

論文96-33B-9-9

# 손 제스처 인식을 위한 상호작용 시각정보 추출

## (Interactive Visual Knowledge Acquisition for Hand-Gesture Recognition)

梁善玉\*, 崔瀾一\*

(Seon-ok Yang and Hyung-il Choi)

## 요 약

컴퓨터 시각에 기반한 제스처 인식시스템의 작업은 일반적으로 분할, 추적, 판단의 3단계로 이루어진다. 그러나 다양한 조명과 배경의 입력 영상으로부터 제스처에 이용되는 신체 부분을 분할하는 것은 어려운 작업이다. 본 논문에서는 컴퓨터 시각에 기반한 손 제스처 인식 시스템의 향상을 위해서 분할을 위한 사전 정보들을 학습하는 방법을 기술한다. 정확한 훈련 자료와 학습된 분할 정보를 획득하기 위하여 시스템은 미리 정의된 계획에 따라서 사용자와 상호작용을 한다. 계획의 과정을 살펴보면 시스템이 사용자에게 모델을 제시한 후 사용자의 반응을 카메라로 입력 받아 분석한다. 입력된 영상을 분석할 때는 모델 정보와 사전 지식을 이용한다. 시스템은 분석된 결과에 따라 사용자에게 적절한 메시지를 전달하기도 하고 분할 정보를 위한 학습모듈을 구동시키기도 한다.

## Abstract

Computer vision-based gesture recognition systems consist of image segmentation, object tracking and decision. However, it is difficult to segment an object from image for gesture in computer systems because of various illuminations and backgrounds. In this paper, we describe a method to learn features for segmentation, which improves the performance of computer vision-based hand-gesture recognition systems. Systems interact with a user to acquire exact training data and segment information according to a predefined plan. System provides some models to the user, takes pictures of the user's response and then analyzes the pictures with models and a prior knowledge. The system sends messages to the user and operates learning module to extract information with the analyzed result.

## 1. 서 론

사람이 의사 교환을 위하여 가장 많이 사용하는 방법은 음성이다. 이와함께 제스처, 얼굴 표정 등을 포함하는 사람의 몸짓도 의사 교환의 상당한 부분을 차지한다. 때에 따라서는 전달하고자 하는 정보의 대부분이 이러한 행동에 포함되기도 한다. 예를들어 사람들의 관심을 어떤 물체로 집중시키고자 할 때의 가리키는 동작, 친구에게 인사할 때의 손 흔드는 동작, 갯수를 표

시할 때의 손가락 제스처 등이 있다. 그러므로 컴퓨터가 인간과 자연스럽게 상호작용을 할 수 있도록 컴퓨터 시스템에 제스처를 인식하는 능력을 부과해야 한다. 시각에 기반한 컴퓨터 시스템의 입장에서 제스처를 이해하려면 사용자의 동작 중에서 의미를 가지는 신체의 특정 부분(손)을 분할해 낸 후, 이 부분에서 취해진 모양이나 동작을 추적하여 추적된 특징들이 미리 정해진 동작 중에서 무엇을 표현하는가를 판단해 낼 수 있어야 한다. 분할, 추적, 판단의 3가지 문제 중 신체의 일부분이면서 비정형적인 모양을 갖는 손을 분할하기란 어려운 문제이기 때문에 많은 연구자들은 특정한 색깔을 입힌 장갑을 이용하거나, 간단한 손 동작만으로 제

\* 正會員, 崇實大學校 컴퓨터學部

(School of Computing, Soongsil University)

接受日字:1995年11月29日, 수정완료일:1996年8月19日

한하거나, 배경에 대해 일정한 정보를 유지하도록 제한하는 등 문제를 단순화하여 연구하고 있다<sup>[1,2,4]</sup>.

예를 들어, Davis와 Shah는 특별히 표기된 장갑을 사용하여 이진 영상으로부터 손가락을 분할하였고<sup>[1]</sup>, Francis와 Quek은 손 뒷면에 흑판을 둠으로써 장면으로부터 손을 쉽게 분할할 수 있도록 하였다<sup>[2]</sup>. 그러나 이러한 제한 조건들은 제스처를 이용하는 실용적인 사용자와의 인터페이스의 역할을 감소시킨다. 실용적인 사용자 인터페이스가 되기 위해서는 제스처 인식이 실시간에 이루어져야 할 뿐 아니라 사용자의 변화나 환경의 변화에도 적응할 수 있어야 한다. Kjeldsen과 Kender는 이러한 점을 보완하기 위해 템플리트를 이용하여 손을 분할한 후 손의 피부색깔(skin-color)에 대한 히스토그램을 훈련하는 방법을 제시하였다<sup>[3]</sup>. 이 시스템은 손 모양의 템플리트를 사용자에게 제시하고, 사용자는 카메라와 일정한 거리를 유지하면서 템플리트에 손을 채운다. 시스템은 템플리트에 채워진 화소들을 손의 색상으로 인정하여 손의 색깔을 이루는 채도와 명도에 대한 히스토그램을 작성하면서 손에 대한 히스토그램을 훈련시키고 있다.

본 논문에서는 시스템과 사용자가 협력하여 분할을 위한 특징을 추출하는 방법을 제시한다. 논문의 구성은 다음과 같다. II절은 적응적 분할을 위해 요구되는 학습시스템의 구성요소들과 그들간의 관계를 소개하고, III절에서는 학습시스템에서 사용되는 지식들이 어떻게 표현되는가와 지식들이 어떻게 운용되는지를 기술하고, IV절에서는 제스처 분할을 위한 규칙들이 어떻게 추출되는가를 설명한다. 마지막 V절에서는 실험 및 결과 분석과 향후 연구 계획에 대해 기술한다.

## II. 학습시스템 개요

가변적 환경에 적응하는 손 제스처 인식시스템을 위해서는 환경 변화에 따른 정보를 얻을 수 있어야 한다. 본 절에서는 손 제스처를 인식하는 과정에 필요한 정보를 얻기 위해서 요구되는 훈련 자료들을 사전에 획득하여 학습할 수 있는 시스템을 소개한다. 컴퓨터의 사용이 점차 대중화되어 감에 따라 인간과 컴퓨터간의 인터페이스는 매우 중요한 쟁점이 되고 있다. 특히 제스처에 대한 컴퓨터의 인식 능력은 인간과의 자연스러운 인터페이스를 제공한다는 측면에서 컴퓨터 시각을 기반하는 인터페이스의 개발이 점차 중요시 되고 있다.

그러나 카메라로부터 입력된 사용자의 제스처를 인식할 때 다양한 조명 및 배경으로부터 사용자의 제스처를 분할한다는 것은 어려운 작업이다. 더우기 훈련되지 않은 사용자일 경우에는 사용자의 미숙으로 정확하지 않은 데이터가 카메라를 통해 입력될 수도 있고 같은 의미의 제스처라고 해도 사용자마다 그 움직임이 조금씩 달라질 수도 있기 때문에 분할은 더욱 어려운 작업이 된다. 그러므로 시스템은 분할 작업을 수행하기 전에 환경 및 사용자에 따른 입력 영상을 분석하여 손 제스처의 분할 특징들을 획득할 수 있어야 한다. 뿐만 아니라 제스처 인식시스템의 성능 및 융통성을 위하여 시스템은 사용자로부터 원하는 제스처를 학습하거나 시스템에 정의된 제스처 명령어를 학습자에게 훈련시키는 방법도 고려되어야 한다.

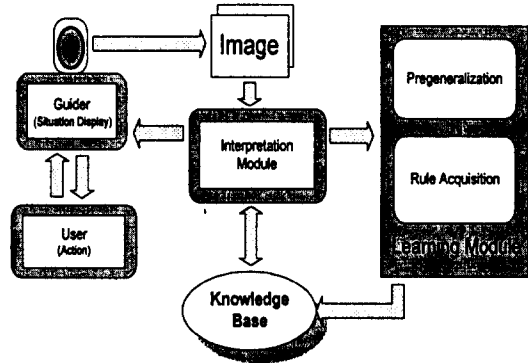


그림 1. 학습시스템 구조  
Fig. 1. Learning system architecture.

본 논문에서 제안하는 학습 시스템의 전체적인 구조는 그림 1에서 보여주듯이 시스템을 사용하는 사용자와 사용자의 행동을 지지하는 가이드, 사용자의 동작을 해석하는 해석모듈 및 학습 모듈과 지식베이스로 구성된다. 여기서 가이드는 해석모듈에 의해 제어되며, 해석모듈의 제어에 따라서 필요한 상황을 사용자에게 설명한다. 해석모듈은 학습을 위해서 어떤 상황을 가이드에게 전달할 지를 선택하는 작업과 카메라로부터 입력되는 사용자의 영상이 제시된 상황에 적합한 자료인지를 검증하는 작업 및 학습모듈을 구동시키는 역할을 한다. 카메라로부터 입력된 영상이 제시된 상황에 적합한 지를 검증하는 과정에서 시스템은 사전에 알려진 정보들을 이용한다. 여기에 대한 자세한 내용은 III절에서 소개한다. 학습 모듈은 전처리 과정과 규칙을 획득하는 과정으로 이루어진다. 학습 모듈의 전처리 과정을

통해서는 해석 모듈에 의해 검증된 자료로부터 잡음 및 모호성을 제거하여 학습에 필요한 훈련 자료들이 획득되며, 규칙을 획득하는 과정에서는 손 제스처의 분할에 필요한 규칙들이 얻어진다. 학습모듈에 대해서는 IV절에서 자세히 소개한다.

### III. 프레임지식에 기반한 해석모듈의 계획제어

본 절에서는 시스템과 사용자와의 상호작용을 계획하고 그에 따른 사용자의 반응을 해석하여 동작을 결정하는 부분인 해석모듈의 계획 수행 과정과 각 단계별 생성된 가설에 대한 검증 방법을 기술한다. 대부분의 컴퓨터 비전시스템은 문제 영역에 대한 사전지식을 이용하여 주어진 문제를 해결하려는 노력을 기울여왔다<sup>[5,6,7]</sup>. 지식에 근거한 시스템들은 지식베이스에 보관된 생성규칙을 이용하여 새로운 가설을 생성하게 되는데 생성규칙은 전제부와 결론부로 구성되며 전제부는 생성된 가설의 검증되어야 할 조건들을 의미하고, 결론부는 검증 결과 다시 생성된 가설을 기술한다. 그리고 지식의 표현 방법 중에서 틀에 짜여진 상황이나 물체를 나타내는 데는 일반적으로 프레임 형식을 많이 이용한다. 그것은 프레임의 구성이 슬롯과 이에 해당하는 슬롯값들로 이루어지면서 프레임의 슬롯값을 계산하기 위하여 프로시저어를 호출할 수도 있고 슬롯값들로서 다른 프레임의 이름을 사용하여 계층구조의 지식도 표현할 수 있기 때문이다<sup>[8]</sup>.

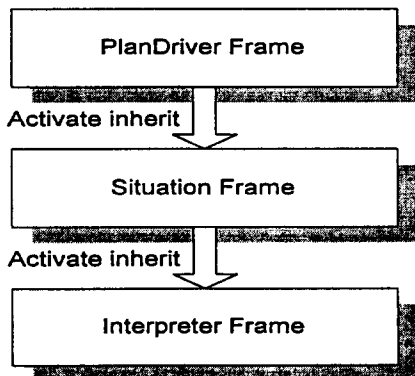


그림 2. 해석모듈을 구성하는 프레임 지식의 계층구조  
Fig. 2. Hierarchy of frame knowledge in interpreter module.

본 논문에서는 생성규칙의 전제부를 구성하는 조건

들을 슬롯에 기술하고 전제부의 검증결과에 따라 생성된 가설을 기술하는 결론부를 후처리기의 *success*와 *fail*로 구분하여 프레임 지식에 가설 생성규칙을 포함시키므로 프레임 지식간에 계층구조를 형성한다. 그림 2는 해석 모듈내에 기술된 프레임 지식간의 계층구조를 보여준다. 그림에서 보듯이 상위 프레임에 의해 하위 레벨의 가설이 생성될 경우, 생성된 가설을 검증하기 위해서 기술된 전처리기들은 상위레벨의 프레임 지식으로부터 검증된 슬롯값을 받아 하위 레벨의 슬롯 조건들을 정의하게 된다.

본 논문에서 사용되는 프레임 지식의 기본 형태는 그림 3과 같다. 그림에서 type은 프레임 지식이 가질 수 있는 값의 범위, 디폴트 값과 예외적인 상황에서의 프레임 지식을 제어할 수 있는 사건을 기술하며 preprocessors는 프레임 지식의 슬롯값들을 구하기 위해 선행되어야 하는 함수들을 기술하고 postprocessors는 가설의 검증 결과에 따라 처리해야 할 후처리기들로서 가설 생성규칙을 의미한다.

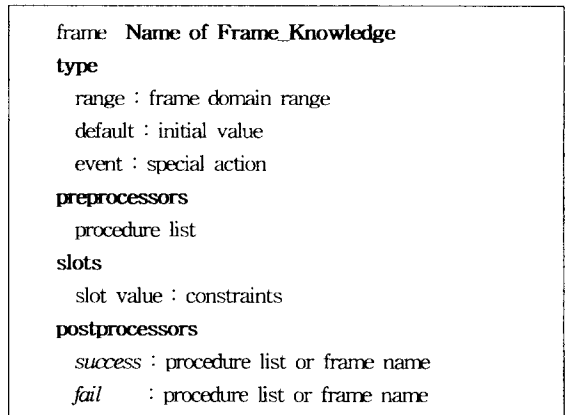


그림 3. 프레임지식의 기본 구조  
Fig. 3. Basic architecture of frame knowledge.

그림 4는 제스처 인식시스템에서 계획기반한 학습시스템의 프레임 지식을 이용하여 훈련자료를 획득하는 과정을 보여준다. 그림에서 보듯이 계획유도 프레임은 사용자의 요구에 따라 학습시스템을 구동하지 않고 기존에 학습된 요소를 가지고 제스처 인식시스템으로 제어될 수도 있고, 정의된 계획에 따라 훈련 자료를 획득하는 과정으로 이동될 수도 있다. 해석모듈의 각 단계별 가설의 생성은 가설 생성규칙에 의해서, 가설의 검증은 가설 검증기에 의해서 수행되지만 계획유도 가설은 제스처 인식시스템이 구동될 때 자동적으로 생성

된다.

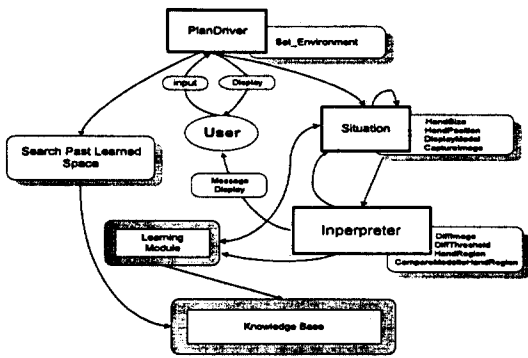


그림 4. 학습시스템의 제어이동  
Fig. 4. Control flow in learning system.

계획유도 프레임(PlanDriver frame)은 어떤 상황을 사용자에게 제공할 지를 결정하는 역할을 한다. 그림 5에서 제시하는 계획유도 프레임은 시스템이 수행할 상황을 두가지로 분류한다. 최적의 훈련자료를 얻기 위한 사용자와의 상호작용의 절차를 계획하는 것과 사용자의 의사가 시스템에 반영되도록 하는 것의 두가지이다. 전자의 경우, 시스템은 정의된 템플리트 중에서 어떤 것을 선택하여 사용자에게 디스플레이 할 지와 카메라로부터 입력된 영상의 분석 결과에 따른 다음 단계로의 전이를 결정하는 절차를 계획하게 되며, 후자는 사용자가 학습과정을 무시하고자 하여 시스템에게 자신의 의도를 전달할 때, 시스템은 사용자의 의사를 받아들여 과거의 학습 내용 중에서 적당한 지식을 탐색하여 제스처 인식을 위한 특징으로 설정하는 경우를 들 수 있다. 이 때 탐색 제약 조건으로 현재의 환경은 과거의 학습 환경과 동일하다는 가정을 사용하게 된다. 계획유도 프레임은 시스템에 대한 사용자의 요구사항이 입력되지 않을 경우에 전처리기의 set\_environment를 수행하여 카메라로 획득될 훈련자료의 위치를 설정하기 위한 사용자 시스템의 환경 정보를 검증한다. 이 정보는 시스템에 의해 정의된 사용자의 가이드용 모델들의 화면상 위치를 조정하기 위해 사용된다. 후처리는 success와 fail 두개의 슬롯으로 구성되며, success는 계획유도 프레임이 활성화되어 생성된 가설의 검증이 성공할 경우에 새로운 가설을 생성하기 위한 슬롯이고, fail은 실패했을 경우에 생성할 가설을 기술한다. 본 논문에서는 가설의 검증 결과 성공적일 경우 사용자에게 상황을 디스플레이하는 과정을 제공하

는 하위 프레임인 상황 프레임을 호출한다.

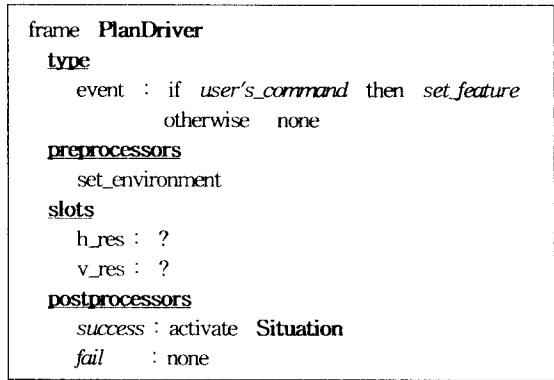


그림 5. 계획유도 프레임의 구조  
Fig. 5. Architecture of PlanDriver frame.

상황 프레임(Situation frame)은 어떤 형태의 템플리트가 화면의 어느 위치에 디스플레이 될 지를 결정하여 사용자에게 제시하고 사용자의 제스처를 카메라로부터 획득하는 역할을 한다. 사용자에게 제시될 템플리트를 선택하는 방법은 다음의 3가지 경우로 나누어 생각할 수 있다. 학습시스템이 시작되어 상황 프레임이 처음 활성화되는 경우에는 디폴트 형태의 템플리트가 선택되고, 학습 모듈로부터 제어가 이동된 경우에는 select\_hand 함수에 의해 이미 정의된 형태의 템플리트 중 하나가 선택되며 해석(Interpreter) 프레임으로부터 제어가 이동될 경우에는 전단계에서 선택된 템플리트가 다시 설정된다. 상황 프레임의 전처리기는 계획유도 프레임에 의해 상속된 화면과 영상의 가로 및 세로의 크기에 대한 정보를 이용하여 템플리트의 가로, 세로 길이를 계산하는 hand\_size, 템플리트의 위치를 계산하는 hand\_position, 템플리트의 형태와 메시지를 화면상에 디스플레이하는 display\_model과 사용자의 동작을 카메라로부터 획득하는 capture\_image로 구성된다. 상황 프레임의 가설의 검증이 성공할 경우, 후처리의 success는 해석 프레임을 활성화한다. 그림 6은 상황 프레임의 지식 형태를 보여준다.

초기의 시스템 사용자는 시스템의 계획에 익숙치 않으므로 제시된 상황에 대하여 정확한 행동을 취하지 못할 수 있다. 그러므로 사용자로부터 획득된 정보가 인식을 위한 학습 자료로 사용하기에 적당한지를 판단하기 위한 작업이 필요하게 되는데 이것은 해석 프레임에 의해 수행된다. 사용자와 시스템간의 상호작용이 정확하게 이루어졌는지는 카메라로부터 획득된 영상을

diff\_image, diff\_threshold, hand\_region의 전처리기를 수행하여 각 슬롯들(slots)의 제약조건(constraints)이 만족되는 가를 조사하여 판단한다. 여기서 슬롯들의 제약조건은 상황 프레임 단계에서 사용자에게 제시된 템플릿의 위치와 크기 정보를 이용하여 정의된다. 전처리의 diff\_image는 카메라로부터 획득된 시간적으로 인접한 2개의 영상프레임에 차연산을 적용하여 차분 영상을 구하는 역할을 하고 구해진 차분영상으로부터 움직임 부분을 검출하기 위한 이진화의 작업은 diff\_threshold에 의해 수행된다. hand\_region은 이진화가 이루어진 영역에 레이블링을 수행함으로써 최소사각영역을 구하는 역할을 한다. 해석 프레임의 활성화가 성공할 경우, 해석 프레임은 사용자의 제스처 영역의 특징을 학습하기 위한 학습모듈을 구동시키고 실패할 경우에는 message\_display에 의해 실패 원인에 대한 메시지를 사용자에게 디스플레이한 후 상황 프레임을 활성화한다. 그림 7은 해석프레임의 지식형태를 보여준다.

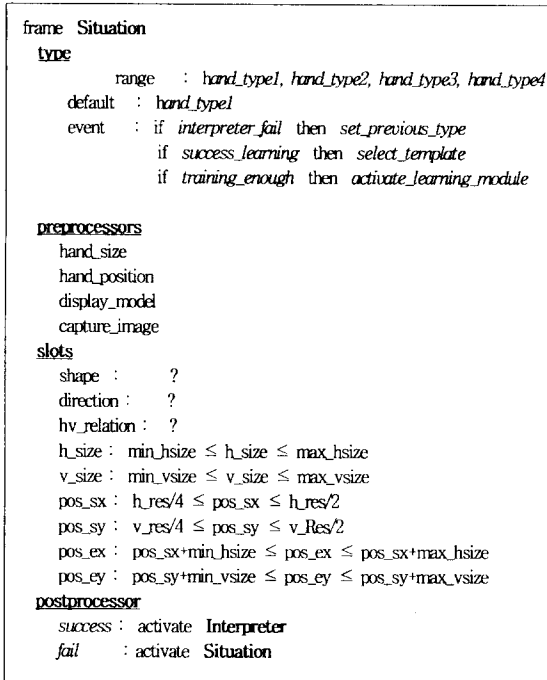


그림 6. 상황 프레임의 구조

Fig. 6. Architecture of Situation frame.

#### IV. 분할요소를 추출하는 학습모듈

학습시스템을 이루는 구성요소 중 전반부에 해당하

는 자료의 전처리과정은 차영상의 영역을 분석하여 잡음이나 부당한 정보를 제거하여 사용자의 제스처 영역에 해당하는 화소들만을 추출하는 역할을 한다.

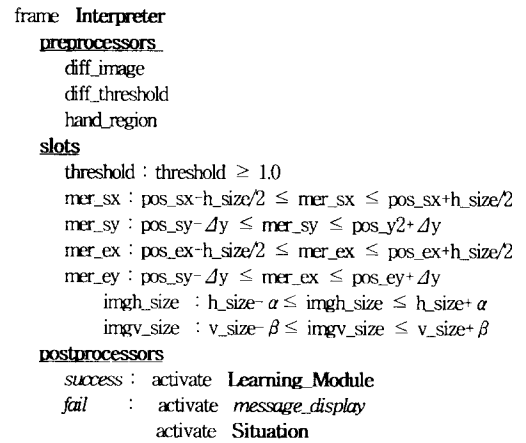


그림 7. 해석 프레임의 구조

Fig. 7. Architecture of Interpreter frame.

전처리의 첫번째 작업은 이진화된 차영상의 레이블링된 영역을 조사하여 사용자 제스처의 움직임 영역이 하나의 레이블로 이루어진 단일한 영역으로 나타나는지 또는 두개의 레이블로 구성된 복합영역으로 나타나는지를 파악하는 것이다. 이렇게 단일영역과 복합영역으로 분석하는 이유는 하나의 물체가 병진운동을 할 때, 차영상은 드러나는 배경영역과 감추어지는 배경영역으로 이루어지며 물체의 특성과 움직임에 따라 하나의 영역을 만들거나 비슷한 크기의 두개의 영역으로 나누어진다는 사실에 기인한다<sup>[9]</sup>.

하나의 레이블로 이루어진 경우에는 움직임 영역과 입력영상의 에지영상에 AND 오퍼레이션을 수행하여 동작에지를 구하고 동작에지들로 연결된 경계의 내부에 포함되는 영역을 제스처 영역에 대한 타당한 화소들로 정의한다. 반면에 두개의 레이블로 이루어진 경우에는 방향정보를 이용하여 레이블 영역의 순서를 정한 후 대응하는 입력 영상의 화소들을 학습을 위한 타당한 화소들로서 정의한다<sup>[10]</sup>. 그림 8은 이러한 전처리 과정을 보여준다.

인간의 시각이 물체를 분별할 수 있는 가장 큰 요소가 되는 것이 칼라이기 때문에 본 논문에서는 제스처에 이용되는 영역의 특징을 칼라로 구분하고자 한다. 칼라를 표현하는 방법은 여러가지 있으나 본 논문에서는 NTSC(National Television System Commit-

tee)에서 규정된 합성칼라 비디오 표준 방식인 YIQ 칼라공간을 이용한다.

성이 많기 때문이다.

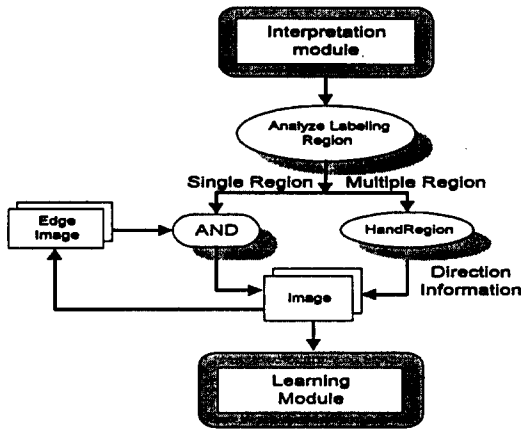


그림 8. 훈련 자료의 전처리 과정  
Fig. 8. Preprocessing of training data.

RGB 칼라 영상을 YIQ 방법으로 표현할 경우 피부 색깔을 갖는 영역이 강화되는데 이것은 YIQ 칼라 공간의 I요소가 -150에서 +150까지의 값으로 정의되는 오렌지색부터 남색까지의 색상을 포함하면서 살색 (skin-color)에 강하게 반응하기 때문이다<sup>[11]</sup>. RGB로 표현된 화소들을 YIQ로 변환하기 위한 방법은 식 (1)과 같다.

$$\begin{aligned}
 Y(x, y) &= 0.299 * R(x, y) + 0.587 * G(x, y) + 0.114 * B(x, y) \\
 I(x, y) &= 0.596 * R(x, y) - 0.275 * G(x, y) - 0.321 * B(x, y) \quad (1) \\
 Q(x, y) &= 0.212 * R(x, y) - 0.523 * G(x, y) + 0.311 * B(x, y)
 \end{aligned}$$

학습 알고리즘은 크게 두 단계로 나누어진다. 첫단계는 현재 계획단계에서 구한 각각의 화소값(I)이 분할에 기여하는 정보량을 계산하고 두번째 단계에서는 현재까지 산출된 모든 화소값에 대한 분할에 기여하는 정보량을 고려하여 색상임계치 정보를 산출한다. 알고리즘 (1)은 첫단계에 대한 처리 과정을 기술하고 있다. 계획 단계에서 추출된 모든 화소들은 손의 색상을 결정하는데 긍정적인 역할을 하지만 정확한 임계치를 결정하기 위해서는  $\alpha$ 에 속하는 화소들에 대하여만 고려한다. 즉, 손의 영역을 표현하는 색상 중에서 적은 수의 화소를 포함하는 화소값은 학습 과정에서 제외된다. 이는 비교적 균일한 색상을 갖는 어떤 물체의 색상 분포에 대한 히스토그램이 그림 9와 같을 경우 평균에서 먼 거리에 있는 화소들은 조명에 의한 잡음일 가능

```

algorithm (I)
input : I of every training pixel, threshold  $\alpha$ 
output : set of (I, contribution)
procedure calculate_contribution
  compute average_I for all pixel;
  do {
    find the near_I of the nearest distance from average_I for all I;
    add the number of pixel with near_I to accumulated_pixel;
  } while (accumulated_pixel <  $\alpha$ );
  find highfrequency for all I;
  compute contribution(I) for all I included accumulated_pixel;
end procedure.
    
```

계획단계에서 획득된 훈련자료인 각 화소들은 식 (1)을 통해서 I값이 계산되어 알고리즘 (1)에 입력된다. 알고리즘 (1)은 입력된 모든 화소들에 대하여 평균 화소값(I)을 구한 후, 평균 화소값을 중심으로 가까운 거리에 있는 화소값의 화소수로부터 전체 화소수의  $\alpha$ 에 해당될 때까지 모두 더한다. 화소값의 기여도(contribution)는 식 (2)에서 보듯이 화소값(I)의 화소수(frequency(I))를 최대화소수(high\_frequency)로 나눈 값이 된다.

$$\text{contribution}(I) = \text{frequency}(I) / \text{high\_frequency} \quad (2)$$

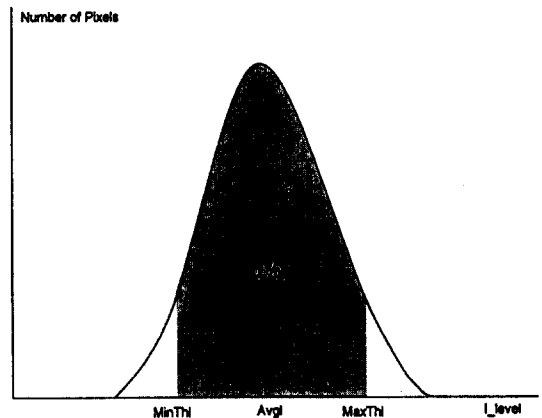


그림 9. 단일 색상을 갖는 물체의 히스토그램  
Fig. 9. Histogram of object with one color.

계획단계에서 제시되는 각각의 템플릿에 따라 추출된 사용자 손의 영역에 대한 화소값의 분포와 화소값의 화소수는 손의 모양과 조명에 대한 상대적 위치로 인해 조금씩 다르게 된다. 이러한 변화를 분할 임계

치에 반영하기 위하여는 각 템플릿에 따라 획득되는 손의 영역의 모든 혼련 화소들에 대해 누적된 분할기여도를 계산해야 한다. 알고리즘 (2)는 알고리즘 (1)에서 산출된 화소값의 분할기여도와 학습시스템의 초기부터 현재까지 누적된 분할기여도를 입력으로 분할을 위한 임계치를 구하는 과정을 기술한다. 학습시스템의 초기부터 현재까지 산출된 모든 화소값(I)에 대한 누적된 분할기여도(accumulated\_contribution(I))와 알고리즘 (1)에서 산출된 화소값의 분할기여도(contribution(I))를 모두 더한 후, 가장 큰 값을 갖는 누적분할기여도로 각각의 누적된 분할기여도를 정규화한다. 정규화된 누적분할기여도의 값이  $\beta$  이상인 것들에 대해 최대화소값과 최소화소값을 구한 후 모든 화소값(I)의 정규화된 누적분할기여도 accumulated\_contribution(I)와 함께 출력한다.

```

algorithm (2)
input : set of (I, contribution) of ith template, threshold  $\beta$ 
output : set of (I, accumulated_contribution)
        lowest_J, largest_J
procedure calculate_accumulation
    add contribution(I) to accumulated_contribution(I);
    normalize accumulated_contribution(I)
    with the largest accumulated_contribution(I) ;
    find lowest_contribution and largest_contribution
    for the normalized accumulated_contribution(I) >  $\beta$ ;
    compute lowest_J and largest_J of lowest_contribution
    and largest_contribution;
end procedure.
  
```

### V. 실험 및 결론

본 절에서는 III절과 IV절에서 기술한 방법을 이용하여 제스처에 이용되는 사용자의 손 영역에 대한 색상 임계치의 학습 내용과 결과를 기술한다. 실험에 사용된 컴퓨터는 IBM-PC/586이며, 사용자에게 제시된 템플릿의 종류는 그림 10과 같이 4가지로 제한하였는데, 그 이유는 사용자 손의 색상 정보를 얻기 위해서는 제시된 4가지 형태로도 충분하기 때문이다. 시스템은 카메라로부터 입력되는 연속한 영상들로부터 2개의 영상을 획득하였다. 사용된 화면의 해상도는 640×480이며, 한 화소가 표현할 수 있는 색상의 범위는 R,G,B 각각이 0에서 255 사이가 되므로 YIQ의 I 요소가 가질수

있는 값의 범위는 -150에서 +150이 된다.



그림 10. 학습을 위한 템플릿  
Fig. 10. Template for learning.

그림 11은 사용자에게 제시되는 템플릿과 메시지의 화면 모습이며, 그림 14, 그림 15는 사용자의 영상이 카메라로부터 그림 12와 같이 입력될 경우 해석단계의 각 프레임 지식에 포함된 가설 생성규칙에 의해 생성된 가설의 상태와 검증된 상태의 프레임 지식을 보여준다. 해석프레임 지식에 의해 영상이 처리된 결과는 그림 13과 같으며, 이 경우에 생성된 해석가설을 검증한 결과가 Img\_Hsize와 Img\_Vsize의 값이 80과 110으로 제약조건  $100 \leq \text{Img\_Hsize} \leq 120, 130 \leq \text{Img\_Vsize} \leq 150$ 의 범위를 벗어나므로 fail이 됨을 그림 15-b에서 볼 수 있다. 그러므로 시스템은 사용자에게 “당신의 손이 템플릿보다 작습니다. 템플릿에 정확히 손을 올려놓으세요.”라는 메시지를 사용자에게 출력하고 상황가설을 생성하게 된다.

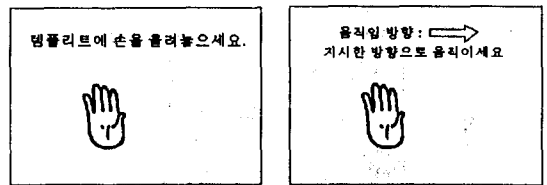


그림 11. 디스플레이 모델  
Fig. 11. Display model.



그림 12. 카메라 입력영상  
Fig. 12. captured image.

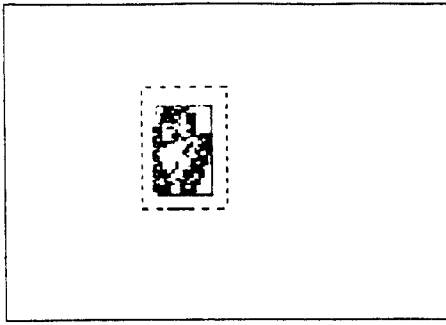
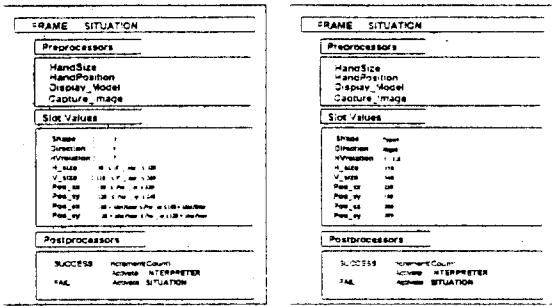
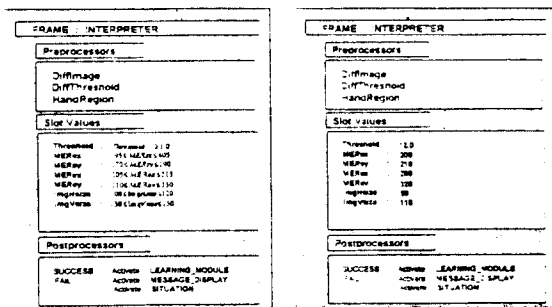


그림 13. 차영상 및 최소사각영역  
Fig. 13. Difference image and MER.



(a) (b)

그림 14. 상황프레임  
(a) 가설 검증 전 (b) 가설 검증 후  
Fig. 14. Situation frame  
(a) Before hypotheses verification  
(b) After hypotheses verification.



(a) (b)

그림 15. 해석프레임  
(a) 가설 검증 전 (b) 가설 검증 후  
Fig. 15. Interpreter frame  
(a) Before hypotheses verification  
(b) After hypotheses verification.

습된 사용자의 최소 색상 임계치, 최대 색상 임계치 정보를 원래 영상에 적용했을 때 검출한 사용자 손의 모습을 보여준다. 실험에서는 정확한 임계치를 추출하기 위해 계획단계에서 획득된 훈련화소 중 85%에 속하는 화소들에 대하여 고려하였고, 누적분할기여도의 임계치  $\beta$ 는 0.3으로 정하였을 때 최소색상임계치 5와 최대색상임계치 36의 값이 추출되었다. 훈련화소수를 선택하는 임계치  $\alpha$ 와 누적분할기여도의 임계치  $\beta$ 에 따라 최소, 최대색상 임계치가 달라지는데 본 실험에서는 경험에 의해서 결정하였다.

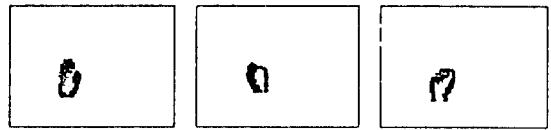


그림 16. 훈련된 색상정보를 이용하여 검출된 손 영역  
Fig. 16. Extracted hand region which used trained color information.

본 논문에서는 컴퓨터 시각에 기반한 손 제스처 인식시스템의 성능 향상을 위해서 시스템과 사용자와의 상호작용에 따라 입력되는 영상을 해석하여 훈련에 적합한 자료들을 획득함으로써 최적의 분할 임계치값을 학습하는 방법을 제안하였다. 템플리트만을 이용하여 사용자 손의 색상정보를 추출하는 방법 [3]에 비해 본 논문에서 제안하는 방법은 분할 정보를 학습하기 위한 훈련 자료를 획득할 때 시스템은 시스템의 일정계획에 따라 사용자와 협력하여 훈련자료를 구하므로 분할을 위한 정확한 자료를 얻을 수 있었다. 뿐만아니라 카메라로부터 입력된 영상을 검증한 결과, 제시된 상황과 적합치 않은 영상일 경우에는 사용자에게 다시 한 번 동일한 상황을 제시하여 새로운 입력 영상을 획득하므로 훈련자료의 정확성을 향상시킬 수 있었다.

본 논문에서는 환경 적응적 손 제스처 인식시스템의 성능 향상을 위해서 필요한 학습시스템의 구조와 각 지식의 표현 방법 및 응용 기법들을 소개하였고, 특히 분할 단계에서 필요한 사전 정보를 학습하는 방법을 제안하였다. 분할을 위한 사전 정보로서 본 논문에서는 색상 정보만을 고려하였다. 실험에서 얻은 결과들을 볼 때, 피부 색상과 다른 배경을 갖는 영상에서는 비교적 좋은 결과를 얻을 수 있었으나, 배경 영역에 피부 색상이 포함될 경우에는 손을 분할할 수가 없었다. 따라서 배경 영역에 피부 색상이 포함될 경우에 손을 분할하

그림 16은 4장에서 기술한 알고리즘을 이용하여 학



기 위한 방법이 앞으로 연구되어야 할 것이다. 또한 본 논문에서 제안하는 시스템에서 훈련자료를 획득하는 과정은 시스템에 의해 정의된 일관적인 계획에 따라 템플릿이 결정되고 사용자가 수행할 동작이 정의되는데 매 단계 학습된 결과에 따라 필요한 템플릿이 선택되며 사용자의 동작이 정의되도록 계획을 수정하는 방법도 앞으로 보완되어야 할 과제이다.

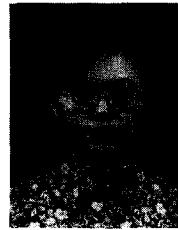
#### 참 고 문 헌

- [1] Davis J., Shah M., "Gesture Recognition", University of Central Florida, Dept. of CS, Tech Rep #CS-TR-93-11.
- [2] Francis K., H. Quek, "Hand Gesture Interface for Human-Machine Interaction", *Proceedings of Virtual Reality Systems*, Fall, pp. 13-19, 1993.
- [3] R. Kjeldsen, J. Kender, "Visual Hand Gesture Recognition for Window System Control", *International Workshop on Automatic Face- and Gesture-Recognition*, Zurich, pp. 184-188, 1995.
- [4] Shinichi Tamura, Shingo Kawasaki, "Recognition of sign language motion images", *Pattern Recognition*, Vol. 21, No. 4, pp. 343-353, 1988.
- [5] 이상원, 고일주, 김계영, 최형일, "이동 물체의 탐지 및 분석을 위한 프레임 시스템", *한국정보과학회 논문지*, 제21권, 제8호, pp. 1480-1489, 1994
- [6] Hidehiro Ohki, Tsutomu Endo, "A Knowledge Base for Natural Scene Understanding", *Proc. of First Korea-Japan Joint Conference on Computer Vision*, pp. 70-76, 1991.
- [7] John K. Tsotsos, John Mylopoulos, H. Dominic Covvey, and Steven W. Zucker, "A Framework for Visual Motion Understanding", *IEEE Trans. on PAMI*, Vol. PAMI-2, No. 6, November, pp. 563-573, 1980.
- [8] 최형일, *컴퓨터 비전 입문*, 홍릉과학출판사, pp. 171-200, 1991
- [9] 김계영, 이은주, 최형일, "차영상 분석에 의한 동작 정보의 추출", *한국정보과학회 논문지*, 제21권, 제8호, pp. 1480-1489, 1994
- [10] Lars Westberg, "Hierarchical Contour-Based Segmentation of Dynamic Scenes", *IEEE Trans. on PAMI*, Vol. 14, No. 9, pp. 946-952, 1992.
- [11] Ying Dai, Yasuaki Nakano, "Extraction of Facial Images from Complex Background Using Color Information and SGLD Matrices", *International Workshop on Automatic Face- and Gesture-Recognition*, Zurich, pp. 238-242, 1995.

#### 저 자 소 개

崔濼一(正會員) 第33卷 第2號 參照

현재 숭실대학교 컴퓨터학부 교수



梁善玉(正會員)

1968년 7월 6일생. 1991년 2월 숭실대학교 전자계산학과 졸업(공학사). 1993년 2월 숭실대학교 대학원 전자계산학과 졸업(공학석사). 1995년 2월 ~ 현재 숭실대학교 대학원 전자계산학과 박사과정 재학중. 관심분야는 컴퓨터 비전, 영상처리, 지식획득 등임.