

일반논문-09-14-3-10

스테레오 카메라를 이용한 이동객체의 실시간 추적과 거리 측정 시스템

이동석^{a)‡}, 이동욱^{a)}, 김수동^{a)}, 김태준^{a)}, 유지상^{a)}

Real-time moving object tracking and distance measurement system using stereo camera

Dongseok Lee^{a)‡}, Dongwook Lee^{a)}, Sudong Kim^{a)}, Taejune Kim^{a)}, Jisang Yoo^{a)}

요 약

본 논문에서는 스테레오 카메라로부터 획득된 좌우 영상을 이용하여 3차원 공간좌표(x, y, z)를 획득하고, 이를 이용하여 제어 되는 가상공간을 통하여 사용자에게 현실감을 제공하는 실시간 시스템을 구현한다. 일반적으로 관심영역의 변이를 추정할 때 관심영역내의 모든 화소(pixel)의 변이를 추정하지만, 제안한 시스템에서는 관심영역의 2차원 중심좌표(x, y)만을 변이추정에 사용하여 실시간으로 변이를 추정한다. 추정된 변이로부터 깊이정보(depth)를 구하여 관심영역의 3차원 공간좌표를 획득한다. 시스템은 손을 관심영역으로 설정하여 실시간으로 손의 움직임 정보를 획득하고, 가상공간(virtual space)에 적용하여 사용자가 가상공간을 조작할 수 있도록 한다. 실험을 통해 제안하는 실시간 시스템이 150cm 거리(distance) 내에서의 깊이측정 시 0.68cm의 평균오차를 가지고 손동작 인식률은 90% 이상 보이는 것을 검증하였다.

Abstract

In this paper, we implement the real-time system which extracts 3-dimensional coordinates from right and left images captured by a stereo camera and provides users with reality through a virtual space operated by the 3-dimensional coordinates. In general, all pixels in correspondence region are compared for the disparity estimation. However, for a real time process, the central coordinates of the correspondence region are only used in the proposed algorithm. In the implemented system, 3D coordinates are obtained by using the depth information derived from the estimated disparity and we set user's hand as a region of interest(ROI). After user's hand is detected as the ROI, the system keeps tracking a hand's movement and generates a virtual space that is controlled by the hand. Experimental results show that the implemented system could estimate the disparity in real-time and gave the mean-error less than 0.68cm within a range of distance, 1.5m. Also It had more than 90% accuracy in the hand recognition.

Key Words : Stereo vision, Hand segmentation, Disparity estimation, Virtual space

a) 광운대학교

Electronic Engineering Dept. Kwangwoon University

‡ 교신저자 : 이동석(dlehdtr86@kw.ac.kr)

※ 본 연구는 지식경제부 및 정보통신연구진흥원의 IT산업원천기술개발사업 [2009-F-208-01, 대화형 디지털 홀로그램 통합서비스 시스템의 구현을 위한 신호처리 요소 기술 및 SoC 개발]과 서울시 산학연 협력사업 (NT080528)의 일환으로 수행하였음.

· 접수일(2009년3월4일), 수정일(2009년5월8일), 게재확정일(2009년5월26일)

1. 서론

이동하는 특정 물체를 실시간으로 인식하고 움직임을 추적(tracking)하는 기술은 오랫동안 컴퓨터 비전(computer vision) 분야에서 연구가 진행되어 왔다. 컴퓨터 비전은 궁극적으로 인간의 시각 시스템을 컴퓨터로 구현하려는 목표

를 가지고 있다. 인간의 시각 시스템에서 중요한 것은 3차원 공간에서 목표물을 인식하고 추적하는 것과 목표물까지의 거리를 측정하는 것이라고 할 수 있다. 특히 목표물까지의 거리를 측정함으로써 차후의 행동에 대한 의사 결정을 할 수 있으며, 이 같은 특성을 컴퓨터에 도입함에 있어 중요한 것이 목표물까지의 거리, 즉 깊이 정보임을 알 수 있다¹⁻³⁾.

깊이 정보를 획득하는 방법은 여러 가지가 있는데 초음파나 레이저센서 등을 이용한 능동적 방법은 주변 환경에 민감하게 반응하기 때문에 잡음이나 간섭에 의해 정확한 정보를 얻기 힘들다. 반면에 수동적 방법은 주변 환경의 변화에 상대적으로 간섭이 적어 잡음이 많은 환경에서 효과적으로 사용될 수 있다. 스테레오 비전(stereo vision)은 대표적인 수동적 방법 중 하나로 3차원 공간상에 설치된 스테레오 카메라로부터 얻어낸 좌우 영상에서 상호간의 정합점을 찾아내고 정합점(matching point)간의 차이를 이용하여 깊이 정보를 추출한다. 스테레오 비전은 인간의 시각 시스템을 컴퓨터로 구현하려는 컴퓨터 비전 분야에서도 매우 중요한 부분으로 현재 많은 연구가 진행되고 있으며 환경의 특별한 제약이 없고 일반적인 자연환경을 대상으로 할 수 있는 장점 때문에 여러 분야에 걸쳐 응용되고 있다⁴⁾.

영상의 정합은 스테레오 비전에서 3차원 공간상의 한 점이 두 영상에 투영되는 점의 대응 관계를 찾아내는 중요한 과정이다. 이 과정을 스테레오 정합(stereo matching)이라 부르며 두 영상에서 선택된 점들이 대응점을 가지는가를 판단하고 그 점에서의 깊이 정보의 지표가 되는 변이를 계산하는 것이 이 과정의 핵심이다.

스테레오 정합은 좌우 영상의 대응점을 찾는 작업을 통하여 단순화 될 수 있으며, 그 정합방법에 따라 영상에서 특징을 포함한 일정 영역의 영상조각을 가지고 다른 영상과의 상관관계를 이용하는 영역기반 정합(area-based matching), 화소의 밝기에 대한 미분 연산자를 이용하여 1차원 또는 2차원 특징 값을 얻어내고 명암대비의 평균값이나 경계의 방향 등을 특징으로 선택하여 정합에 사용하는 특징기반 정합(feature-based matching), 베이시안 추론을 적용한 에너지 함수를 이용하거나 정합을 정칙화(regularization) 문제로 표현하여 반복법으로 변이를 구하는 에너지기반 정합(energy-based matching) 정합으로 나뉜다. 이

중 영상에서 정합될 관심영역에 포함된 모든 화소에 대하여 다른 영상과의 상관관계를 구하는 영역기반 정합방법이 널리 사용된다. 이런 방법은 전체 변이를 한 번에 얻을 수 있는 반면 계산 복잡도가 높아 연산에 상당한 시간이 소요되는 단점이 있다^{5,6)}.

본 논문에서는 좌우 영상의 관심영역을 대표하는 특징점을 얻어내고 특징점만을 비교하여 변이추정(disparity estimation)에 필요한 대응영역을 줄여 속도를 향상시키는 기법을 제안한다. 정합에 필요한 위치를 선택하기 위해 사용하는 비용 함수는 SSD(sum of squared difference)를 이용한다. 또한 계산된 관심영역의 움직임은 3차원 좌표로 추적하여 사용자에게 제공한다. 손의 움직임에 따라서 동작하는 마우스 포인터와 3차원 좌표의 움직임을 추적하는 가상 공간을 통하여 사용자는 손으로 직접 조작하는 것 같은 현실감을 느낄 수 있다. 실험을 통하여 제안한 실시간 깊이추정 알고리즘과 손동작 인식의 정확성을 확인하였다. 제안하는 알고리즘과 기존의 스테레오 카메라를 이용한 거리추정 알고리즘과의 결과비교를 통해 제안하는 알고리즘의 깊이추정 성능이 향상됨을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 전체적인 구조도와 세부적인 동작을 설명하고, 3장에서는 제안한 시스템의 실험 조건 및 결과에 대해 설명하고, 4장에서는 결론을 서술한다.

II. 제안하는 3차원 공간인식 시스템

그림 1은 본 논문에서 제안하는 3차원 공간인식 시스템의 전체적인 구조이다. 먼저 스테레오 카메라로부터 얻은 좌우 영상에서 피부색의 특징을 이용하여 손 영역을 분리한다. 손 영역의 분리 후, 조명의 영향으로 발생하는 잡음 성분을 제거하기 위한 과정을 거친다. 정합과정에 쓰일 관심영역의 특징점을 구하여 실시간으로 변이를 추정하며 깊이 정보를 포함한 3차원 좌표를 계산한다. 동시에 손가락의 개수를 계산하여 관심영역인 손의 동작을 인식하여 마우스 포인터에 연계시켜서 손으로 마우스를 제어하도록 한다. 사용자는 계산된 3차원 공간좌표와 인식된 손동작의 정보

를 통해 가상공간을 제어한다.

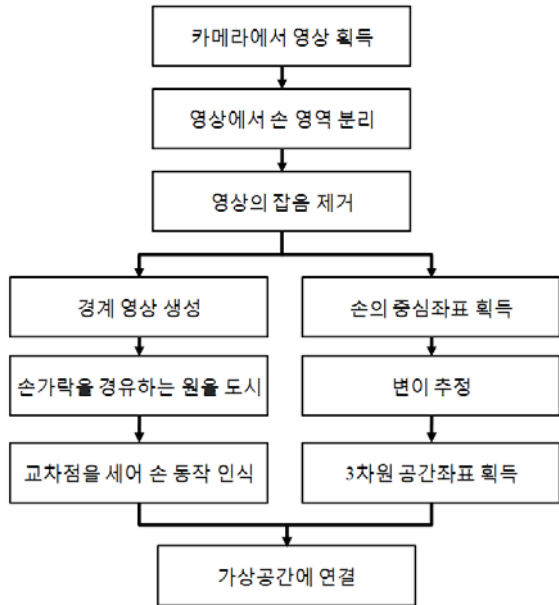


그림 1. 제안하는 시스템 구조도
Fig 1. The proposed system structure

1. 관심영역 분리

제안한 시스템에서는 관심영역만을 따로 분리하여 변이를 추정하는 방법을 사용한다. 이를 위해서는 관심영역이 비관심영역과 정확히 구분 되어야 한다. 본 논문에서는 관심영역의 분리를 위해서 그림 2와 같은 과정을 거친다.

카메라에서 획득된 영상은 가로 1280개, 세로 960개의 화소개수로 구성되며 이 크기의 영상은 연산량이 많아 실시간으로 처리하기에 부적합하다. 연산량을 줄이기 위해 영상을 가로와 세로로 각각 4배 씩 다운 샘플링(down sampling)을 수행하여 영상의 크기를 각각 가로 320개, 세로

240개로 줄이고 이 축소된 영상에서 관심영역을 추출하고 변이를 추정하게 된다. 제안한 시스템에서는 사용자가 직접 다룰 수 있도록 손 영역을 관심영역으로 정하였다.

손 영역을 추출하기 위하여 피부색의 특징을 이용하여 색차 평면상에서 자연계의 다른 물체와 구분할 수 있도록 한다. 사람과 인종마다 피부색 차이는 존재하지만 색차 평면상의 그 분포는 매우 협소한 영역을 차지하고 있고, 밝기의 차이만이 존재하므로 색차 성분을 이용하면 쉽게 피부영역을 추출할 수 있다. 이 경우 살색 계통의 물체가 배경영역에 존재하거나 비슷한 색상의 반사광이 피부영역과 함께 검출되는 문제점이 발생할 수 있다. 보다 정확한 검출을 위해서 제안하는 시스템에서는 YCbCr 색상 구조를 사용한다. YCbCr 색상 구조에서 피부색은 매우 좁은 영역에 분포하므로 다른 색상 구조에 비해 효과적으로 피부색을 추출할 수 있다. 또한 카메라가 제공하는 RGB 색상 구조는 조명의 영향에 따라 같은 위치상의 화소 각각의 RGB값이 크게 변동하는 반면에 YCbCr 색상 구조는 명암 성분(luma component) Y와 색차 성분(chrominance components) Cb, Cr로 나누어서 색을 표현하기 때문에 조명 환경의 변화에도 색차 성분 Cb, Cr은 비교적 안정적인 값을 가지는 장점이 있다^[7].

$$\begin{aligned} 100 &\leq C_b \leq 129 \\ 128 &\leq C_r \leq 137 \end{aligned} \tag{1}$$

식 (1)에서 Cb는 파란색의 색차 성분(blue chrominance)을, Cr은 붉은색의 색차 성분(red chrominance)을 의미한다. 식 (1)은 실험적으로 피부색에 해당하는 색차 성분의 범위를 구한 것이다. 다운 샘플링을 수행하여 크기가 줄어든 좌우 영상에 식 (1)의 조건을 이용하여 이진화를 수행한다. 식 (1)의 조건에 충족하는 화소의 밝기값은 255로 하여 화소의 밝기값을 흰색으로 변경하고, 그렇지 않은 화소의 밝기값은 0으로 하여 화소의 색을 검은색으로 변경한다. 그림



그림 2. 관심영역을 분리하는 과정
Fig 2. Process for estimating a region of interest

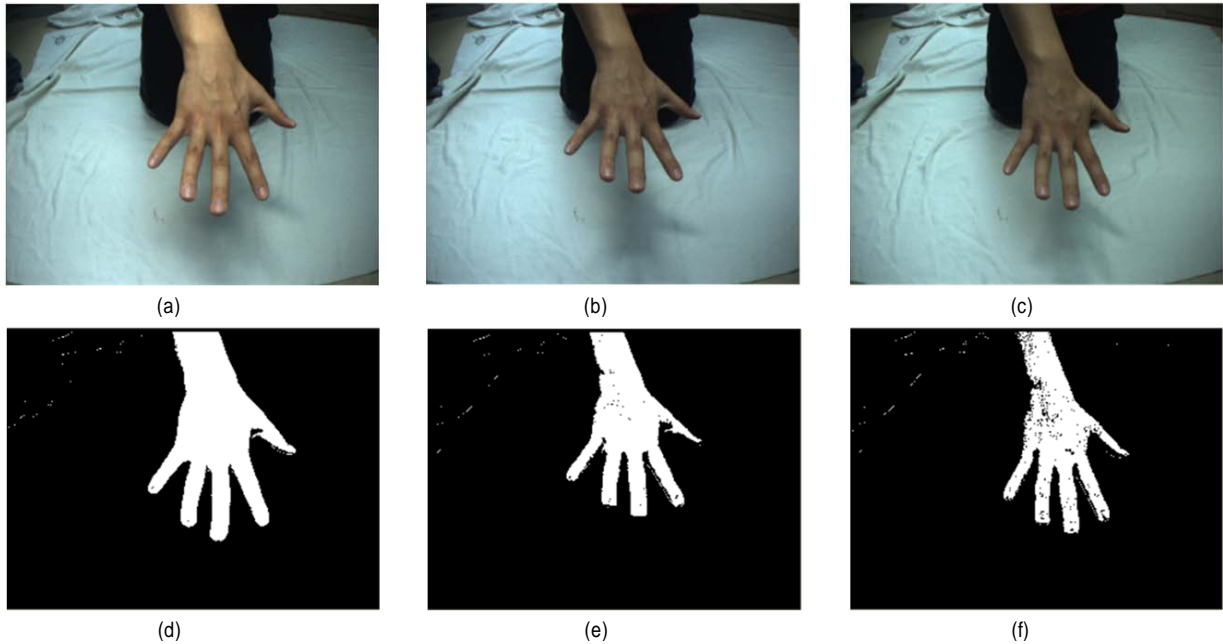


그림 3. 색차 정보의 특징을 이용하여 이진화한 손 영역 영상 (a) 조명이 밝을 때, (b) 조명이 중간일 때, (c) 조명이 어두울 때, (d), (e), (f) (a), (b), (c)의 이진화 영상

Fig 3. Binary images obtained by using color information (a) well lighted, (b) medium lighted, (c) poorly lighted, (d), (e) and (f) binary images of (a), (b), (c)

3은 조명의 강도가 각기 다른 환경에서의 원영상과 식 (1)의 조건으로 이진화를 수행하여 손 영역을 얻어낸 영상을 보여준다.

그림 3을 통해 각기 다른 밝기의 조명 조건에서도 YCbCr 색상 구조상의 식 (1)의 조건을 이용하여 수행한 이진화가 비교적 손의 형태를 잘 찾아내는 것을 확인할 수 있다. 그러나 그림 3(a), 그림 3(b), 그림 3(c)와 같이 광원의 방향과 강도에 따라 그림 3(d), 그림 3(e), 그림 3(f)와 같이 이진화를 수행한 후의 관심영역 내외에 잡음의 발생 정도가 다른

것을 알 수 있다. 일정량 이상의 조명이 비춰지는 환경에서는 그림 3(d)와 같이 손 영역 내외에 잡음이 거의 발생하지 않지만 그림 3(e), 그림 3(f)와 같이 점점 어두워질수록 발생하는 잡음의 양이 늘어나게 된다.

주위의 조명 환경에 따라 발생할 수 있는 이러한 잡음을 제거하기 위해 홀 처리(hole filling)와 스무딩 필터링(smoothing filtering) 과정을 수행한다. 홀 처리는 영상을 수직과 수평, 두 방향으로 탐색하면서 흰색의 화소간의 간격이 임계치보다 작으면 두 화소 사이의 모든 화소의 색을

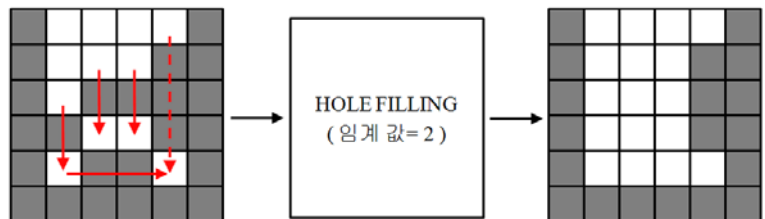


그림 4. 홀 처리 과정
Fig 4. Hole filling process

흰색으로 변경시키고 임계치보다 크면 두 화소 사이의 모든 화소의 색을 검은색으로 변경시키는 동작을 수행한다. 그림 4는 임계치가 2로 설정된 홀 처리의 동작을 보여준다.

흰색의 밝기값을 가지는 화소사이의 간격이 임계치보다 작거나 같을 경우, 그 사이 화소들은 모두 흰색으로 변경된다. 홀 처리는 정렬 과정이 없어 수행속도가 빠르고 관심영역 내의 잡음은 효과적으로 제거하는 장점을 가지지만 관심영역 이외의 부분에 생성된 잡음은 제거하지 못하고 잡음을 더하게 되는 단점을 가지고 있다. 그림 5는 각각 다른 임계값을 적용한 홀 처리의 결과를 보여준다. 그림 5에서 홀 처리가 적용되는 영상은 잡음제거의 성능을 확인하기 위해 이진화 수행 후 가장 잡음이 많이 발생한 그림 3(f)의 영상을 사용하였다.

그림 5(a)처럼 임계값이 너무 작으면 제거되지 못한 잡음이 존재하게 되고, 반대로 그림 5(c)와 같이 임계값이 너무 크면 잡음은 제거되지만 손가락 사이의 공간이 메워져 손 영역을 완벽히 구별해 내지 못한다. 그림 5를 통해 가장 적당한

홀 처리의 임계값이 5임을 알 수 있다. 그림 5(b)에서 검지와 중지 사이나 배경영역의 잡음과 같이 홀 처리로 제거하지 못한 관심영역 외의 잡음을 제거하기 위해서 스무딩 필터링을 수행한다. 식 (2)는 스무딩 필터의 동작을 보여준다.

$$I(x, y) = \begin{cases} 255, & \sum_{k=-N/2}^{N/2} \left\{ \sum_{l=-N/2}^{N/2} I(x+k, y+l) \right\} \geq \frac{255 \times N^2}{2} \\ 0, & \text{다른 경우} \end{cases} \quad (2)$$

식 (2)에서 $I(x, y)$ 는 영상의 (x, y) 좌표에서 화소의 밝기 값을 의미한다. 화소 단위로 영상을 탐색하면서 중심화소를 둘러싼 일정 크기의 블록내의 모든 화소 값을 더한 값이 식 (2)의 임계값(블록의 크기×화소 밝기의 최대값÷2)이상이면 블록의 중심화소의 색을 흰색으로 변경하고 임계값 이하이면 검은 색으로 변경한다. 스무딩 필터링은 저역통과 필터를 수행한 것처럼 영상이 다소 흐려지는 현상이 발생하지만 홀 처리를 통해 제거하지 못하는 관심영역 외의

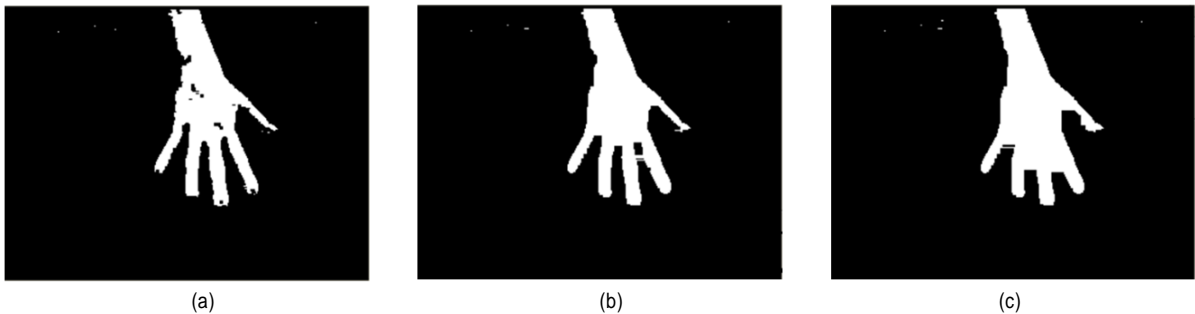


그림 5. 홀 처리를 수행한 영상 (a) 임계값이 2일 때, (b) 임계값이 5일 때, (c) 임계값이 10일 때
Fig 5. Obtained images with a hole filling process (a) when critical value is 2, (b) 5 and (c) 10

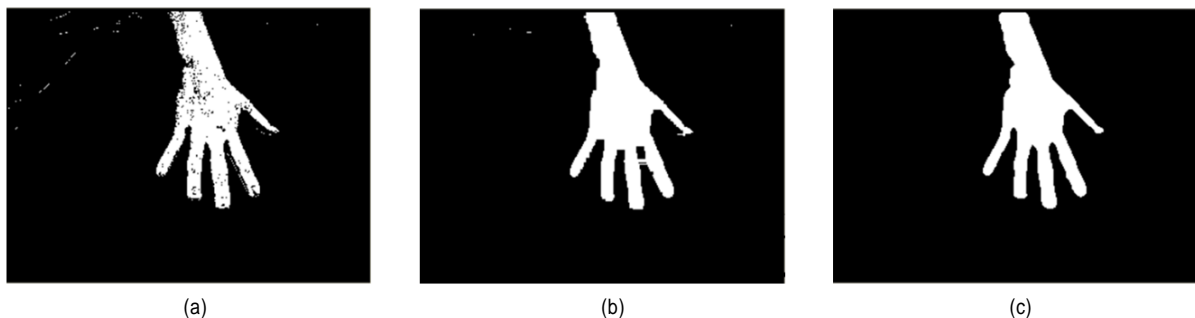


그림 6. 손 영역 분할 과정 (a) 이진화 영상, (b) 홀 처리를 수행한 후의 영상, (c) 홀 처리와 스무딩 필터링을 수행한 후의 영상
Fig 6. segmentation process for hand region (a) binary image, (b) after a hole filling, (c) after a hole filling and a smoothing filtering

잡음을 제거하는 데 효과적이다. 그림 6은 그림 3(f)의 영상에 홀 처리와 스무딩 필터링을 차례로 수행하는 손 영역 분할의 전 과정을 보여준다.

그림 6(c)의 결과 영상을 보면 그림 3(f)의 영상에 존재하는 잡음 성분이 많이 제거된 것을 알 수 있다. 그림 6을 통해 홀 처리와 스무딩 필터링 과정을 차례대로 적용하는 것이 관심영역을 추출하는 데 있어서 효과적이라는 것을 확인할 수 있다.

2. 변이추정

스테레오 카메라로부터 획득된 좌우 영상은 시점의 차이 때문에 동일한 물체에 대한 좌우 영상의 좌표값의 차이가 발생한다. 이를 변이(disparity)라고 한다. 일반적으로 관심 영역의 변이를 추정할 때 관심영역내의 모든 화소(pixel)의 변이를 추정하지만, 본 연구에서는 다양한 변이추정 기법 중에서 일정 영역의 영상 조각(image patch)을 가지고 다른 영상과의 상관관계를 이용하여 변이를 추정하는 영역기반 정합기법을 이용한다. 영역기반 정합은 주로 블록 단위로 비교하여 변이를 추정하는 형태이기 때문에 블록의 크기에 따라 결과가 가변적이다. 블록의 크기가 작으면 속도가 빠르지만 잡음에 취약하고 블록의 크기가 크면 잡음에 강하지만 속도가 느리고 일종의 저역통과 필터의 역할을 하여 세밀한 처리가 불가능하다. 이 같은 문제점을 보완하기 위해 일반적으로 그림 7과 같이 관심영역을 분할하여 이 영역에서만 변이추정을 수행함으로써 속도와 정확도를 높인다⁵⁾.

본 논문에서는 그림 7과 같이 모든 관심영역에 대하여 변이를 추정하는 대신 좌우 관심영역의 특징점을 추출하고 비교하는 방식으로 변이를 추정하여 실시간 처리가 가능한 기법을 제안한다. 관심영역으로 판단된 화소의 좌표값들을 모

두 더하고 그 개수로 나누어 수직, 수평적으로 평균이 되는 중심좌표를 얻어내고 관심영역의 특징점으로 사용한다. 관심영역의 2차원 중심좌표를 얻는 방법은 다음 식 (3)과 같다.

$$(X_c, Y_c) = \frac{\sum_{m=0}^N (X_m, Y_m)}{N} \tag{3}$$

식 (3)에서 X_m 과 Y_m 은 관심영역의 (x, y) 좌표의 값을 누적한 값이고, N 은 관심영역내의 화소의 총 개수를 의미한다. 이와 같이 관심영역의 모든 화소들의 (x, y) 좌표의 순서쌍을 모두 더하여 화소의 총 개수 N 으로 나누어주면 좌우 영상에서 관심 영역의 중심좌표 (X_c, Y_c) 가 계산된다. 그림 8은 식 (3)을 통해 계산된 관심영역의 중심좌표 (X_c, Y_c) 를 보여준다.

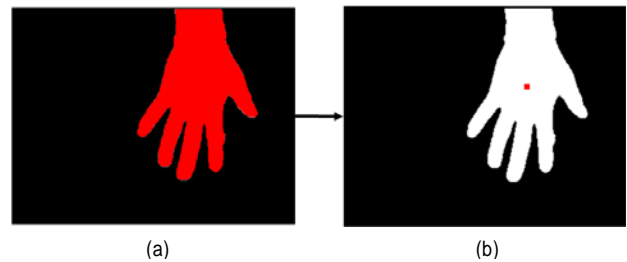


그림 8. 변이추정을 위한 비교 범위 (a) 기존의 방법, (b) 제안한 시스템에서 사용하는 방법
Fig 8. The range of correspondence for disparity estimation (a) Segmentation-based method, (b) The proposed method

좌우 관심영역 각각의 중심좌표를 식 (4)에 대입하여 깊이 정보를 얻어낸다. 식 (4)에서 Z 는 계산된 거리를 의미한다. 식 (3)을 이용하여 얻은 2차원 중심좌표에 Z 를 추가하여 3차원 공간좌표를 완성한다.

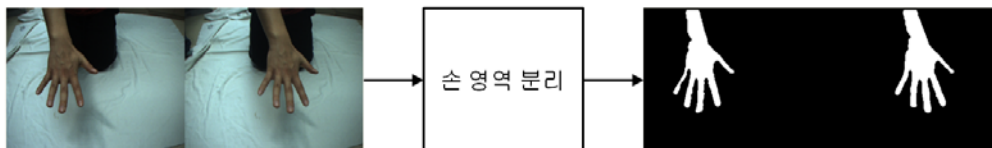


그림 7. 영역 분할 기법을 이용한 정합
Fig 7. Stereo matching by segmentation

$$Z = \alpha f \frac{B}{X_{CR} - X_{CL}} \quad (4)$$

여기서 B 는 두 카메라 렌즈 중심 간의 거리를 나타내고 f 는 두 렌즈의 초점거리를 나타낸다. X_{CR} 과 X_{CL} 은 좌우 관심영역의 x축 중심좌표이다. y축 변이는 없다고 가정하였다. α 는 실험을 통해 계산되는 가중치로 원하는 단위의 값으로 환산할 때 사용한다. 실험을 통해 깊이 정보를 센티미터(cm) 단위로 측정하기위한 α 의 값은 15.74임을 확인하였다.

3. 손동작 인식

제안하는 시스템은 식 (3)을 통해 얻은 좌우 관심영역의 중심좌표에 식 (4)를 이용하여 구한 측정 거리를 추가하여 구성된 3차원 공간좌표를 마우스 포인터가 추적하도록 하여 관심영역인 손을 이용하여 마우스를 제어한다. 또한 손가락 개수를 계산하여 손동작을 인식하고 지정된 손동작을 마우스의 클릭 동작과 연계하여 손으로 마우스의 동작을 구현한다. 스테레오 카메라에서 획득된 좌우 영상중 하나의 영상을 지정하여 손 영역의 경계를 구하고, 식 (3)에서 계산된 중심좌표를 중심으로 하는 원과의 교점의 개수를 이용하여 손가락의 개수를 계산한다. 먼저 추출된 영역에 소벨 필터(Sobel filter)를 거쳐 경계 영상을 구하고 획득된 경계는 원과 한 개의 교점만을 가지도록 세선화(edge thinning) 과정을 거친다. 그림 9의 세선화 과정에서는 획득된 경계 영상과 소벨 필터를 거치기 전의 손 영역 분할 영상과

의 뺄셈 연산이 수행된다.

이렇게 하면 획득된 경계 영상에서 관심영역으로 분리된 안쪽 화소들이 지워져 관심영역을 둘러싼 가는 경계만 남는다. 뺄셈 연산으로 인해 발생하는 음의 화소값들은 clipping 기법을 사용하여 그 밝기값을 0으로 한다. 손가락 개수를 구하기 위한 원의 반지름은 카메라로부터의 거리에 따른 손 영역의 크기에 따라 가변적으로 변하게 한다. 즉, 식 (5)과 같이 중심좌표로부터 22.5° 씩 16방향을 검색하여 경계와 만나는 점과 중심좌표와의 거리를 비교하는 방식으로 추출된 손 영역의 크기를 추정하고 원의 크기를 결정한다.

$$R = \alpha \sum_{k=2}^{13} \frac{D_k}{12} \quad (5)$$

식 (5)에서 D_k 는 경계와 만나는 16개의 점들과 중심좌표 간의 거리 값을 내림차순으로 정렬한 배열의 k 번째 값을 의미한다. 경계 영상과의 16개의 교점들 중 그림 10에서 일부 노란색 점과 같이 손목이나 손가락의 끝 부분에서의 중심으로부터 거리가 긴 교점이나 반대로 손가락 사이의 부분에서의 중심으로부터 거리가 짧은 교점은 다섯 손가락을 모두 경유하는 적절한 반지름 R 을 추정하는데 방해요소로 작용할 수 있다. 그래서 손가락을 모두 경유하는 적절한 반지름 R 을 구하기 위해 식 (5)와 같이 D_k 중 가장 큰 값 2개와 가장 작은 값 2개를 제외한 12개의 D_k 값의 평균값을 이용한다. 이때 가중치 α 는 사용자가 손을 뗐을 때 원이 다섯 개의 손가락과 모두 겹쳐지도록 반지름 R 의 값을 조절하는데 사용한다.

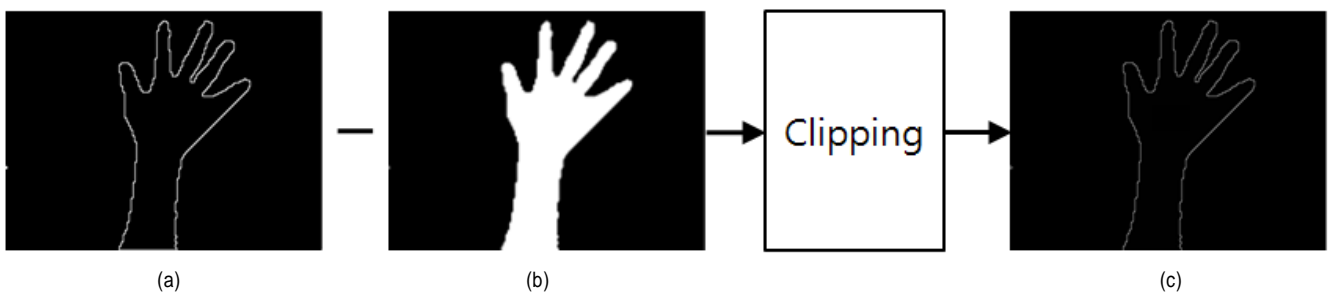


그림 9. 경계 세선화 과정 (a) 소벨 필터를 이용한 경계 영상, (b) 손 영역 분할 영상, (c) 세선화된 경계 영상

Fig 9. Edge thinning process (a) edge image with Sobel filtering, (b) segmented hand image and (c) Image after an edge thinning process

원과 만나는 교점의 개수는 식 (6)을 이용하여 손가락의 개수 N 으로 변환된다. 원과 경계의 교점은 다섯 개의 손가락과 손목에 각각 2개씩 12개가 생성되며 손목에서의 교점을 제거하기 위해 총 교점의 개수에서 2를 뺀다. 그 후에 손가락 하나의 경계와 원은 두 번 교차하기 때문에 교점을 2로 나누었으며 그 값을 소수 첫째자리에서 반올림을 수행하는 $ROUND$ 함수에 대입하여 자연수로 나타낸다.

$$N = ROUND\left(\frac{Intersection\ points - 2}{2}\right) \quad (6)$$

그림 10은 제안한 손동작 인식 알고리즘을 적용한 영상을 보여준다. 노란색 점들은 중심좌표로부터 검색된 16방

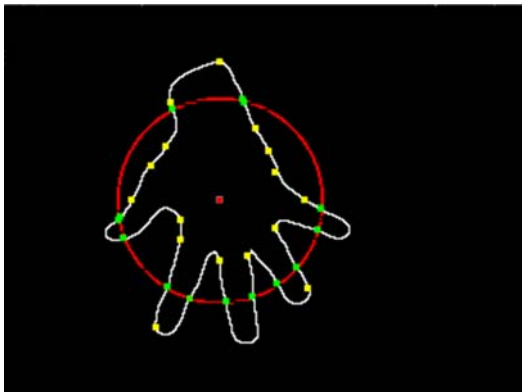


그림 10. 경계와의 교점을 이용하여 원을 그린 영상
Fig 10. Circle by using intersected points with an edge map

향에서 경계와 만나는 점을 나타내고 연두색 점들은 원과 경계와의 교점을 나타낸다. 식 (3)을 통해 중심좌표와 식 (5)를 통해 얻은 반지름 R 을 이용하여 원을 그리고 22.5°씩 16방향으로 영상을 탐색하여 경계와의 교점을 영상에 표시한다. 식 (6)을 사용하여 손가락의 개수 N 을 얻어내어 손의 모양을 인식하고 마우스에서의 특정 동작과 연계한다.

4. 3차원 가상공간의 구현

3차원 좌표가 마우스 포인터를 추적하는 동작을 수행하는 플레이어를 제작하였다. 또한 마우스 포인터의 움직임으로 제어할 수 있는 가상공간과 플레이어를 제작하였다. 3차원 가상공간은 $OPENGL$ API(application programming interface)를 사용하여 제작하였다. 가상공간은 마우스 포인터의 움직임을 따라 미로 찾기를 할 수 있도록 제작하였으며 관심영역의 상하, 좌우, 전후의 3차원적인 움직임을 따르도록 제작하여 사용자는 손의 움직임으로 가상공간 상의 위치를 조정할 수 있도록 하였다.

그림 11은 제안한 모든 알고리즘과 동작을 포함하여 제작한 시스템의 플레이어를 보여준다. 버튼을 통해 카메라로부터 얻은 좌우 원 영상과 알고리즘을 거쳐 변이추정에 쓰이는 영상을 보여준다. 3차원 좌표를 마우스 포인터와 연결시키는 추적 동작, 그리고 $OPENGL$ 을 이용하여 제작한 가상공간의 구동은 별도의 버튼으로 동작 시키도록 하였다. 간단한 마우스의 동작을 설명하는 박스와 좌우 영상의 2차

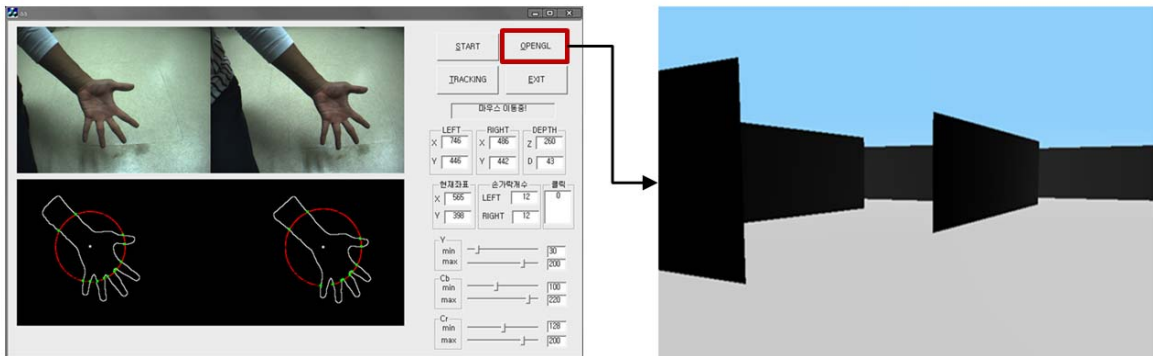


그림 11. 가상공간을 제어하는 시스템
Fig 11. System to control the virtual space

원 중심좌표, 그리고 알고리즘을 통해 추정된 좌우 특징점 간의 변이값과 식 (4)를 이용하여 얻은 실제거리의 측정치를 실시간으로 보여주도록 제작하였다. 조명에 영향을 비교적 덜 받는 YCbCr 색차 구조를 사용하여 피부를 추출하지만 사용 환경에 따라 피부로 인식되는 색차 정보의 값도 유동적이기 때문에 색차 구조의 조건을 조절할 수 있도록 컨트롤러로 구현하였다.

III. 실험 조건 및 결과

본 논문에서는 제안하는 실시간 변이추정 알고리즘의 성능을 검증하기 위해 거리에 따라 얻은 깊이 정보와 실제거리 값과의 오차를 측정하는 실험을 진행하였다. 실험에는 범용 스테레오 카메라(모델명 : BUMBLEBEE XB3)를 이용하였고, 홀 처리의 임계치가 되는 화소의 개수는 5이고, 스무딩 필터의 마스크의 크기는 5×5를, 임계값은 3188 (25×255÷2)을 사용하였다. 처리 속도는 초당 약 10~11 프레임의 속도를 보였다.

그림 12(a)의 실제거리에 따른 좌우 영상의 중심점간의 차이, 즉 식 (4)에서 $X_{CR} - X_{CL}$ 를 그래프로 보여준다. 거리에 따라 획득된 $X_{CR} - X_{CL}$ 과 실험에 사용된 카메라의 파라미터 값, $B(12)$ 와 $F(60)$ 를 식 (4)에 대입하여 측정거리 Z 를

획득한다. 측정거리 Z 를 센티미터(cm) 단위로 변경하기 위한 α 값은 실험적으로 평균 15.74를 가짐을 확인하였다. 그림 12(b)는 식 (4)를 통해 측정된 거리 Z 와 실제거리와의 차이를 그래프로 보여준다. 좀 더 정량적인 실험결과를 얻어내기 위해 식 (7)과 식 (8)을 통해 획득한 깊이 정보와 실제거리의 오차와 평균 오차거리를 측정하였다.

$$\text{오차}(cm) = |\text{실제 거리} - \text{측정 거리}| \tag{7}$$

$$\text{평균 오차}(cm) = \frac{\text{시행한 오차의 총합}}{\text{총 실험 시행 횟수}} \tag{8}$$

실제거리와 측정거리와의 차이의 절대값을 오차로 계산하여 반복 시행하였다. 또한 식 (8)을 통해 같은 실제거리를 두고 계산된 오차간의 평균을 계산하여 좀 더 정확한 성능을 측정하고자 하였다. 이와 같이 정량화한 실험결과를 표 1에 정리하였다.

표 1. 측정거리의 평균 오차
Table 1. The average error of measured distance

카메라로부터의 거리(cm)	제안한 알고리즘의 평균 오차거리(cm)	제안한 알고리즘의 평균 오차율(%)
20 ~ 40	1.75	5.83
40 ~ 150	0.5	0.53
150 ~ 250	6.8	3.4

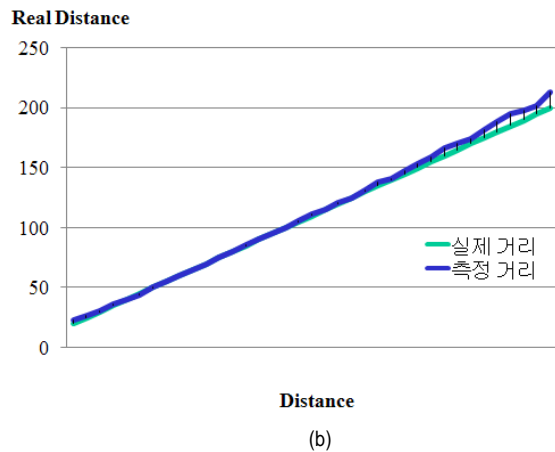
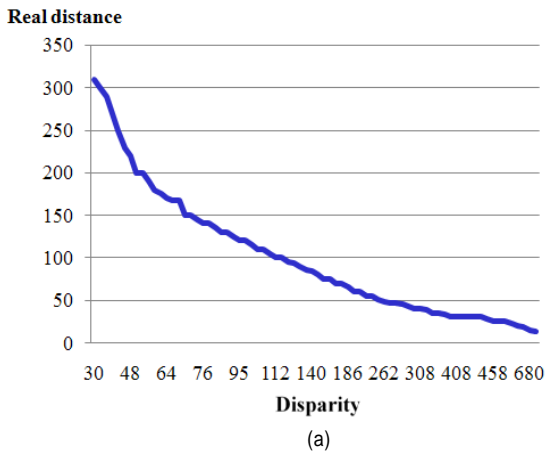


그림 12. 거리에 따라 측정결과 (a) 변이, (b) 실제거리와 측정거리의 오차
Fig 12. Experimental results with distance (a) disparity and (b) the error between measured and real distances

실험결과 근거리(40~150cm)에서는 높은 신뢰도를 보였지만 40cm 미만의 거리에서 큰 오차율을 가졌다. 이는 제한적인 스테레오 카메라의 좌우 화각으로 인해 카메라에 관심 물체가 점점 근접할수록 좌우 영상에 맞는 관심 물체의 대응영역이 점차 줄어들기 때문이다. 또한 200cm 이상의 거리에서도 오차율이 증가하는데 이는 거리에 증가함에 따라 좌우 영상에 맞는 관심 물체의 대응영역 정보량이 점차 감소하기 때문이다.

또한 손가락의 개수에 따라 마우스의 동작을 연결시키고 인식의 정확성을 측정하는 실험을 수행하였다. 마우스 동작 구분은 손가락의 개수에 따라 다르게 대응시켰다. 손가락의 개수가 1개 이상일 때는 마우스의 이동 동작에 대응시키고 1개일 때는 좌클릭 동작에 대응시켰다.

표 2. 손가락 개수 인식률
Table 2. The correct recognition rate

손가락 개수	인식률 (%)
1	89
2	87
3	90
4	97
5	88
평균(%)	90.2

표 3. 손동작 인식 성공률
Table 3. The success rate of hand shape recognition

마우스 동작 구분	성공률(%)
이동	97
좌클릭	73

표 2는 실험을 통해 손가락의 개수를 세는 동작의 인식률을 보여준다. 평균적으로 90.2%의 인식률을 보였다. 표 3을 통해 마우스의 이동 동작에서는 97%의 성공률을 보였으나 좌클릭 동작의 성공률은 낮아지는 것을 확인할 수 있었다.

표 4는 제안하는 알고리즘과 유사한 방식으로 관심영역의 깊이를 추정하는 알고리즘들과의 평균 오차율을 비교한다. 비교하는 첫 번째 알고리즘은 관심물체주위의 비슷한 색상이나 반사광에 의해서 물체가 뚜렷하게 인식되지 못하여 오차가 생기는 문제점을 개선하기 위해서 레이저 광선

을 물체에 주사하여 그 점광원을 인식하여 거리를 구한다^[11]. 두 번째 알고리즘은 좌우 각각의 영상으로부터 관심 물체의 중심점을 찾고, 이들 간의 대응성을 비교한 변이를 추출하여 이동하는 물체의 3차원 좌표의 실시간 획득을 목표로 한다. 빠른 연산을 위해 영상의 크기를 축소한 후, 관심 물체의 색상의 특징을 이용하여 이진화를 수행하고 중심점을 찾는 부분도 제안하는 알고리즘과 상당히 유사하다^[12].

표 4. 기존 연구와의 결과비교
Table 4. Performance comparison of the proposed algorithm with a conventional algorithm

카메라로부터 거리(cm)	기존 알고리즘의 평균 오차율(%) ^[11]	기존 알고리즘의 평균 오차율(%) ^[12]	제안한 알고리즘의 평균 오차율(%)
20~40	6.2	4.8	5.8
40~60	5.6	4.4	0.6
60~80	4.6	3.3	0.2
80~100	3.6	2.9	0.2

표 4를 통해 관심영역의 특징점을 이용한 깊이추정에 있어서 기존의 알고리즘에 비해 제안한 알고리즘이 향상된 성능을 가지는 것을 확인할 수 있다.

IV. 결론

3차원 공간에서 한 점이 좌우 영상에 투영되는 대응점을 찾아내는 정합과정은 스테레오 비전에서 매우 중요한 과정이다. 정합할 관심영역에 포함된 모든 화소를 가지고 대응점을 구하는 영역기반의 정합방식은 연산량이 많아서 처리에 상당한 시간이 소요된다는 단점이 있다.

본 논문에서는 영역기반의 정합과정의 연산량을 줄이기 위해 관심영역의 중심좌표만을 추출하고 정합과정에 사용하여 실시간으로 깊이 정보를 얻어내는 시스템을 구현하였다. 시스템은 피부색이 가지는 색차 구조상의 특징을 이용하여 손의 중심좌표를 획득하고 정합에 사용하여 손의 깊이 정보를 얻어낸다. 중심좌표에 깊이 정보를 추가한 3차원 좌표를 획득하고 좌표의 움직임을 마우스 포인터가 추적하도록 하였다.

제한한 특징점 기반의 변이추정 방법은 깊이 정보의 실시간 획득을 위해 좌우 영상의 대응영역을 상당히 축소한다. 깊이추정에 사용되는 영상의 정보가 현격히 줄어들었음에도 불구하고 기존의 실시간 거리 측정 알고리즘과 비교할 때 근거리에서 높은 신뢰도를 보임을 확인할 수 있었다. 그러나 본 논문에서 정합에 사용되는 영상 정보의 축소는 빠른 깊이의 추정을 가능케 하지만 반대로 정확한 깊이 추정이 가능한 거리는 줄어드는 결과를 가진다. 앞으로 깊이추정의 정확성과 실시간성을 유지하면서 깊이 추정이 가능한 거리영역을 확장하는 효과적인 방법에 대한 연구가 향후 이루어 져야 할 것이다.

참 고 문 헌

- [1] 김수인, 이재수, 손영우, "영상감시 시스템에서 평행식 스테레오 카메라를 이용한 다중 이동물체의 거리측정," 조명전기설비학회논문지, 제18권 제1호, pp. 137~145, Jan, 2004.
- [2] 김수인, 남궁재찬, "스테레오 비전 시스템에서 레이저 포인터를 이용한 표적물까지의 거리 측정," 호 한국통신학회논문지, 제27권 5B호, pp. 541~550, May, 2002.
- [3] 이재수, 홍권의, 손영우, "방법 설비의 스테레오 카메라 시스템에서 이동객체의 추적과 거리 측정," 조명전기설비학회논문지, 제17권 제4호, pp. 106 ~ 115, Jul, 2003.
- [4] 이정수, 양승구, 김준성, "실시간 스테레오 비전 시스템을 위한 SAD 정합연산기 설계," 전자공학회 논문지, 제45권 CI편, 제1호, pp. 55 ~ 61, Jan, 2008.
- [5] 박상백, 문종섭, "동적계획법을 적용한 효과적인 폐색영역제거를 위한 고속 스테레오 정합법," 고려대학교 대학원 전자정보공학과 학위논문, 2001.
- [6] 김계영, 장석우, "유형기반 스테레오 정합을 통한 영상 변이 측정," 한국컴퓨터정보학회 논문지, 제11권, 제2호, pp. 83 ~ 92, May, 2006.
- [7] 김장운, 김송국, 장한별, 배기태, 이철우, "에지 방향성 히스토그램 데이터를 이용한 손 형상 인식," 한국 정보과학회 학술발표논문집, 제33권, 제2호(B), pp. 49 ~ 53, Oct, 2006.
- [8] 김종민, 이철우, "윈도우 플레이어 제어를 위한 에지 방향성 히스토그램 손 형상 인식," 한국정보과학회 2003년도 가을 학술발표논문집, 제30권, 제2호(II), pp. 628 ~ 630, Oct, 2003.
- [9] C. Shan, Y. Wei, T. Tan and F. Ojardias, "Real Time Hand Tracking By Combining Particle Filtering and Mean Shift," Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FG'04), 669-674, Seoul, South Korea, May, 2004.
- [10] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing," Second Edition.
- [11] 김균수, "양안 카메라를 이용한 이동 목표물의 3차원 거리 측정 방법 연구," 호남대학교 전자공학과 학위논문, 2002.
- [12] 이찬주, "스테레오 카메라를 이용한 입체영상화 및 거리 측정에 관한 연구," 충남대학교 전기공학과 학위논문, Aug, 2005.

저 자 소 개



이 동 석

- 2009년 2월 : 광운대학교 전자공학과 학사
- 2009년 3월 ~ 현재 : 광운대학교 전자공학과 석사과정
- 주관심분야 : Video Coding, Image Processing



이 동 욱

- 2009년 2월 : 광운대학교 전자공학과 학사
- 주관심분야 : Image Processing, Stereo Matching

저 자 소 개



김 수 동

- 2009년 2월 : 광운대학교 전자공학과 학사
- 2009년 3월 ~ 현재 : 광운대학교 전자공학과 석사과정
- 주관심분야 : Image Processing, 2D/3D Conversion



김 태 준

- 2008년 2월 : 광운대학교 전자공학과 학사
- 2008년 3월 ~ 현재 : 광운대학교 전자공학과 석사과정
- 주관심분야 : Stereo matching, Computer Vision, 동영상 코덱



유 지 상

- 1985년 2월 : 서울대학교 전자공학과 학사
- 1987년 2월 : 서울대학교 대학원 전자공학과 석사
- 1993년 5월 : Purdue 대학교 전기공학과 박사
- 1993년 9월 ~ 1994년 8월 : 현대전자산업(주) 산전연구소 선임연구원
- 1994년 9월 ~ 1997년 8월 : 한림대학교 전자공학과 조교수
- 1997년 9월 ~ 2001년 8월 : 광운대학교 전자공학과 조교수
- 2001년 9월 ~ 2006년 8월 : 광운대학교 전자공학과 부교수
- 2006년 9월 ~ 현재 : 광운대학교 전자공학과 정교수
- 주관심분야 : 웨이블릿 기반 영상처리, 영상압축, 영상인식, 비선형 신호처리