

# 단안영상에서 움직임 벡터를 이용한 영역의 깊이추정

## A Region Depth Estimation Algorithm using Motion Vector from Monocular Video Sequence

손정만\*, 박영민\*\*, 윤영우\*\*\*

Joung Man Sohn\*, Young Min Park\*\*, Young Woo Yoon\*\*\*

### 요약

2차원 이미지로부터 3차원 이미지 복원은 각 픽셀까지의 깊이 정보가 필요하고, 3차원 모델의 복원에 관한 일반적인 수작업은 많은 시간과 비용이 소모된다. 본 논문의 목표는 카메라가 이동하는 중에, 획득된 단안 영상에서 영역의 상대적인 깊이 정보를 추출하는 것이다. 카메라 이동에 의한 영상의 모든 점들의 움직임은 깊이 정보에 종속적이라는 사실에 기반을 두고 있다. 전역 탐색 기법을 사용하여 획득한 움직임 벡터에서 카메라 회전과 배율에 관해서 보상을 한다. 움직임 벡터를 분석하여 평균 깊이를 측정하고, 평균 깊이에 대한 각 영역의 상대적 깊이를 구하였다. 실험결과 영역의 상대적인 깊이는 인간이 인식하는 상대적인 깊이와 일치한다는 것을 보였다.

### Abstract

The recovering 3D image from 2D requires the depth information for each picture element. The manual creation of those 3D models is time consuming and expensive. The goal in this paper is to estimate the relative depth information of every region from single view image with camera translation. The paper is based on the fact that the motion of every point within image which taken from camera translation depends on the depth. Motion vector using full-search motion estimation is compensated for camera rotation and zooming. We have developed a framework that estimates the average frame depth by analyzing motion vector and then calculates relative depth of region to average frame depth. Simulation results show that the depth of region belongs to a near or far object is consistent accord with relative depth that man recognizes.

**Key Words** : Depth Estimation, Motion Estimation, Global Motion Estimation

### I. 서론

정보처리 기술이 비약적으로 발달한 현대 정보화 사회에서 효과적인 정보전달을 위한 수단으로 음성과 시각을 동시에 이용하는 동영상의 위치는 더욱 더 중요하게 부각되고 있다[1-3]. 동영상을 처리하는데 있어서도 보다 사실적이고 실세계에 가까운 입체적인 형태로 변화시켜 처리하고자 하는 연구가 전 세계적으로 진행되고 있으며

이런 입체적인 형태를 표현하는 것도 기존의 2차원적인 영상 표현 방법에서 탈피하여 실세계에 보다 가깝게 느껴지는 3차원적인 영상의 중요성이 더욱 증가하고 있다 [4-5]. 3차원 영상을 이용하면 다양한 멀티미디어 정보를 거리와 시간과 공간을 초월하여 효율적으로 전달할 수 있으며, 여행, 쇼핑 등을 실제로 현장에 가지 않고도 체험할 수도 있다.[6].

이러한 3차원 영상으로 구현하는 기법으로는 여러 개의 카메라를 사용하여 이미지를 획득하고, 획득한 이미지에서 두 점의 상관관계를 이용하여 깊이 정보를 추출하는 스테레오(stereo)영상 모델과 3차원 디지털이저를 이용하여 실제물체의 3차원 정보를 레이저 장비로 직접 획득

\*강릉영동대학 디지털미디어계열

\*\*경운대학교 멀티미디어공학부 \*\*\*영남대학교 전자정보공학부

접수 일자 : 2003. 10. 15      수정 완료 : 2004. 4. 03

논문 번호 : 2003-4-19

득하는 3차원 레이저 모델이 있다. 또한, 현재 대부분의 TV, 영화와 같은 영상물들은 단안(monocular)카메라를 사용하여 만들어진 단안 영상들인데, 이런 단안 카메라를 사용하여 나온 단안 영상들에서 기하학적 구조부터 깊이를 추정하는 단안 영상 모델이 있다[7].

단안 영상에서 깊이 정보를 추정할 수 있다면 기존의 TV, 영화와 같은 2차원 영상물을 3차원 영상으로 복원할 수 있다[7]. 하지만, 실세계의 물체가 2차원 평면영상으로 원근 투영되면서 많은 정보가 손실되기 때문에 기존의 2차원 평면영상을 이용하여 3차원 영상으로 완벽하게 복원하는 것은 여전히 미해결 과제로 남아 있다[8].

본 논문은 단안 카메라로 촬영한 단안 영상에 대하여 움직임 벡터를 추정하여 카메라 움직임 성분 중에서 회전과 배율의 영향을 보정한 후, 이동에 따른 움직임 벡터의 상관관계를 이용하여 프레임 내부의 전체 평균 깊이를 구하고, 각 영역간의 상대적인 깊이 정보를 추정한다.

본 논문은 II장에서는 관련 연구를 소개하고, III장에서 제안하는 영역의 깊이 정보 추출기법에 대해서 기술하고, IV장에서 실험 및 결과 분석, 마지막으로 V장에서 결론과 향후연구를 기술한다.

## II. 관련 연구

본 논문은 3차원 구현 기법에 기초를 두고 있으므로 3차원 영상 구현을 위한 기존의 기법들에 대해서 살펴본다. 또한 카메라 움직임과 영역의 움직임에 대해서 알아보고, 두 움직임간의 상관관계에 대해서도 살펴본다.

### 2.1 3차원 구현 기법

우리가 살아가는 실세계는 3차원으로 되어있다. 3차원 모델링에 관한 작업을 수작업으로 하는 것은 많은 시간과 노력이 소요되기 때문에 3차원 객체를 자동으로 재구성하는 여러 가지 기법들이 연구 되고 있으며 스테레오 영상 모델과 단안 영상 모델로 나눌 수 있다.

#### 2.2.1 스테레오 영상 모델

스테레오 영상 모델의 기본 개념은 인간이 두 눈으로 사물을 보는 원리이다. 인간의 시각은 위치가 다른 두 눈으로부터 다른 관점의 두 영상을 인식한다. 이렇게 인식한 두 영상간의 시각차이로부터 깊이 정보를 유도할 수가 있다[9]. 이 모델은 좌 우 카메라로 획득한 입력영상으로부터 두 영상간의 정합 정보인 변이를 예측하고, 생성하고자 하는 모델의 특징이 되는 부분을 추출하여, 추출

된 특징에 대해 예측된 변이 정보를 깊이 정보로 부여하여 최종적인 3차원 모델을 생성하는 기법이다[10].

스테레오 기하의 기본적인 형태를 살펴보면 그림 1과 같이 나타낼 수 있다.

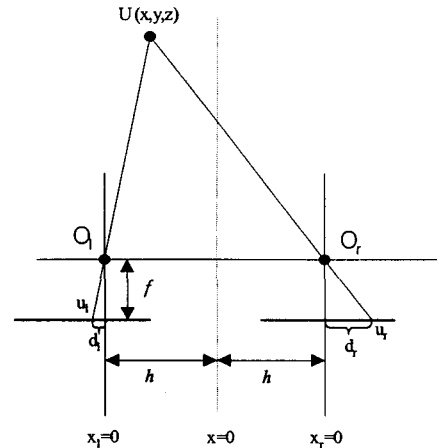


그림 1. 스테레오 기하 다이어그램  
Fig 1. Stereo geometry diagram

스테레오 기하 형태에서 두 대의 카메라에 대해서, 광축이 서로 평행하고, 상호간의 거리는  $2h$  라고 가정했을 때, 실세계 점  $U(x,y,z)$ 를 투영시킨 두 점을 각각  $u_l$ 과  $u_r$ 이라고 하고, 점까지의 거리가  $d_l$ 과  $d_r$ 일때, 카메라로부터의 거리  $z$ 를 나타내면 아래와 같이 나타낼 수 있다. 여기서  $f$ 는 카메라 초점 거리를 나타낸다.

$$z = \frac{2hf}{d_r - d_l} \quad (1)$$

#### 2.2.2 단안 영상 모델

단안 영상 모델을 살펴보면, 실세계의 한점  $U(X,Y,Z)$ 은 2차원 영상에서 한점  $u(x,y)$ 로 투영된다. 카메라가 움직인 후에, 다시  $U$ 는  $u'(x',y')$ 에 투영되게 되는데, 이 때 두 점  $u, u'$ 의 기하학적 구조로부터 깊이를 추정하는 기법이다[11]. 그림 2는 단안 영상 모델의 기본적인 형태를 나타내고 있다.

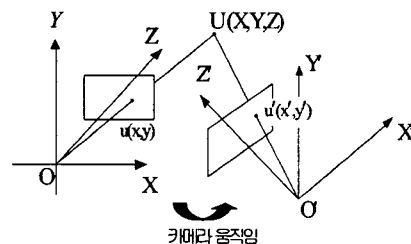


그림 2. 단안 영상 모델  
Fig 2. Monocular image model

단안 영상 모델에서 영상의 밝기 값의 강도를 이용하여 각 영역의 절대적 깊이 보다는 카메라 초점에서의 상대적인 깊이를 추정하는 방법이 많이 연구되고 있다. 어떤 한 특징점의 깊이 정보를 얻게되면 픽셀들간의 상관관계를 이용하여 다른 모든 점들의 깊이를 계산할 수 있다. 이렇게 형상정보를 재구성하는 방법을 "Shape from X"라 부르며 광류, 역초점, 윤곽등의 정보를 이용하여 영상프레임내의 형상을 재구성하는 다양한 기법들이 표 1에 나타나 있다.

표 1. 다양한 형상재구성 기법  
Table 1. Shape from X

X의 종류	광류	역초점	윤곽
시각정보의 근원	밝기 변화	영상의 흐린 영역	다양한 관점
사전 지식	카메라 이동거리	초점 거리	다중 카메라
용도	3차원영상의 표면특성 재구성	초점거리가 다른 영상의 깊이 추정	형상인식의 제약조건
단점	단안영상에 적용불가	초점거리가 긴 배경에 적용불가	윤곽이외의 정보이용불가

단안 영상 모델은 거리 영상으로부터 구조를 추출할 때 한 장의 입력 영상을 사용하여 모델을 생성하는 장점을 가지고 있다. 현재 대부분의 TV나 영화와 같은 단안 영상을 입력으로 하여 깊이 정보 추출 하게 되면, 3차원으로 복원하여 다양한 분야에 적용이 가능하다.

### 2.3 카메라 움직임과 영역의 움직임

영상에서 움직임은 3차원 실세계에서 관찰자인 인간에 의해서 전개되는 하나의 주요한 성질이고, 실세계에서의 움직임은 상당히 복잡하다. 객체의 관점에 따른 상대적 움직임, 관찰자에 대한 상대적인 이동과 회전, 그리고 움직이는 객체에 대한 관찰자의 상대적 이동과 같은 모든 정보는 형상(shape)과 깊이(depth)를 해결하는 실마리가 된다. 이런 영상에서 움직임을 분류해보면 카메라에 의한 움직임인 전역 움직임과 객체에 의한 움직임인 국부 움직임으로 분류된다. 움직이는 객체에 속하는 영역은 카메라의 움직임에 독립적인 움직임을 나타내지만, 움직이지 않는 객체에 속하는 영역은 카메라의 움직임에 따라 움직임의 변화량이 종속적이다.

카메라의 움직임에 대해서 살펴보면 그림 3과 같이 크게 회전(rotation)성분, 이동(translation) 성분, 그리고 배율(zooming) 성분으로 구분할 수 있다. 회전성분에는 X축 회전각(tilting,  $\theta$ ), Y축 회전각(panning,  $\psi$ ), Z축 회전각(rolling,  $\phi$ )으로 나눌 수 있다. 이동성분에는 수평이동(tracking,  $T_x$ ), 수직이동(booming,  $T_y$ ), 전후이동(dollying,  $T_z$ )이 있다.

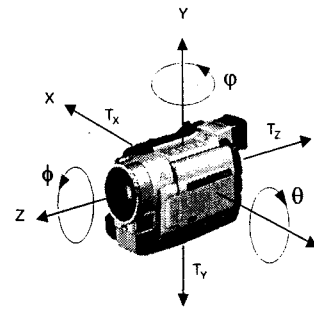


그림 3. 카메라 움직임 매개변수  
Fig 3. Several parameters for camera movement

그림 4에서 깊이 정보는 카메라의 이동에는 종속적이지만, 회전 및 배율과는 독립적이다. 카메라가 이동할 때는 움직임 추정의 정확도를 기대할 수 있으나, 회전과 배율 같은 비이동형 움직임에서는 정확한 움직임 추정이 어렵다[12].

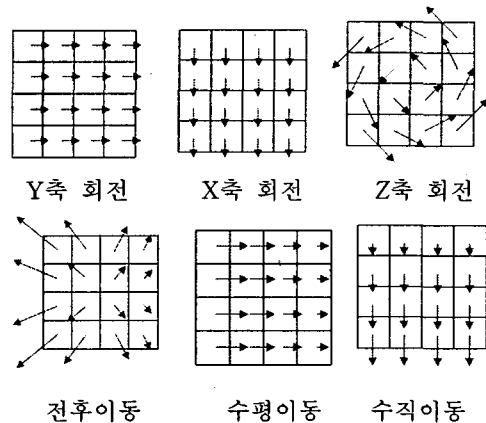


그림 4. 카메라 움직임에 대한 영상 효과  
Fig 4. A effects on an image for camera movement

카메라가 일정한 움직임을 가지고 이동할 때 프레임 내부의 각 영역에 대하여 깊이 정보가 일정하지 않기 때문에 영상 프레임 내부의 각 영역들의 움직임은 일정하지 않다.

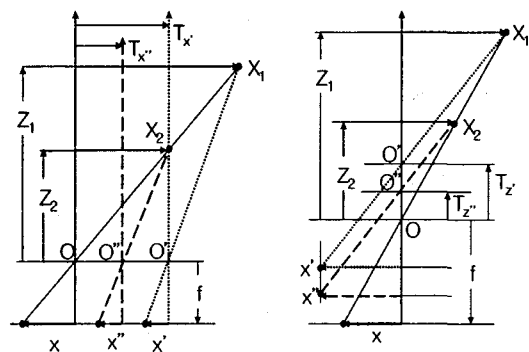


그림 5. 카메라 이동과 벡터  
Fig 5. Camera movement and it's vector

그림 5에서 실세계의 점  $X_1, X_2$ 는 각각 깊이( $Z_1, Z_2$ )가 각각 다른 점들이지만, 영상에서는 같은 위치에 맺히게 된다. 이 점들이 영상에서 같은 이동 벡터량( $x', x''$ )를 보이기 위해서는 카메라가 깊이에 따라 얼마나 이동( $T_{x'}, T_{x''}$ )해야 하는지를 보여준다. 좌측그림은 수평이동일 때이고, 우측그림은 전후이동일 때이고, O는 카메라 초점이다.

### III. 영역의 상대적인 깊이 추정 알고리즘

현재 TV, 영화와 같은 많은 영상들은 단안 카메라를 사용하여 만들어진 단안 영상이다. 이러한 단안 영상을 3차원 구조 복원을 하기 위해서는 깊이 정보의 추출이 필수적이다. 본 알고리즘의 목적은 카메라가 이동하면서 촬영을 한 단안 영상의 각 영역에 대한 상대적인 깊이를 추출하는 것이다.

그림 6은 본 알고리즘의 전체 흐름도를 나타내고 있다. 전체 흐름도에서 나타낸 것과 같이 제안한 알고리즘은 카메라로부터 획득된 2차원 단안 영상에 대하여 총 세단계로 나누어서 처리를 한다. 1단계에서는 먼저 잡음(noise)을 없애고 경계선의 보존성도 우수한 중간값 필터(Median filter)를 사용하여 잡음을 제거하고 전역 탐색 기법을 사용하여 움직임 벡터를 구한다. 전역 탐색 기법은 국부최소(Local Minima)에 빠질 위험이 가장 적고, 정확한 움직임 벡터를 구할 수 있다는 장점이 있다[13]. 영상의 깊이 정보는 카메라의 이동성분에는 종속적이지만, 회전과 배율에는 독립적이므로 본 알고리즘에서는 카메라의 회전과 배율성분을 보상하여 처리한다.

본 논문에서 구현한 알고리즘은 유리식으로 구성되어 있다. 분모의 값이 0을 가지는 경우에는 깊이 정보 추출이 불가능하게 되므로 이런 특이점에 대한 처리가 필요하다. 따라서 2단계에서는 보상된 움직임 벡터를 분석하여 특이점(움직임 벡터의 x성분  $v_x=0$  또는 y성분  $v_y=0$ )을 먼저 처리하고 일반적인 경우에 대해서는 카메라 이동성분 중에서 전후이동이 있는 경우와 없는 경우를 나누어서 처리한다. 전후이동이 없는 경우에는 영상 프레임 전체에 균등하게 영향을 주지만, 전후이동이 있는 경우에는 카메라의 광선(optical ray)을 중심으로 프레임 전체에 다르게 영향을 주기 때문에 분리하여 영역의 상대적인 깊이를 계산한다.

마지막 3단계에서는 프레임간의 관계를 처리하여 준다. 영상에서 같은 샷에서는 전후이동이 없는 경우와 있는 경우를 나누어서 샷의 처음 프레임에 대한 현재 프레임의 깊이변화 관계를 구하여 최종적인 영역의 상대적인 깊이정보를 구한다.

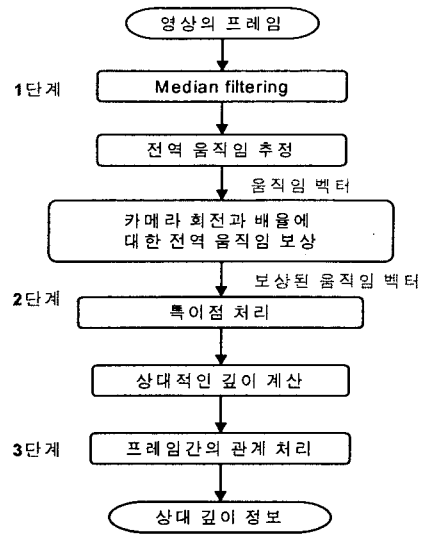


그림 6. 깊이 추정 알고리즘의 전체 흐름도

Fig 6. Block diagram of depth estimation algorithm

### 3.1 Median filtering

영상내의 잡음을 줄여서 전체 알고리즘의 성능을 높이기 위하여 본 논문에서는 중간값 필터를 사용하여 잡음을 제거한다. RGB 컬러 공간에서 중간값을 각각 구하는 방법은 실제 존재하는 중간값이 아니라 가상의 중간값이므로 실제 존재하는 중간값을 구하기 위해 그림 7과 같이 나타낼 수 있다.

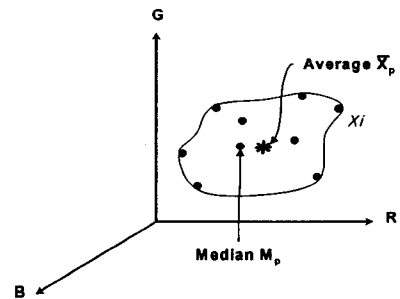


그림 7. RGB 공간에서의 중간값 필터

Fig 7. Median filter on RGB color space

먼저, 화소값  $X_i$ 와 점 P에 인접한 화소의 평균값  $\overline{X_P}$ 를 식 (2)과 같이 정의한다.

$$X_i = [R_i, G_i, B_i]^T \quad \overline{X_P} = [\overline{R_P}, \overline{G_P}, \overline{B_P}]^T \quad (2)$$

여기서,  $\overline{R_P}, \overline{G_P}, \overline{B_P}$ 는 각각 점 P에 인접한 Red, Green, Blue 화소의 평균값을 말하며 식 (3)와 같이 정의한다.

$$\begin{aligned} \overline{R}_p &= \frac{1}{n} \sum_{i=1}^n R_i, \quad \overline{G}_p = \frac{1}{n} \sum_{i=1}^n G_i \\ \overline{B}_p &= \frac{1}{n} \sum_{i \in N(p)} B_i \end{aligned} \quad (3)$$

$D_i$ 는 점 P에 인접한 화소와 평균값과의 거리를 나타내며, 식 (4)와 같이 계산량을 줄이기 위하여 근사치  $\overline{D}_i$ 를 사용한다.

$$D_i = |X_i - \overline{X}_p| = \sqrt{(R_i - \overline{R}_p)^2 + (G_i - \overline{G}_p)^2 + (B_i - \overline{B}_p)^2}$$

Let,

$$\overline{D}_i = |R_i - \overline{R}_p| + |G_i - \overline{G}_p| + |B_i - \overline{B}_p| \quad (4)$$

$$M_P = \underset{X_i}{\text{Min}} \{ \overline{D}_i \} \quad (5)$$

식 (5)는 점 p에 인접한 n개 화소의 평균값에 가장 가까운 화소값은 RGB 중간값과 유사하다는 것을 나타낸다.

### 3.2 카메라 이동과 움직임 벡터

일반적으로 영상은 많은 객체들로 구성되어 있다. 이들 객체들은 움직임이 있는 객체와 움직임이 없는 객체로 다시 나눌 수 있다. 객체들은 카메라 움직임에 상당히 종속적이며, 객체와 카메라 움직임간의 관계를 살펴보면 다음과 같이 나눌 수 있다. 정지 카메라와 정적인 객체, 정지 카메라와 동적인 객체, 움직이는 카메라와 정적인 객체 그리고, 움직이는 카메라와 동적인 객체의 관계로 표현된다[14]. 정지 카메라와 정적인 객체인 경우에는 움직임이 전혀 없으므로, 움직임 벡터의 값이 항상 0이다. 따라서 움직임 벡터를 이용해서 깊이 추정을 할 수가 없다. 정지 카메라와 동적인 객체의 경우는 움직임 벡터 값을 살펴보면 일정한 특성이 없이 나타나며, 객체의 움직임을 예측할 수가 없으므로 깊이 정보 추출을 할 수가 없다. 카메라가 일정한 방향으로 움직이고 객체가 정지 상태인 경우에는 카메라 초점으로부터 객체까지의 깊이 정보에 따라서 움직임 벡터가 일정한 특성을 나타낸다. 이런 경우에는 움직임 벡터의 특성을 분석하여 깊이 추정이 가능하다. 마지막으로, 카메라도 일정한 방향으로 움직이고, 객체도 움직이는 경우에는 움직임 벡터를 분석하여 카메라에 의한 움직임인지, 객체의 움직임인지를 알아서 움직이는 객체를 구해낼 수 있다. 다시 말해서, 카메라의 움직임이 없는 경우에는 움직임 벡터를 적용하여 깊이를 추정할 수가 없다.

단안 영상에는 카메라 이동과 회전뿐만 아니라 배율과 같은 다양한 카메라 움직임을 포함하고 있다. 그림 2에 나타난 단안 영상 모델에서 실세계의 한 점  $U$ 는 2차원 영상평면의 점  $u(x, y)$ 에 투영된 후, 카메라 움직임이 발생하여, 새로운 점  $u'(x', y')$ 에 투영되게 되는데 이것을 식으로 나타내면 다음과 같이 나타낼 수 있다.

$$SRU - T = U' \quad (6)$$

이 식에서  $S$ 는 카메라의 배율성분을,  $R$ 은 회전성분을 나타낸다. 그러면 이 식을 카메라에 대한 점  $U$ 의 상대적 위치벡터를  $[X, Y, Z]^T$ 라 하고 카메라가 움직인 후의 점  $U'$ 의 위치벡터를  $[X', Y', Z']^T$ 로 나타내고, 카메라의 이동 성분을  $T = [T_x, T_y, T_z]^T$ 로 나타내면 다음 식과 같이 된다.

$$SR[X \ Y \ Z]^T - [T_x \ T_y \ T_z]^T = [X' \ Y' \ Z']^T \quad (7)$$

식 (7)에서 카메라에 의한 회전과 배율성분을 보상하게

되면  $S = R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ 가 되고, 최종적으로 다음과 간단히 나타낼 수 있다.

$$[X \ Y \ Z]^T - [T_x \ T_y \ T_z]^T = [X' \ Y' \ Z']^T \quad (8)$$

설명 of 단순화를 위하여 X축만 고려하면, 실세계의 좌표  $X, Z$ 축과 영상에서의  $x$ 사이에는 그림 8과 같은 관계가 성립된다. 여기서 삼각형의 닮음비를 이용하여 관계를 나타내면 다음과 같다. ( $O$ 는 카메라 초점,  $f$ 는 초점거리)

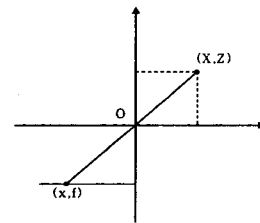


그림 8. 원근 투영 기하  
Fig 8. Perspective Projection

$$\frac{X}{Z} = \frac{x}{f}, \quad X = \frac{Z}{f} x$$

이런 관계를 적용하여 식 (8)을 나타내면 식 (9)와 같이 된다.

$$\frac{Z}{f}x - T_x = \frac{Z}{f}x' \quad (9)$$

이 식에서  $\frac{Z}{f}x' = \frac{Z - T_z}{f}(x + v_x)$ 이므로 적용하면 해서 나타내면 식 (10)과 같은 식이 나온다.

$$\frac{Z}{f} = \frac{1}{v_x} \left\{ \frac{T_z}{f}(x + v_x) - T_x \right\} \quad (10)$$

같은 방법으로 Y축으로 계산하면 다음과 같다.

$$\frac{Z}{f} = \frac{1}{v_y} \left\{ \frac{T_z}{f}(y + v_y) - T_y \right\} \quad (11)$$

### 3.3 카메라 이동 분석

앞에서 살펴본 카메라 이동과 움직임 벡터와의 관계에서 유도해낸 식 (10)과 (11)을 이용하면 아래 식을 얻을 수 있다.

$$f \frac{T_y}{T_z} = \frac{1}{v_x} (y v_x - x v_y + v_y f \frac{T_x}{T_z}) \quad (12)$$

식 (12)을 살펴보면  $f \frac{T_x}{T_z}$ ,  $f \frac{T_y}{T_z}$ 는 한 프레임에서는 일정하게 영향을 주는 전역변수들이다. 이 전역변수를 구하기 위하여 프레임 내의  $\forall p, q (p \neq q)$ 에 대해서 적용을 시키면 아래와 같은 수식을 얻어낼 수 있다. 여기서  $x_p, y_p$ 는  $p$ 의 좌표를 나타내며,  $v_x(p), v_y(p)$ 는  $p$ 의 움직임 벡터값을 의미한다.

$$f \frac{T_y}{T_z} = \frac{1}{v_x(p)} [y_p v_x(p) - x_p v_y(p) + v_y(p) f \frac{T_x}{T_z}] \quad (13)$$

$$= \frac{1}{v_x(q)} [y_q v_x(q) - x_q v_y(q) + v_y(q) f \frac{T_x}{T_z}] \quad (14)$$

$$f \frac{T_x}{T_z} = \frac{(y_q - y_p) v_x(p) v_x(q) + x_p v_x(q) v_y(p) - x_q v_x(p) v_y(q)}{v_x(q) v_y(p) - v_x(p) v_y(q)} \quad (15)$$

$$f \frac{T_y}{T_z} = \frac{(x_q - x_p) v_y(p) v_y(q) + y_p v_y(q) v_x(p) - y_q v_y(p) v_x(q)}{v_y(q) v_x(p) - v_y(p) v_x(q)} \quad (16)$$

위 두 식을 살펴보면, 만일  $T_x \neq 0 \wedge T_z = 0$  이면, 식 (15)는 무한대( $\infty$ )가 되고, 마찬가지로  $T_y \neq 0 \wedge T_z = 0$ 이면, 식 (16)은 무한대로 접근한다. 식 (15)와 (16)의 분자는 절대 무한대로 접근할 수 없지만, 분모는 무한대가 될 수 있다. 따라서, 아래와 같은 판별식이 나온다.

$$v_x(p) v_y(q) = v_x(q) v_y(p) \rightarrow T_z = 0 \quad (17)$$

프레임의 임의의 두 점이 판별식을 통과하면 전후이동이 없는 것으로 판단하게 된다.

### 3.4 수평-수직이동인 경우

전후이동이 없는 경우( $T_z = 0$ ) 즉, 판별식  $v_x(p) v_y(q) = v_x(q) v_y(p)$ 를 만족하는 경우, 위의 식 (10)와 식 (11)은  $T_z = 0$ 이므로, 다음과 같이 된다.

$$\text{식 (10)} \rightarrow \frac{Z(p)}{f} = - \frac{T_x}{v_x(p)} \quad (18)$$

$$\text{식 (11)} \rightarrow \frac{Z(p)}{f} = - \frac{T_y}{v_y(p)} \quad (19)$$

식 (18)과 (19)에서  $f$ 와  $T_x, T_y$ 는 한 프레임 안에서 일정한 상수이고,  $Z(p)$ 와  $v_x(p)$  및  $v_y(p)$ 는 점  $p$ 에 따라 변화하는 변수이다. 따라서 식 (18), (19)를 이용하여 한 프레임의 평균값이  $\bar{Z}$ 를 구하면 다음과 같다.

$$\text{식 (18)} \rightarrow \bar{Z} = \frac{1}{N} \sum_{p=1}^N Z(p) = -f T_x \frac{1}{N} \sum_{p=1}^N \frac{1}{v_x(p)} = -f T_x \overline{v_x^{-1}}$$

$$\text{식 (19)} \rightarrow \bar{Z} = \frac{1}{N} \sum_{p=1}^N Z(p) = -f T_y \frac{1}{N} \sum_{p=1}^N \frac{1}{v_y(p)} = -f T_y \overline{v_y^{-1}}$$

$$\therefore \bar{Z} = -f T_x \overline{v_x^{-1}} = -f T_y \overline{v_y^{-1}} \quad (20)$$

앞의 식 (18), (19)와 식 (20)을 이용하여 프레임의 평균값이  $\bar{Z}$ 에 대한 임의의 점  $p$ 의 상대적인 값이인  $z(p)$ 를 구하면 다음과 같다.

$$z(p) = \frac{Z(p)}{\bar{Z}} = \frac{1}{v_x(p) \overline{v_x^{-1}}} = \frac{1}{v_y(p) \overline{v_y^{-1}}} \quad (21)$$

이 식에서도 한 프레임 내에서  $\overline{v_x^{-1}}$ 와  $\overline{v_y^{-1}}$ 는 일정한 상수이고,  $z(p)$ 와  $v_x(p)$  및  $v_y(p)$ 는 점  $p$ 에 따라 변화하는 변수이다. 따라서 임의의 점  $p$ 가 아래의 조건을 만족하면  $p$ 의 상대적인 값이  $z(p)$ 는 다음과 같이 구할 수 있다.

$$\text{조건} : v_x(p) \overline{v_x^{-1}} = v_y(p) \overline{v_y^{-1}}$$

$$z(p) = \frac{1}{v_x(p) \overline{v_x^{-1}}} \quad (22)$$

위의 조건을 만족하지 않는 점  $p$ 는 움직이는 영역에 속한 점이다. 같은 샷에서의 프레임과 프레임 사이의 평균 값

이는 상관관계가 높다. 현재 프레임  $k$ 와 프레임  $k-1$ 사이의 평균 깊이는 카메라의 전후이동이 없으므로, 식 (23)와 같이 평균 깊이 변화가 발생하지 않는다.

$$\bar{Z}_k = \bar{Z}_{k-1} \quad (23)$$

### 3.5 일반적인 경우

전후이동이 있는 일반적인 경우 즉, 판별식  $v_x(p)v_y(q) \neq v_x(q)v_y(p)$ 의 관계가 성립된다. 이런 경우 파라메타  $\alpha, \beta$ 를 아래와 같이 정의한다.

$$\alpha \equiv f \frac{T_x}{T_z}, \quad \beta \equiv f \frac{T_y}{T_z}$$

새로 정의된 파라메타를 이용하여 앞의 식 (10), (11)을 나타내면 다음과 같이 나타낼 수 있다.

$$\text{식(10)} \rightarrow \frac{Z(p)}{T_z} = \frac{1}{v_x(p)} (x_p + v_x(p) - \alpha) \quad (24)$$

$$\text{식(11)} \rightarrow \frac{Z(p)}{T_z} = \frac{1}{v_y(p)} (y_p + v_y(p) - \beta) \quad (25)$$

파라메타  $\gamma$ 를 아래와 같이 정의하고, 평균깊이  $\bar{Z}$ 를 나타내면 다음과 같다.

$$\gamma \equiv \frac{1}{N} \sum_{p=1}^N \frac{Z(p)}{T_z} = \frac{\frac{1}{N} \sum_{p=1}^N Z(p)}{T_z} = \frac{\bar{Z}}{T_z}$$

$$\bar{Z} = \gamma T_z \quad (26)$$

식 (24), (25), (26)을 이용하여 프레임의 평균깊이  $\bar{Z}$ 에 대한 임의의 점  $p$ 의 상대적인 깊이인  $z(p)$ 를 구하면 다음과 같다.

$$z(p) = \frac{Z(p)}{\bar{Z}} = \frac{T_z \frac{(x_p + v_x(p) - \alpha)}{v_x(p)}}{\gamma T_z} = \frac{T_z \frac{(y_p + v_y(p) - \beta)}{v_y(p)}}{\gamma T_z}$$

$$= \frac{x_p + v_x(p) - \alpha}{\gamma v_x(p)} = \frac{y_p + v_y(p) - \beta}{\gamma v_y(p)} \quad (27)$$

식 (27)을 이용하여 프레임 내의 임의의 점  $p$ 가 다음의 조건을 만족한다면, 상대적인 깊이  $z(p)$ 를 아래와 같이 구해낼 수 있다. 반대로 조건에 만족하지 않는다면 점  $p$ 는 움직이는 영역에 속한 점이다.

조건 :

$$v_y(p)(x_p + v_x(p) - \alpha) = v_x(p)(y_p + v_y(p) - \beta)$$

$$z(p) = \frac{x_p + v_x(p) - \alpha}{\gamma v_x(p)} \quad (28)$$

같은 샷에서 현재 프레임  $k$ 와 프레임  $k-1$ 사이의 평균 깊이의 상관관계를 살펴보면, 전후이동이 있으므로, 다음과

같은 변화가 생긴다. 여기서,  $T_{zk}$ 는 프레임  $k$ 에서의 카메라 전후이동을 나타낸다.

$$\bar{Z}_{k-1} = \bar{Z}_k + T_{zk} = \bar{Z}_k \left(1 + \frac{T_{zk}}{\bar{Z}_k}\right) = \bar{Z}_k \left(1 + \frac{1}{\gamma_k}\right) \quad (29)$$

### 3.6 프레임간의 관계

영상에서 같은 샷에서는 샷의 처음프레임과 현재 프레임 사이에는 전후이동의 유무에 따라 평균깊이의 변화가 다른 관계를 가지게 된다. 전후이동이 없는 경우에는 앞의 식 (23)에서와 같이 깊이변화가 발생하지 않으므로, 다음과 같은 관계가 성립된다. 여기서  $d_k$ 는 샷의 처음 프레임의 평균 깊이에 대한 프레임  $k$ 의 평균 깊이의 비를 나타낸다.

$$d_k = \frac{\bar{Z}_k}{\bar{Z}_1} = \frac{\bar{Z}_{k-1}}{\bar{Z}_1} = d_{k-1} \quad (30)$$

전후이동이 있는 일반적인 경우에는 프레임 사이에 평균 깊이변화가 발생하게 된다. 앞의 식 (29)에서의 관계를 적용하면 다음과 같은 관계가 발생한다.

$$d_k = \frac{\bar{Z}_k}{\bar{Z}_1} = \frac{\bar{Z}_{k-1}}{\bar{Z}_1} \cdot \frac{\bar{Z}_k}{\bar{Z}_{k-1}} = d_{k-1} \cdot \frac{\bar{Z}_k}{\bar{Z}_{k-1}}$$

$$= d_{k-1} \cdot \frac{\gamma_k}{\gamma_{k-1} + 1} \quad (31)$$

프레임 사이의 관계로 적용해서 샷의 처음 프레임의 평균 깊이에 대한 현재 프레임  $k$ 의 점  $p$ 의 깊이 비 ( $z'(p)$ )를 구하면 다음과 같다.

$$z'(p) = \frac{Z_k(p)}{\bar{Z}_1} = \frac{d_k Z_k(p)}{\bar{Z}_k} = d_k z_k(p) \quad (32)$$

## IV. 실험결과 및 분석

본 논문의 카메라 이동이 있는 단안 영상에서 움직임 벡터를 이용하여 영역간의 상대적인 깊이 정보를 추출하는 알고리즘의 효용성을 검증하기 위해서 MPEG 표준영상인 flower garden 영상을 사용하였다. flower garden 영상을 설명하면 카메라의 수평이동만 있고, 원근감이 깊이에 따라 표현된 복잡한 자연 영상이다.

실험과정은 median 필터를 사용하여 잡음이 제거된 영상에서 전역 탐색 기법을 사용하여 움직임 벡터를 구한다. 카메라의 움직임 중에서 회전과 배율성분은 깊이정보와 독립적이므로 카메라 회전과 배율성분을 보상하여 주었다.

그림 9는 flower garden의 6-7번 프레임의 원영상이고, 그림 10은 두 프레임의 움직임 추정을 한 후 카메라의 회전과 배율 성분을 보상하고 나온 움직임 추정 결과

그림이다.

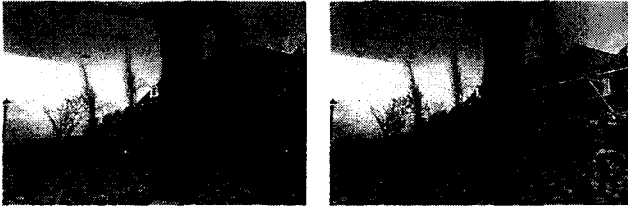


그림 9. Flower garden #6-#7  
Fig 9. Flower garden #6-#7

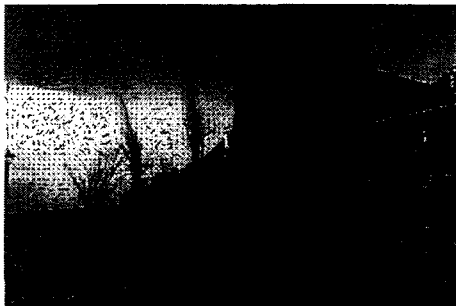


그림 10. flower garden 움직임 추정 결과  
Fig 10. A result of motion estimation for flower garden

그림 11은 움직임 벡터를 이용하여 상대적인 깊이 값에 따라 여러 가지 색상이 부여한 것이다. 깊이 값의 최대값은 20으로 제한을 하였고, 각 깊이의 값을 7단계로 축소하여 표현하였다. 그림 12는 상대적인 깊이 값을 인간이 인식하는 깊이를 유사한 방법으로 표현한 거리 영상의 그림이다. 원영상과는 달리 깊이 정보만을 표현하기 때문에, 프레임 내의 객체 정보는 존재하지 않는다.

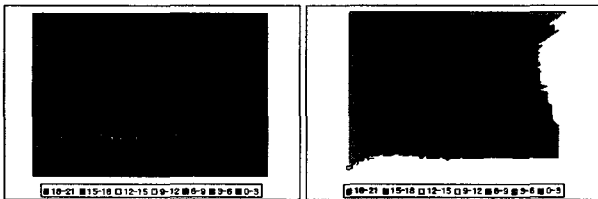


그림 11. 상대적인 깊이 정보    그림 12. 거리 영상  
Fig 11. Relative depth information    Fig 12. Depth image

그림 11과 그림 12를 보면, 대체로 원영상의 깊이 정보를 정확히 표현하고 있으며, 카메라가 수평 이동을 하면서 획득된 영상에 대하여 인간의 깊이인식과 유사한 결과를 보인다. 결과 영상의 좌측 상단 부분에 깊이 정보가 원영상과는 달리 상당히 복잡하게 나타나는 것은 움직임 추정의 결과가 정확하지 않아 영상의 복잡도가 낮거나 객체들간의 경계 영역에서 완벽한 움직임 추정이 불가능한 경우에 깊이 정보값이 정확하게 표현되지 않을 보여준다.

본 알고리즘의 효용성 검증하기 위해서 전후이동과 여러 가지 카메라 이동이 있는 일반적인 비디오 영상들

을 사용하였다.

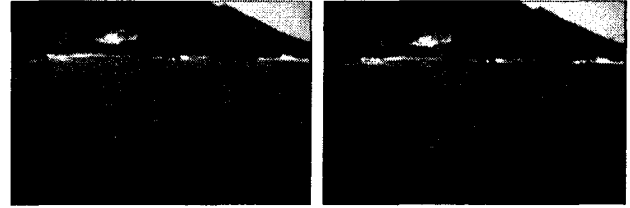


그림 13. National Geographic #8-#9  
Fig 13. National Geographic #8-#9

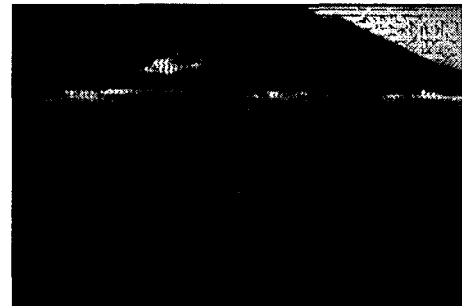


그림 14. 움직임 추정 결과  
Fig 14. A result of motion estimation

그림 15는 상대적인 깊이의 최대값을 30으로 제한하고, 각 깊이의 값을 6단계로 색상을 부여하여 표현하였다. 그림 16은 상대적인 깊이 값을 인간이 인식하는 깊이를 유사한 방법으로 표현한 거리 영상의 그림이며, 카메라의 이동이 있는 일반적인 영상에서도 깊이정보가 정확하게 표현되지만 바다부분의 움직임 추정에 오류가 발생한 것을 알 수 있다.

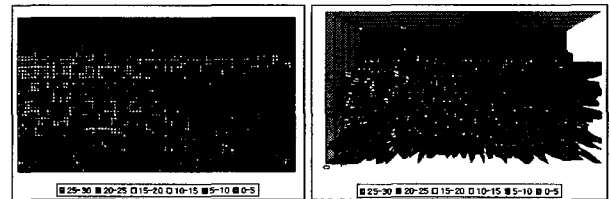


그림 15. 상대적인 깊이 정보    그림 16. 거리 영상  
Fig 15. Relative depth information    Fig 16. Depth image

그림 17은 사람을 뒤에서 따라가면서 찍은 장면이며, 카메라의 이동과 객체의 움직임이 모두 있는 일반적인 영상이다.



그림 17. National Geographic #1-#2  
Fig 17. National Geographic #1-#2



그림 18은 움직임 추정 결과영상이다.

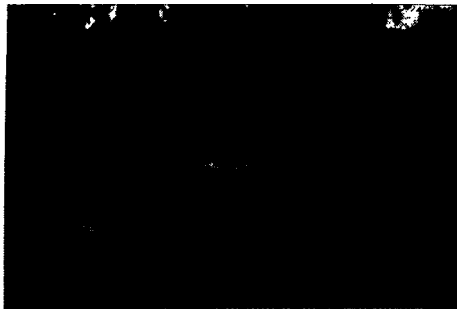


그림 18. 움직임 추정 결과  
Fig 18. A result of motion estimation

그림 19와 그림 20은 제한기준을 앞의 일반적인 경우와 동일하게 주었을 때 상대적인 깊이정보와 거리영상을 나타낸 것이다.

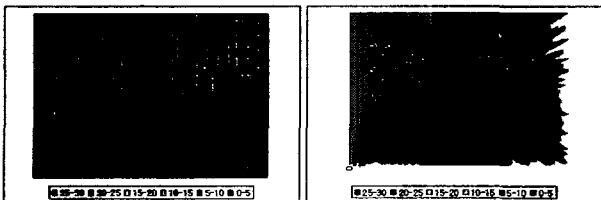


그림 19. 상대적인 깊이 정보    그림 20. 거리 영상  
Fig 19. Relative depth information    Fig 20. Depth image

그림 19와 그림 20을 보면 깊이정보도 정확히 표현되었으며, 영상의 가운데 움직이는 객체도 정확히 검출되었다.

### V. 결론 및 향후 연구

본 논문에서는 카메라 움직임이 있는 단안 영상을 이용하여 움직임 벡터를 산출하고, 산출된 움직임 벡터를 분석하여 영역의 상대적인 깊이를 구하는 알고리즘을 구현하였다. 상대적인 깊이를 구하기 위하여 사용되는 움직임 벡터 값은 시간적인 오차를 줄이기 위하여 바로 이전 프레임으로부터 구하였고, 실시간 처리가 가능하고 가장 정확한 움직임 벡터 값을 산출하는 전역 탐색 기법으로 구하였다. 산출된 움직임 벡터는 카메라의 모든 움직임 성분을 가지고 있는데, 카메라의 움직임 성분 중에서 회전과 배율 성분은 깊이 정보에는 독립적인 영향을 미치지 않으므로, 보상하여서 깊이 정보에 종속적인 카메라 이동 성분만 남겼다. 이렇게 보정한 순수한 카메라 이동 성분만 남은 움직임 벡터를 분석하여 프레임의 전체 평균 깊이를 산출하고, 전체 평균 깊이에 대한 각 영역의 상대적인 깊이를 구하였다.

제안 알고리즘을 실험하기 위해서 MPEG표준 영상인

flower garden 영상과 일반적인 비디오 영상을 사용하였다. 실험결과는 산출된 상대적인 깊이 정보에 따라 다양한 색상을 부여하여 나타내었고, 또한 상대적인 깊이 값을 실제 인간이 인식하는 거리 영상으로 표현하였다. 이렇게 나타낸 상대적인 깊이 값은 인간이 인식하는 상대적인 깊이와 일치한다는 것을 알 수 있었고, 움직이는 객체를 추출하는 것도 가능하였다. 실험결과에서 일부 깊이 정보가 부정확한 것은 잘못된 움직임 벡터 값에 의해서 깊이 정보가 잘못 나온 것인데, 그 이유는 영상의 복잡도가 낮거나 객체들 간의 경계 영역에서는 완벽한 움직임 추정이 안 될 수 있기 때문에 깊이 정보값이 부정확하게 표현되었다.

본 논문에서 제안하는 알고리즘은 계산량이 적고, 알고리즘이 간단하여 하드웨어 제작에 적용하기가 쉽다는 것이 특징이다. 기존의 깊이 추출 알고리즘과 비교하면, 스테레오 영상을 이용하여 깊이 추정 알고리즘에서는 두 카메라 사이의 거리 정보와 초점 거리에 관한 정보가 반드시 필요하지만, 본 논문에서 제안하는 알고리즘은 단안 영상에 대해서 사전에 어떠한 정보도 필요하지 않다는 것이 장점이라고 할 수 있다. 실험 결과는 기존의 단안 카메라에 의해서 촬영된 2차원 영상을 복원하여 3차원 영상으로 제작하는 분야에 적용할 수 있다.

향후 연구 과제로는 일부 깊이 정보가 부정확하게 나오는 오류를 해결하기 위하여 영역을 분할하여 객체를 추출한 후, 객체의 움직임을 분석하여 나온 움직임 벡터를 이용하여 객체의 상대적인 깊이 정보를 추출하는 알고리즘 개발이 요구된다.

### 참 고 문 헌

- [1] H.261 ITU-T Recommendation H.261, "Video codec for audiovisual services at p×64 kbit/s", Mar.1993.
- [2] MPEG-1 ISO/IEC 11172-2 "Information Technology-Coding of Moving Picture and Associated Audio for Digital Storage Media at up to about 1.5mbit/s : Part 2 Video", Aug.1993.
- [3] MPEG-2 ISO/IEC 13818-2, "Information Technology-Generic coding of Moving Picture and Associated Audio Information : Video", Darft, Mar. 1994.
- [4] R.Koch, "Model-Based 3D Scene Analysis from Stereoscopic Image Sequences," ISPRS '92, Vol.29, Part B5, Washington, pp.427-437, Oct. 1992.
- [5] 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.

- [6] Tony Jebara, Ali Azarbayejani, Alex Pentland, 3D Structure from 2D Motion, MIT Media Lab, Cambridge MA 02139, 1999.
- [7] 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.
- [8] Milan Sonka, Vaclav Hlavac, Roger Boyle, Image Processing, Analysis and Machine Vision, PWS Publishing, 1999.
- [9] E. Thoopson, "The Projective Theory Of Relative Orientation," Photogrammetria, 23(1): pp.67-75, 1968.
- [10] D. Marr and T. Poggio, "Cooperative Computation Of Stereo Disparity," Science, 194: pp.282-287, Oct. 1976.
- [11] R Mohr. *Projective Geometry And Computer Vision*, Handbook of Pattern Recognition and Computer Vision, World Scientific, Singapore, pp.369-393, 1993.
- [12] Joon-Ho Park, Dae-Seong Jeoune, Sang-Gon Kim, Young- Woo Yoon, "Estimation of the Depth Information from Single View Image Sequence with Camera Translation", CISST 2000, Vol. I, pp.179-183, June 2000.
- [13] 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 Rotation and Zooming Factor", CISST2000, Vol. I, pp. 163-169, June 2000.
- [14] Ramesh Jain, Rangachar Kasturi, Brian G. Schunch, Machine Vision, MIT Press and McGRAW-Hill, Inc., 1995.



손 정 만(Joung-Man Sohn)  
1985년 경북대학교 전자계산기과 졸업 (공학사)  
1990년 영남대학교 정보처리과(교육학석사)  
1997년 영남대학교 컴퓨터공학과(박사수료)  
1985년~1993년 구미전자공고

993년~현재 강릉영동대학 디지털미디어계열 조교수  
관심분야 : 영상처리, 컴퓨터구조



박 영 민(Young-Min Park)  
1993년 영남대학교 전산공학과(공학사)  
1995년 영남대학교 전산공학과(공학석사)  
2001년 영남대학교 컴퓨터공학과(박사수료)  
1995년~1999년 삼성전자(주) 정보통신연구소

2000년~현재 경운대학교 멀티미디어공학부 전임강사  
관심분야 : 컴퓨터비전, ASIC설계, 컴퓨터구조



윤 영 우(Young-Woo Yoon)  
1972년 영남대학교 전자공학과(공학사)  
1974년 영남대학교 전자공학과(공학석사)  
1984년 영남대학교 전자공학과(공학박사)  
1988년~현재 영남대학교 전자정보공학부 교수

관심분야 : 컴퓨터비전, 멀티미디어시스템, 컴퓨터구조