계산량이 적고 하드웨어 구현이 용이한 MAD를 선택하였다[13].

직교(Cartesian) 좌표계에서 3차원 회전은 세 좌표축에 대한 회전의 오일러(Eulerian) 각으로 표시되거나, 회전축과 각으로 나타낼 수 있다[4]. 식 (9)와 식 (10)에서 구한 움직임 벡터를 포함하여 제안한 알고리즘에서는 카메라 회전 성분인 회전각을 포함한 움직임 벡터를 식 (11)과 식 (12)로 재정 의하였다.

$$v_x = v_x - \frac{-\phi f y + \phi f^2 + \phi x^2 - \theta x y}{f - \phi x + \theta y} \quad (11)$$

$$v_y = v_y - \frac{-\phi f x + \phi f^2 + \phi x y - \theta y^2}{f - \phi x + \theta y} \quad (12)$$

이 방정식은 카메라 회전각에 의한 움직임이 제거된 움직임 벡터를 나타낸다[6]. 여기서, v'x와 v'y는 전역움직임 탐색 후 산출된 움직임 벡터이다.

## 6. 카메라 이동에 대한 기하

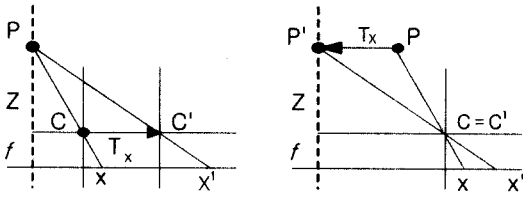
방정식 (9)와 (10)에서 초점거리 f를 제외한 다른 모든 변수의 값을 구할 수 있으므로 변수 Z의 값을 구할 수 있다.

$$Z = \frac{1}{v_x}(-fT_x + xT_z) + T_z = \frac{1}{v_y}(-fT_y + yT_z) + T_z \quad (13)$$

움직임 추정으로부터 각 블록에 대한 움직임 벡터 (vx, vy)를 구하고, 프레임내에서 일정한 값을 가지는 Tx, Ty, Tz와 f는 전역 변수이고, Z, x, y, vx, vy는 화소의 위치에 따라서 변화하는 값을 가진다.

(그림 5)는 카메라의 위치가 C에서 C'로 이동하였을 때, 영상 평면 위의 화소의 위치가 x에서 x'로 수평 이동되는 관계를 나타내었다.

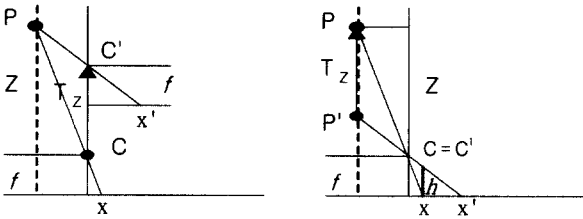
$$v_x = \frac{fT_x}{Z} \quad (14)$$



(그림 5) 카메라 수평 이동에 대한 기하

여기에서, 움직임 벡터의  $x$  성분 값인  $v_x$ 를 알 수 있다. 동일한 방법으로 움직임 벡터  $y$  성분 값을 알 수 있다.

$$v_y = \frac{fT_Y}{Z} \quad (15)$$



(그림 6) 카메라의 전후 이동에 대한 기하

(그림 6)은 카메라의 위치가 C에서 C'로 이동하였을 때, 화소의 위치가 이동되는 관계를 도식화하였다.

$$\frac{h}{f} = \frac{v_x}{x + v_x} \quad (16)$$

위의 (그림 6)과 식 (16)으로부터 다음의 식 (17)을 유도하였다.

$$\frac{v_x}{x + v_x} = \frac{T_Z}{Z} \quad (17)$$

식 (14), 식 (15)와 식 (17)에서 전역 변수인  $T_X, T_Y$ 와  $T_Z$ 는  $f$ 에 대해서 상대적인 값이기 때문에  $f, T_X, T_Y$ 와  $T_Z$ 에 관한 정확한 실제 값을 알 수가 없다. 즉, 프레임내의 각 화소에 대하여  $Z$ 에 대한 전역 변수의 비율인  $fT_X/Z, fT_Y/Z, T_Z/Z$ 만을 산출할 수 있다. 영상복원 연구에서는 식의 단순화를 위해서 초점 거리  $f$ 의 값을 일반적으로 1로 가정하는 경우가 많다. 식 (13)으로부터 서로 다른 화소간의 상대적인 깊이의 비를 알 수 있다. 식 (13)을 재정리하여 다음의 식 (18)을 구하였다.

$$\frac{1}{v_x} \left( -\frac{fT_X}{Z} + x \frac{T_Z}{Z} \right) + \frac{T_Z}{Z} = \frac{1}{v_y} \left( -\frac{fT_Y}{Z} + y \frac{T_Z}{Z} \right) + \frac{T_Z}{Z} = 1 \quad (18)$$

### 7. 상대적 깊이 정보 추출

식 (18)에서 각 화소를 상대적인 깊이의 비로 표현할 수가

있었다.  $r_x \equiv (fT_X/Z), r_y \equiv (fT_Y/Z), r_z \equiv (T_Z/Z)$ 를 각각 정의하면, 카메라의 움직임이  $x, y$ 축상을 이동하는 경우에 대하여 다음 식 (19)와 같이 유도할 수 있다.

$$\begin{aligned} \frac{v_y}{v_x} &= \frac{r_y}{r_x} = \frac{T_Y}{T_X} \\ \frac{Z}{fT_X} &\equiv \frac{1}{r_x} = -\frac{1}{v_x} \\ \frac{Z}{fT_Y} &\equiv \frac{1}{r_y} = -\frac{1}{v_y} \end{aligned} \quad (19)$$

즉,  $x$ 와  $y$ 축 방향의 움직임이 크면 클수록 깊이 정보는 낮은 값을 나타낸다. 이는 움직임 벡터 값이 크면 객체가 가까이 있고, 작을수록 멀리 있는 객체임을 의미한다.

$r_z \neq 0$ 의 경우, 즉  $(v_y/v_x) \neq$  상수인 경우에서,

$$\alpha \equiv \frac{r_x}{r_z} = \frac{fT_X}{T_Z} \quad \beta \equiv \frac{r_y}{r_z} = \frac{fT_Y}{T_Z}$$

를 각각 정의하면,  $\alpha$ 와  $\beta$ 는 동일한 프레임내에서는 일정한 값을 가지며, 식 (18)에서 재정리하면 다음 식 (20)이 성립한다.

$$\frac{1}{v_x}(-\alpha + x) + 1 = \frac{1}{v_y}(-\beta + y) + 1 = \frac{1}{r_z} \quad (20)$$

위의 식 (20)과  $\alpha$ 와  $\beta$ 의 정의로부터 각  $x, y, z$  방향으로 카메라의 이동에 대하여 상대적인 깊이 값을 표현하는 식 (21)을 유도하였다.

$$\begin{aligned} \frac{Z}{T_Z} &\equiv \frac{1}{r_z} = \frac{x + v_x - \alpha}{v_x} = \frac{y + v_y - \beta}{v_y} \\ \frac{Z}{fT_X} &= \frac{1}{r_x} = \frac{1}{\alpha r_z} = \frac{x + v_x - \alpha}{\alpha v_x} \\ \frac{Z}{fT_Y} &= \frac{1}{r_y} = \frac{1}{\beta r_z} = \frac{y + v_y - \beta}{\beta v_y} \end{aligned} \quad (21)$$

단,  $v_x, v_y \neq 0$ 이다.

식 (21)은  $v_x \neq 0, v_y \neq 0$  조건에서 만족한다. 움직임 벡터가  $v_x = 0, v_y = 0$ 인 경우는 방정식의 분모 값이 불능의 형태가 되므로  $v_x = 0, v_y = 0$ , 그리고  $v_{xj}v_{yi} = v_{xi}v_{yj}$ 의 경우에 대하여 방정식을 각각 정의하였다.

$$\begin{aligned} \frac{Z}{fT_Y} &= -\frac{1}{v_y} \\ \frac{Z}{fT_X} &= -\frac{1}{v_x} \\ \frac{v_y}{v_x} &= \frac{r_y}{r_x} = \frac{T_Y}{T_X} \end{aligned} \quad (22)$$

프레임내의 모든 움직임은  $\frac{Z}{T_z}, \frac{Z}{fT_x}$ 와  $\frac{Z}{fT_y}$ 로 구성된 방정식으로 표현될 수 있으며,  $f, T_x, T_y$ 와  $T_z$ 에 따라 그 값

이 결정된다. 또한, 모두 상대적인 값이므로, 깊이에 대한 기준을 프레임 전체의 평균 깊이  $Z_{avg}$ 로 정한다. 본 논문의 최종 목표는 영상 프레임내의 각 영역에 대한 상대적인 깊이를 구하는 것이다. 마지막으로, 한 프레임내의 평균 깊이를 식 (23)과 같이 구하였다. 단,  $Z_{avg}$ 는 각 영역의 깊이  $Z$ 의 평균 값이고,  $N$ 은 각 영역 블록의 총 개수이다.

$$\begin{aligned} \frac{Z_{avg}}{fT_X} &= \frac{1}{fT_X} \frac{\sum Z}{N} = \frac{\sum \frac{Z}{fT_X}}{N} \\ \frac{Z_{avg}}{fT_Y} &= \frac{1}{fT_Y} \frac{\sum Z}{N} = \frac{\sum \frac{Z}{fT_Y}}{N} \quad (23) \\ \frac{Z_{avg}}{fT_Z} &= \frac{1}{fT_Z} \frac{\sum Z}{N} = \frac{\sum \frac{Z}{fT_Z}}{N} \end{aligned}$$

프레임내의 상대적인 깊이는 다음과 같이 정의하였다.

$$Z_{rel} = \frac{\frac{Z}{fT_X}}{\frac{Z_{avg}}{fT_X}} = \frac{\frac{Z}{fT_Y}}{\frac{Z_{avg}}{fT_Y}} = \frac{\frac{Z}{fT_Z}}{\frac{Z_{avg}}{fT_Z}} = \frac{Z}{Z_{avg}} \quad (24)$$

만일 영역이 배경에 속하면,  $Z_{rel}$ 에 관한 세 방정식은 반드시 등식이 성립하고 동일한 양수 값을 가져야 한다. 그렇지 않으면, 세 방정식의 등식은 성립이 되지 않고 각 영역은 다른 영역에 대하여 서로 독립적이며 전경에 속하는 영역으로 간주된다.

### 8. 모의 실험 및 결과 분석

카메라 초점에서 객체까지의 깊이 정보는 카메라 이동에 대하여 종속적이라는 사실에 기반을 두고, 하드웨어적 구현을 용이하게 하기 위해서 영상 프레임은 전경과 배경으로 구성되어 있으며, 배경은 카메라 움직임에 종속적이고 전경은 독립적이라는 것을 가정한다.

카메라 이동이 있는 단안 영상 시퀀스로부터 깊이 정보를 추출하는 알고리즘을 모의 실험하기 위해서 flower garden 영상을 사용하였다. 이 실험은 한 프레임내의 전체 평균 깊이 값에 대하여 각 매크로블록의 상대적인 깊이 정보를 추출한다. 깊이 정보를 추출하기 위한 첫 단계는 움직임 벡터를 구하는 것이다. 움직임 벡터 값은 시간적인 오차를 줄이기 위해서 바로 이전 프레임으로부터 구하였다. (그림 7)은 flower garden 프레임의 원영상 #1이고, (그림 8)은 flower garden 프레임의 원영상 #2이며, 1/30초 후의 프레임이다. (그림 9)와 (그림 10)은 제안한 깊이 추출 알고리즘을 수행하여 산출한 영상 시퀀스의 깊이 정보를 깊이 정보에 따라 색상을 부여하여 표시한 결과이다. 카메라가 왼쪽에서 오른쪽으로 이동하면서 획득

된 영상 시퀀스는 원영상 #1에서 원영상 #2로 카메라가 이동하였을 경우에 해당되며, 상대적인 깊이 값에 따라 다른 색상으로 깊이 값을 (그림 9)로 나타내었다. (그림 10)은 프레임 #2에서 프레임 #1로 카메라가 이동하였을 경우에 해당되며, 카메라가 오른쪽에서 왼쪽으로 이동하는 경우에 획득되는 영상 시퀀스로 간주할 수 있다. (그림 9)와 (그림 10)은 프레임 내부의 전체 프레임의 평균 깊이에 대하여 각 매크로블록 영역의 상대적인 깊이 정보를 표현한 그림이다. 깊이 값의 최대 값은 18로 제한하였고, 깊이의 값을 6단계로 축소하여 표현하였다. (그림 9)와 (그림 10)을 보면, 대체로 원영상의 깊이 정보를 정확히 표현하였고, 카메라가 좌우 이동을 하면서 획득된 영상에 대하여 인간의 깊이 인식과 유사한 결과를 나타내었다. (그림 10)의 좌상 부분의 깊이 정보가 원영상과는 달리 상당히 균일하지 않게 표현되었는데, 이것은 전역 움직임 탐색의 블록 정합 알고리즘의 결과가 정확하지 않기 때문이다. 영상의 복잡도가 낮은 영역에서 완벽한 움직임 추정 불가능하다. 그래서, 그 부분의 깊이 정보 값이 정확하게 표현되지 않은 결과이다. 일반적으로 영상의 공간 복잡도가 상당히 높거나, 아주 낮은 경우에는 영상 압축을 목적으로 하는 움직임 탐색에는 성공할 확률이 높을지라도 실제의 정확한 움직임을 나타낼 확률은 상당히 떨어진다. 또한, 나무의 가지와 같이 아주 작은 객체를 블록 단위의 움직임 추정을 하였기 때문에, 움직임 벡터의 결과에 정확성이 낮다.

본 논문에서는 전경에 속하는 영역의 독립적인 움직임에 대하여서는 깊이 정보의 산출이 어렵다. 그러므로, 다음의 각 경우에 해당되지 않는 영역은 전경에 속하는 것으로 간주하였다.

- 1)  $T_X = 0, T_Y = 0, T_Z = 0$
- 2)  $T_X = 0, T_Y \neq 0, T_Z = 0$
- 3)  $T_X \neq 0, T_Y = 0, T_Z = 0$
- 4)  $T_X \neq 0, T_Y \neq 0, T_Z = 0, \frac{v_y}{v_x} = \frac{r_y}{r_x} = \frac{T_Y}{T_X}$
- 5)  $T_Z \neq 0, \alpha \equiv \frac{r_x}{r_z} = \frac{fT_X}{T_Z}, \beta \equiv \frac{r_y}{r_z} = \frac{fT_Y}{T_Z}$

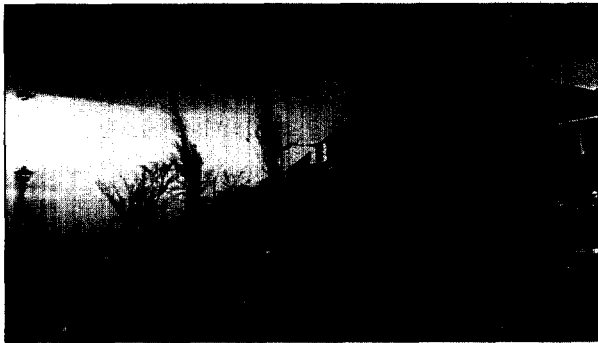
스테레오 영상을 이용하여 깊이를 추정하는 알고리즘은 두대의 카메라에서 획득한 영상이 필요하고, 두 카메라 사이의 거리 정보와 초점 거리에 관한 정보가 반드시 필요하지만, 본 논문에서 제안한 깊이 정보 추출 알고리즘은 스테레오 영상을 사용하지 않고 단안 영상 시퀀스를 사용하였으며, 어떠한 사전 정보도 필요하지 않다는 것이 장점이며, 기존에 제작되어 있는 많은 단안 영상들을 입체 영상으로 변환하는데 적용할 수 있다.

### 9. 결론 및 연구 과제

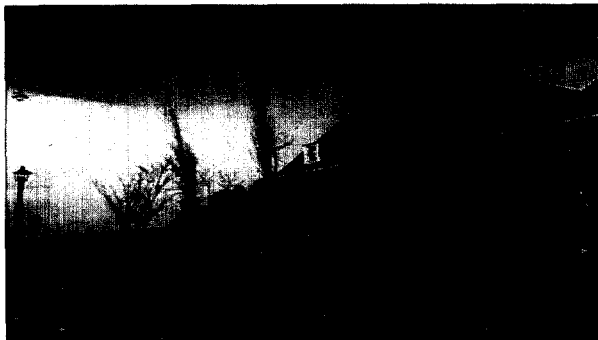
본 논문에서는 카메라 움직임이 있는 영상을 분석하여 카

메라 회전각에 의한 객체의 움직임을 보상하고, 움직임 벡터를 이용하여 프레임의 평균 깊이를 산출하여, 각 매크로블록에 대한 상대적 깊이를 구하였다. 또한 실시간 하드웨어 구현을 위하여 알고리즘의 처리 과정을 단순화하였다. 모의실험 결과를 보면, 배경에 속하면서 근경 혹은 원경에 포함되는 영역의 상대적인 깊이 값은 인간의 시각 체계가 인식하는 상대적인 깊이와 일치한다는 것을 보여주었다.

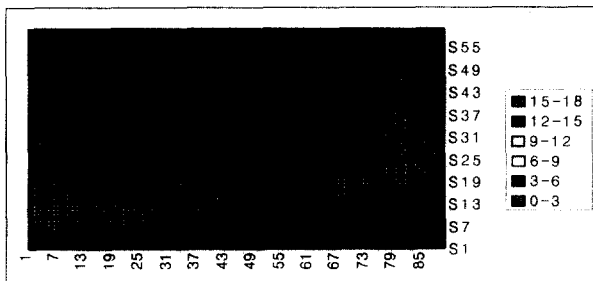
실험 결과는 기존의 단안 카메라에 의해서 촬영된 2차원 영상을 3차원으로 복원하는 입체 영상 시스템의 제작분야에 적용할 수 있다. 향후 연구 과제로서, 보다 정확한 움직임 벡터를 이용하기 위하여 공간 분할을 이용한 영역 분할 과정을 수행하여 객체를 추출한 다음, 동일한 분할영역에는 동일한 움직임 벡터 정보를 적용하여 객체의 균일한 내부 영역에 대한 깊이 정보와 영역간의 경계 부근에서 보다 정확한 깊이 정보를 추출하는 알고리즘의 개발이 요구된다.



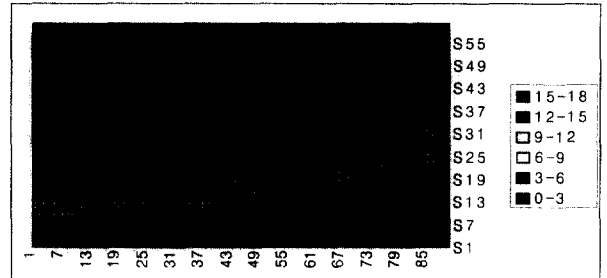
(그림 7) Flower garden 프레임 #1



(그림 8) Flower garden 프레임 #2



(그림 9) Flower garden #1 → #2의 깊이 정보 추출



(그림 10) Flower garden #2 → #1의 깊이 정보 추출

### 참 고 문 헌

- [1] R. Koch, "Model-Based 3D Scene Analysis from Stereoscopic Image Sequences," ISPRS '92, Vol.29, Part B5, Washington, pp.427-437, Oct. 1992.
- [2] Niem, W. Buschmann, "Automatic Modeling of 3D Natural Objects from Multiple Views," European Workshop on Combined Real And Synthetic Image Processing For Broadcast And Video Productions, pp.23-24. Nov. 1994, Hanburg, Germany.
- [3] Tony Jebara, Ali Azarbajejani, Alex Pentland, '3D Structure from 2D Motion', MIT Media Lab., Cambridge MA 02139, 1999.
- [4] A. Murat Tekalp, 'Digital Video Processing', Prentice Hall PTR Upper Saddle River, NJ07458, pp.19-34, 1995.
- [5] J. D. Foley and A. Van Dam, 'Fundamentals of Interactive Computer Graphics, Reading', MA : Addison-Wesley, 1983.
- [6] Dae-Seong Jeoune, Zi-Yeon Song, Joon-Ho Park, Young-Min Park and Young-Woo Yoon, "Global Motion Estimation by Analyzing the Panning/Tilting Angles of Camera," CISST2000, Vol.I, pp.163-170, Jun. 2000.
- [7] Milan Sonka, Vaclav Hlavac, Roger Boyle, 'Image Processing, Analysis, and Machine Vision', PWS Publishing, 1999.
- [8] D. Marr and T. Poggio, "Cooperative Computation Of Stereo Disparity," Science, 194 : pp.282-287, Oct. 1976.
- [9] E. Thoompson, "The Projective Theory Of Relative Orientation," Photogrammetria, 23(1) : pp.67-75, 1968.
- [10] O. Faugeras. "What Can Be Seen In Three Dimensions From An Uncalibrated Stereo Rig?," In Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy, Springer-Verlag, pp.563-578, 1992.
- [11] R. Hartley, R. Gupta, and T. Chang, "Stereo From Uncalibrated Cameras," In Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, pp.761-764, 1992.
- [12] Borko Furht, Joshua Greenberg, Raymond Wet-water, 'Motion Estimation Algorithms For Video Compression', KAP, 1997.
- [13] J. H. Park, M. C. Ryoo, J. H. Kim, S. H. Kim, "Motion Estimation Method Using Hierarchical Search Point Displacement," ITRC, KyungWoon University, Vol.1, No.1, pp. 66-70, Feb. 1999.
- [14] Jae-Woong Yi, Jun-Ho Oh, "Estimation Of Depth And 3D

Motion Parameter Of Moving Object With Multiple Stereo Images," Image and Vision Computing 14, pp.501-516, Oct. 1996.

- [15] 이제도, 박준호, 전대성, 윤영우, 김상곤, "동영상의 시간적 블록기반 영상분할 알고리즘", 한국정보처리학회논문지, 제7권 제5호, pp.1587-1598, 2000.
- [16] 전대성, 박준호, "전역 움직임 추정을 위한 카메라 회전각과 확대율 분석", 경운대학교 산업기술연구소논문지, 제3권 제1호(B), pp.329-338, 2000.



### 박 준 호

e-mail jhpark@kyungwoon.ac.kr

1990년 영남대학교 컴퓨터공학과 졸업  
(공학사)

1994년 영남대학교 대학원 컴퓨터공학과  
졸업(공학석사)

1999년 영남대학교 대학원 컴퓨터공학과  
박사과정 수료

1994년~1997년 현대전자 멀티미디어 연구소  
1999년~현재 경운대학교 컴퓨터공학과 전임강사  
관심분야 : MPEG, 영상처리, 병렬처리 시스템



### 전 대 성

e-mail : dsjeoune@tfc.ac.kr

1991년 영남대학교 컴퓨터공학과 졸업  
(공학사)

1996년 영남대학교 대학원 컴퓨터공학과  
졸업(공학석사)

1999년 영남대학교 대학원 컴퓨터공학과  
박사과정 수료

1996년~1999년 경동정보대학 겸임교수

1999년~현재 대구미래대학 멀티미디어정보과학과 전임강사  
관심분야 : 동영상처리, 영상압축, 데이터통신, 컴퓨터구조 등



### 윤 영 우

e-mail : ywyooun@ynucc.yu.ac.kr

1972년 영남대학교 전자공학과 졸업(학사)

1974년 영남대학교 대학원 전자공학과  
졸업(석사)

1984년 영남대학교 대학원 전자공학과 박사

1988년~현재 영남대학교 컴퓨터공학과 교수

관심분야 : 컴퓨터비전, 컴퓨터구조, 멀티미디어시스템