한국정밀공학회지 제18권 제3호 (2001년 3월)
Journal of the Korean Society of Precision Engineering Vol. 18, No. 3, March 2001.

◆특집◆ Machine Vision과 생산기술

머신비전의 최근 연구동향

한승훈*, 장기정*, 윤국진*, 차준혁*, 노경식**, 권인소*

Recent Developments in Machine Vision Research

Seung Hoon Han*, Gi-Jeong Jang*, Kuk-Jin Yoon*, Joon-Hyuk Cha*, Kyung-Sig Roh**, In So Kweon*

Key Words : Recognition, Motion vision, Video indexing, Projective invariant, Color segmentation, Mosaicing

1. 서론

1970년대 말 인간의 시각기능에 대한 과학적 설명을 시도하면서 시작된 머신비전은, 그 응용에 있어서도 다양한 공정에 적용되었다. 예를 들면, 이차원 영상으로부터 추출된 경계선과 꼭지점 등의 영상특징(image feature)을 이용하여 제품의 양, 불량을 판정하거나 부품의 위치와 종류를 자동으로 인식하는 영상인식기술들이 개발되었다 [1, 2, 3]. 이러한 영상인식기술들은 반도체, 자동차 등의 생산공정 뿐만 아니라, 문자인식, 과일 선별작업, 피자의 품질관리, 의료영상의 인식, 위성영상을 이용한 원격탐사 등에 적용되었다 [4, 5, 6]. 최근 인터넷의 급속한 보급으로 영상인식기술은 멀티미디어 응용, Man/Machine 인터페이스, 지능형 감시(intelligent surveillance), 생체인증 등 복잡하고 다양해진 새로운 응용분야에 적용되고 있다 [7, 8].

이 같은 다양한 응용분야에 적용되어 신뢰성 있는 인식 결과를 얻기 위해서는 머신비전에서 풀어야 할 근본적인 몇 가지 문제가 있다. (1) 영상의 한 픽셀에서의 밝기값은 조명의 종류 및 세기, 카메라의 위치나 자세, 렌즈 광학계의 특성, CCD 센서의 특성, 피사체의 표면 반사특성 등 많은 변수에 의해 결정되며, 그 결과 영상의 밝기는 이 변수들에 민감하게 변화하게 된다. (2) 이차원 영상은 삼차원 물체를 이차원 영상평면에 투사하여 얻어지게 되므로, 이차원 영상을 이용한 삼차원 물체의 인식은 삼차원 물체가 투사되는 방향에 따라 그 형상이나 모습이 변화하게 되어 어려운 문제이다. (3) 대표적인 NTSC 의 경우 초 당 처리해야 할 데이터량은 약 12 Mbytes 로 이를 실시간에 처리하기 위해서는 많은 계산량이 요구된다. (4) 뿐만 아니라 영상인식에는 다양한 종류의 지식(knowledge)이 필요하게 되며, 이를 영상에서 안정되게 추출하는 것은 매우 어려운 문제이다.

이러한 문제를 해결하려는 여러 가지 시도들이 있었지만, 많은 경우 실험실 수준에서의 성공을 보장할 뿐, 이들의 실제 산업현장에의 적용에는 아직도 부족한 면이 많은 것이 현실이다. 이러한 한계를 극복하기 위한 몇 가지 시도들이 있다. 여러 시점에서 잡힌 여러 장의 영상으로부터 삼차원 정보를 추출하여 인식하려는 삼차원 영상인식 [9, 10], 연속으로 입력되는 영상으로부터 움직임 정보를 추출하여 인식하려는 동영상 인식 [11, 12], 조명의 변화에도 변화하지 않는 광도불변량(photometric invariant)을 이용한 영상기법 [13, 14], 인간시각 특성을 이용한 영상인식기법 [15, 16], 투사방향에 따라 변화하지 않는 영상 특징량 중의 하나인 투사불

* 한국과학기술원 전자전산학과 전기및전자공학전공 로보틱스 및 컴퓨터비전 실험실
Tel. 042-869-3465, Fax. 042-869-3410, Email iskweon@kaist.ac.kr
삼차원비전, Mobile robot, Motion vision, Color vision분야에 관심을 두고 연구활동을 하고 있다.
** 삼성종합기술원

변량(projective invariant)을 이용한 영상인식기법 [17, 18, 19] 등이 있다.

본 특집논문에서는 이와 관련하여 지난 수년간 본 실험실에서 개발된 영상인식기술을 소개한다. 2장에서는 새로운 투사불변량에 근거한 한 장의 2-D 영상을 이용한 삼차원 물체인식 [18], 3장에서는 조명이 변화하는 환경에서 강인하게 작동하는 실시간 얼굴추적 기법 [20], 4장에서는 인간시각의 색상감도 특성을 이용한 새로운 영상분할 기법 [16], 5장에서는 멀티미디어 응용의 예로서 카메라의 편집특성을 자동 인식하여 영상데이터 양을 줄이기 위한 비디오분할(video segmentation) 기법 [8], 6장에서는 산업용 비전시스템의 예로서, 연속 입력영상으로부터 움직임을 추정한 후 영상들을 차례대로 합성하여 가스계량기의 숫자를 인식하는 시스템을 기술한다.

2. 한 장의 영상에서의 투사불변량을 이용한 3차원 물체 인식

머신 비전 분야에서 사용되는 대부분의 영상 불변량은 평면에서 평면으로의 대응에 바탕을 두어 왔다. 이러한 평면투사그룹(plane projective group)의 영상 불변량은 깊이 연구되어 왔고 많은 형태는 알려져 있다. 그리고 실제 비전 시스템에 성공적으로 적용되어 왔다[17-19]. 3차원 구조의 영상불변량을 그것의 2차원 투사 이미지로부터 이끌어내는 것은 상당히 어렵고, 이것이 영상 불변량 이론을 비전 분야로 적용하는데 있어서 주요 목표이다.

2.1 투사불변량

Burns [26]은 한 장의 영상으로부터는, 3차원에서 일반적인 위치에 있는 (어떠한 구조를 이루지 않는) 점들에 대한 영상 불변량은 얻어질 수 없다는 것을 보였다. 이 문제의 해(solution)는 다음의 세 가지 부류로 나뉘어 질 수 있다. 첫째는, 두 장의 영상 사이의 에피폴라기하학(epipolar geometry)이 미리 계산되어져 있다는 가정 하에, 두 장의 영상으로부터 영상 불변량을 이끌어내는 것이다[9, 27, 29]. 둘째는, 에피폴라기하학을 계산하지 않고 세 장의 영상으로부터 영상 불변량을 계산하는 것이고 [27, 28], 셋째는, 특수한 구조를 가지는 한 장의 영상으로부터 영상 불변량을 얻는 것이다[21-23]. 이 세 가지의 부류 중 마지막 세 번째는 각 영상의 특징치(feature) 사이의 대응(correspondence) 정보가 필요치 않다.

Rothwell[21]은 두 가지 특별한 구조를 제안하여 그것으로부터 한 장의 영상에서 영상 불변량을 이끌어냈다. 첫번째 구조는 다면체의 꼭지점들을 이용하여 점과 평면 사이의 산술적 계산을 통해 영상 불변량을 얻는 것이고, 둘째는, 양면 대칭(bilateral symmetric) 인 물체에 의한 것이다. 첫번째 구조를 구축하기 위해선 육면체의 꼭지점 중 최소한 일곱 점이 필요하다. 두 번째 경우는 최소한으로 양면 대칭인 여덟 점, 혹은 네 점과 두 직선이 필요하다. Zhu[22, 23]는 이웃하는 두 평면 위의 여섯 점(즉, 두 쌍의 한 평면 위의 네점)으로 이루어진 구조로부터 영상 불변량을 계산하였다. 일곱점이 아닌 여섯점을 이용하기 때문에 Zhu의 방법이 Rothwell의 그것보다 훨씬 덜 제한적이다.

본 연구에서는 Zhu의 방법보다 훨씬 더 일반적인 구조에 대한 새로운 영상 불변량을 제안한다 [18]. 제안된 구조는 한 평면위의 네 점과 임의의 두 점(한 평면 위에 있지 않음)으로 이루어진다. 일반적으로는 이 구조는 에피폴라(epipolar) 구조를 모르는 상태에서도 두 장의 영상을 취하면 영상 불변량을 제공하는데, 본 연구에서는 단지 한 장의 영상을 이용하여 제안된 구조에 대한 영상 불변량 관계식을 이끌어 낸다. 삼차원 물체 인식을 위하여 영상 불변량 관계식과 아울러 기하학적 해싱(geometric hashing)을 이용한 모델 기반 방법을 이용한다.

정규프레임(Canonical frame) 개념을 이용하여 새로운 구조의 영상 불변량을 유도한다. X1~X6 를 물체의 여섯 점이라 하고, x1~x6을 대응되는 여섯 영상점이라 하자. 단, 그림 1 에서 보는 바와 같이 X1, X2, X3, X4 는 한 평면 위의 네 점을 나타내고 X5, 과 X6은 한 같은 평면 위에 있지 않는 임의의 두 점을 나타낸다고 하자. 그러면 물체의 점들과 대응되는 영상점들 사이에는 다음과 같은 평면 방정식으로 표현되는 영상 불변량 관계식이 유도된다.

$$-\frac{(\mathbf{V}_1\times\mathbf{V}_2)}{\|\mathbf{V}_1\times\mathbf{V}_2\|}\bullet\frac{\mathbf{V}_3}{\|\mathbf{V}_3\|}=0 \qquad (1)$$

여기서,

$\mathbf{V}_1=(u_5,v_5,w_5),\ \mathbf{V}_2=(u_6,v_6,w_6),\mathbf{V}_3=(\bar{\alpha},\bar{\beta},\bar{\gamma})$

모든 계산은 정규좌표계에서 표현된 것이다.

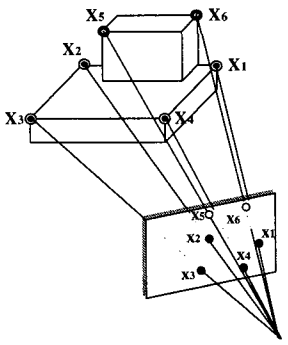


Fig. 1 Projective relation of six points

여섯 점에 대한 정규투영 좌표는 다음과 같이 주어진다.

$$\begin{aligned}
\mathbf{X}_1&=(X_1,Y_1,Z_1,1)\ \rightarrow(1,0,0,0)\\
\mathbf{X}_2&=(X_2,Y_2,Z_2,1)\ \rightarrow(0,1,0,0)\\
\mathbf{X}_3&=(X_3,Y_3,Z_3,1)\ \rightarrow(0,0,1,0)\\
\mathbf{X}_4&=(X_4,Y_4,Z_4,1)\ \rightarrow(\alpha,\beta,\gamma,0)\\
\mathbf{X}_5&=(X_5,Y_5,Z_5,1)\ \rightarrow(0,0,0,1)\\
\mathbf{X}_6&=(X_6,Y_6,Z_6,1)\ \rightarrow(1,1,1,1)
\end{aligned} \qquad (2)$$

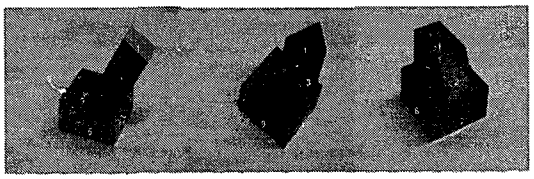
물체 위의 여섯 점과 대응되는 영상점들간의 관계는 다음과 같이 주어진다.

$$\begin{aligned}
\mathbf{x}_1&=(x_1,y_1,1)\ \rightarrow(1,0,0)\\
\mathbf{x}_2&=(x_2,y_2,1)\ \rightarrow(0,1,0)\\
\mathbf{x}_3&=(x_3,y_3,1)\ \rightarrow(0,0,1)\\
\mathbf{x}_4&=(x_4,y_4,1)\ \rightarrow(1,1,1)\\
\mathbf{x}_5&=(x_5,y_5,1)\ \rightarrow(u_5,v_5,w_5)\\
\mathbf{x}_6&=(x_6,y_6,1)\ \rightarrow(u_6,v_6,w_6)
\end{aligned} \qquad (3)$$

앞서 정의된 영상 불변량 관계식으로부터, 물체 위의 점들로부터 얻어진 V3 는 영상에서 얻어지는 V1 과 V2 의 외적(cross product)과 서로 수직이 됨을 알 수 있다. 만약 여섯번째 점 X6 가 (X3, X4, X5)에 의해 만들어지는 평면 위에 있다면, Zhu가 제안한 구조와 같은 구조가 됨을 알 수 있고 동일 평면(coplanar) 조건에 영상 불변량 관계식을 더함으로써 그 구조에 대한 영상불변량을 쉽게 계산할 수 있다.

2.2 실험 결과

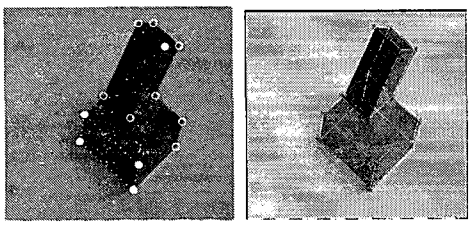
Fig. 2는 제안된 불변량 관계식을 이용한 물체 인식 실험에 사용된 세 개의 모델을 보여주고 있으며, 이 그림에서의 번호는 평면들을 나타내고 있다.



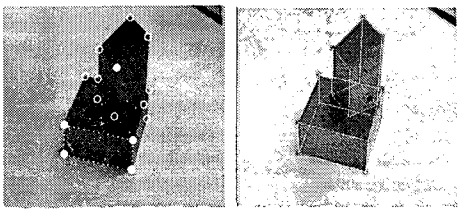
(a) Model I (b) Model II (c) Model III

Fig. 2 Models for recognition

Fig. 3은 실험결과를 보여주고 있다. 왼쪽 영상은 입력영상과 인식에 사용된 코너점들을 보여주고 있다. 오른쪽 영상은 인식결과로 얻어진 모델과 추출된 영상 특징점 사이의 대응점을 이용하여 계산된 모델의 자세를 입력영상에 겹쳐서 보여주고 있다.



(a) For input image I



(b) For input image II

Fig. 3 Recognition results using the proposed invariant relationship.

3. 조명변화에 무관한 얼굴추적

실시간 얼굴 추적 시스템은 HCI (Human Computer Interaction) 부분에 많은 응용 분야를 가지고 있다. 여기서 얼굴 추적은 단순히 컴퓨터의 주의를 사용자의 얼굴에 맞추게 하는 외에 사용자의 얼굴을 인식하기 위한 전제 조건으로서 분할(segmentation)을 수행하기 위하여도 연구되어지고 있다. 이는 인식의 정확도를 높이기 위해서 필수적인 요소이다. 근래 신분 확인, 표정이나 감정 인식, 입 모양 인식(lip-reading)과 같은 인식의 분야에 강인성을 증진시키기 위해서 얼굴 추적을 행하고 있으며 화상회의나 보안시스템에도 널리 적용되고 있다.

3.1 기존의 얼굴추적 기법

현재 다양한 얼굴 추적 알고리즘이 알려져 있다. 가장 간단한 형태는 주어진 배경화면과의 차(background differencing) 를 이용하는 것이다 [7, 30]. 이것은 픽셀 단위의 색상 차이를 이용해서 차영상(difference map)을 만드는데, 주로 주어진 문턱치를 이용해서 이진화한 후 관심 영역을 분리하게 된다. 이는 단지 정적인 카메라에 대해서만 유효하고 배경 조명이 급격히 변하지 않는다는 가정을 만족하여야 한다. 움직이는 카메라에 대해서도 제한적으로 적용이 가능한 차분방법이 있다 [31]. 그러나 이것도 얼굴의 움직임이 없거나 균일한 영역에 대해서는 차영상이 불확실해지는 단점이 있다.

모델 기반 방법 [7, 32, 33] 은 3차원이나 2차원(영상) 공간에서 정의된 얼굴 모델을 기반으로 한다. 추적이 이루어지고 있는 동안 영상정합(image registration)을 통하여 크기, 평행 이동, 회전, 변형, 조명 조건 변화 등의 모델 파라미터들이 계속해서 업데이트 된다. 이러한 방법은 더욱 신뢰성과 정확성을 높일 수는 있으나 계산량이 많아진다는 단점을 가지고 있다. 만약 모델을 단순화시키면 (예를 들어 2차원 평면 상에서 정의) 계산량을 줄일 수는 있지만 일반성의 결여로 인한 추적 강인성의 감소를 피할 수 없다. 칼라 기반 방법 [34-37] 에서 얼굴은 칼라 공간에서 단일분포(unimodal distribution)로 가정할 수 있다. 따라서 칼라 모델은 간단한 통계학적 파라메타(평균, 표준편차)로 표현 가능하고 때로는 칼라 히스토그램으로 표현된다. 이 방법은 비교적 얼굴 방향이나 부분적인 가려짐에 대해서 강인성을 보이며 알고리즘이 간단하고 속도가 빠르다는 장점이 있으나 조명 조건의 변화에 취약하다는 단점을 가지고 있다. 따라서 이러한 문제점을 극복하기 위하여 칼라 모델 자체를 반복필터(recursive filter)를 이용하여 업데이트 시키거나 조명 조건을 표현하는 추가적인 파라메타를 사용하기도 하였다. 그러나 급격한 조명 변화에 대해서는 그다지 만족스러운 결과를 보이지 않는다.

3.2 적응형 컨덴세이션 기법

본 연구에서는 컨덴세이션(Condensation) 알고리즘 [38, 39] 에 기초한 새로운 강인한 얼굴 추적 시스템을 제안한다. 얼굴 칼라를 표현하기 위해서는 2차원 색도칼라공간(chromatic color space)에서 정의하고, 급격한 조명 변화에 대한 적응성과 다양한 얼굴 칼라에 대한 범용성을 확보하기 위하여 적응 컨덴세이션 방법을 새롭게 제안한다.

본 연구에서 제안한 얼굴 추적 시스템에서는 추적 대상의 상태(얼굴의 위치)를 알기 위해 예상되는 위치에 샘플들을 뿌려놓고 지정한 색상척도가 얼마나 주어진 모델(얼굴색)과 가까운지를 수치화한다. 샘플들은 수치화된 값의 크기에 따라 번식과 소멸을 거듭하며 이들이 계속 확률이 높은 곳(얼굴색과 유사한 위치)으로 전파해 가도록 함으로써 추적의 효과를 거둔다. 또한 동적인 특성을 미리 학습시켜 놓아 현재 얼굴의 동적 특성에 맞는 다음 프레임에서의 위치를 예측하여 샘플들을 뿌려줌으로써 시스템의 강인성을 높여주고 있다. Fig. 4는 얼굴 추적 단계에 따른 개략적인 샘플의 전파를 보여주고 있다.

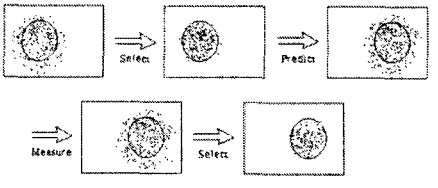


Fig. 4 Sample propagation of the Condensation algorithm.

구체적으로 제안된 방법은 기존의 컨덴세이션 알고리즘이 갖는 한계성을 두 가지 관점에서 개선한다. 첫째 추적 알고리즘에서 참조할 얼굴 칼라 모델을 추적하고자 하는 얼굴의 칼라에 적응시킨다. 이를 위하여 컨덴세이션을 이용하여 목표물을 추적하듯 칼라 공간에서 얼굴 칼라의 영역을 적응 컨덴세이션 알고리즘을 이용하여 추적하도록 한다. 둘째 얼굴의 대략적인 분할을 계산량의 증가가 거의 없이 샘플들이 뿌려질 탐색 영역의 넓이를 결정하는 파라메타의 자동 조정으로 수행할 수 있도록 한다. 이를 통해 추적하고자 하는 얼굴의 크기에 비례하는 적절한 탐색 영역을 결정함으로써 주변 잡음의 영향을 덜 받도록 한다.

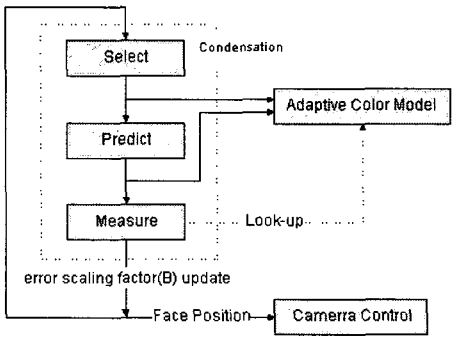


Fig. 5 Diagram of the proposed tracking system using the adaptive Condensation algorithm.

본 연구에서는 제안된 알고리즘을 Pentium III 500Mhz PC에서 320*240크기의 영상을 사용하여 구현하였다. 제안된 알고리즘은 상대적으로 매우 작은 계산량을 필요로 한다. 프래임그래버(frame grabber)로 영상을 잡는 시간을 제외한 순수 처리 시간만을 따졌을 때 얼굴 추적에 충분하다고 생각되는 1000 샘플에 대해서 초당 70개의 영상을 처리 가능하다.

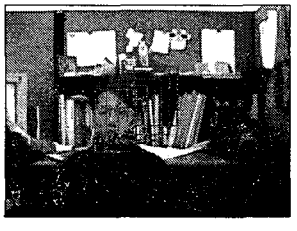


Fig. 6 Tracking result by the proposed adaptive Condensation algorithm.

4. 인간시각특성을 이용한 영상분할

영상 분할은 분할이 일어나는 공간(domain)에 따라 크게 영상 공간을 분할하는 방법과 영상으로부터 얻어진 특징량을 이용하여 특징 공간을 분할하는 방식으로 나누어 볼 수 있다 [40]. K-means 클러스터링 기법은 특징 공간을 이용한 영상 분할 기법 중 대표적인 기법으로 많은 데이터 집합을 최적의 클러스터로 나눌 수 있는 알고리즘으로 알려져 있다 [41]. 하지만 이 방식을 이용한 기존의 방법들은 인간의 시각 특성을 고려하지 않았기 때문에 인간의 시각에 부합하지 않는 분할 결과가 얻어지게 된다. 인간의 시각은 색상 값이 그 색상 공간에서 어디에 위치하는지, 그리고 각 화소가 어떠한 영역에 존재하는지에 따라 그 색상에 대한 인지 감각이 달라진다. 따라서 인간의 이러한 시각 특성을 고려한 클러스터링을 통해 최적의 분할 결과를 얻을 수 있다.

4.1 색상 공간 선택

색상 공간에서 두 색상 간의 차이를 인간이 인식하는 차이에 부합하도록 나타내기 위해 유클리디안 거리가 인간의 시각 차와 유사하다고 알려진 CIELab 색상 공간을 사용한다. CIELab 색상 공간에서 두 색상 값의 차는 다음과 같이 모델링 된다.

$$D(c(i,j),c(x,y)) = 1 - \exp\left(-\frac{E(c(i,j),c(x,y))}{\gamma}\right) \quad (4)$$

이때 $c(i,j)$와 $c(x,y)$는 두 화소를 나타내며, $E(c(i,j), c(x,y))$는 두 화소의 색상 사이의 유클리디안 거리를 나타낸다.

4.2 가중치 계산

Wandell [15]에 따르면 인간의 시각은 색상 패턴의 변화가 적을수록 높은 인지력을 보이기 때문에, 각 화소의 색상에 주변의 색상 패턴 변화에 따른 가중치를 부여하여야 한다. 이는 인간의 시각 특성을 클러스터링 과정에서 고려해 주기 위해서이다. 이를 위해서 먼저 화소 $c(i,j)$ 주변의 색상 변화를 국부 마스크 $\Omega(i, j)$와 가우시안 커널을 이용하여

다음과 같이 측정한다.

$$\varphi(i,j)=\sum_{x,y\in\Omega}G_{\alpha}\left(1-\exp\left(-\frac{E(c(x,y),c(i,j))}{\gamma}\right)\right) \quad (5)$$

이 값이 높을수록 주변에 많은 색상 패턴 변화가 있음을 의미한다. 따라서 이를 이용하여 각 화소가 갖는 색상의 가중치를 다음과 같이 부여한다.

$$w(\varphi(i,j))=\mathrm{k}\cdot\exp\left[-\frac{1}{2\sigma^2}\left(\frac{\varphi(i,j)}{\iint_{x,y\in I}\varphi(x,y)dxdy}\right)^2\right] \quad (6)$$

이때 k는 정규화 상수이고 이 가중치 값은 0과 1 사이의 값을 갖는다.

4.3 클러스터 수의 결정

영상 내에 복잡한 색상 패턴이 존재할 경우에는 인간의 지각 능력이 떨어지기 때문에 적은 수의 클러스터를 사용하는 것이 효율적이다. 반대로 균일한 영역이 많을 경우 인간 시각이 민감하게 작용하므로 많은 수의 클러스터를 사용해야 한다. 따라서 앞서 구해진 각 색상의 가중치를 이용하면 최적의 클러스터 수를 구할 수 있다.

$$K=M\cdot\frac{\iint_{x,y\in I}w(\varphi(x,y))dxdy}{N_I} \quad (7)$$

여기에서 N_I는 영상 내의 전체 화소 수를 나타내고 M은 사용 가능한 최대의 클러스터 수로 영상 분할의 레벨을 결정한다.

4.4 클러스터링

앞서 정의된 가중치와 클러스터의 수를 이용하여 실제로 클러스터링을 수행한다. 먼저 전체 에너지 함수를 다음과 같이 정의한다.

$$J=\sum_{i=1}^{K}\sum_{C\in S_i}\Theta(C)\left\|C-\bar{C}_{S_i}\right\|^2 \quad (8)$$

이때 클러스터들의 초기 중심값은 전체 데이터의 중심에 임의의 노이즈를 발생시켜 결정해준다. 다음과 같이 클러스터와 그 중심값을 반복적으로 바꾸어주면서 클러스터링을 수행한다.

$$C\in S_i^t \text{ if } \left\|C-\bar{C}_{S_i}^t\right\|^2<\left\|C-\bar{C}_{S_j}^t\right\|^2 \quad (9)$$

$$\bar{C}_{S_i}^{t+1}=\frac{\sum_{C\in S_i^t}w(C)\times C}{\sum_{C\in S_i^t}w(C)} \quad (10)$$

4.5 분할 결과

Fig. 7은 영상 분할의 실험 결과를 보여준다. 이때 사용된 클러스터의 수는 22개인데, 같은 수의 클러스터를 사용한 경우 가중치를 사용한 방법이 더 효과적으로 영상을 분할함을 알 수 있다.

(a) input image

(b) weighting value

(c) result without weighting

(d) result with weighting

Fig. 7 Segmentation result for the flower garden image.

5. 비디오 이벤트 인식

방대한 멀티미디어 정보가 인터넷을 통해 사용자에게 공급되고 있다. 최근에는 단순한 문자 정보뿐만 아니라, 시각적인 자극을 가져올 수 있는 영상 및 동영상의 보급이 점차 확산되고 있다. 그러나, 이러한 정보를 단순한 키워드에 의해 찾아내는 데는 이미 한계를 드러내고 있다. 따라서 복잡하고 방대한 영상을 효과적으로 검색하기 위해서는 영상이 가지고 있는 내용 (content)에 대한 색인 및 추출이 이루어져야 한다. 본 장에서는 동영상에 존재하는 객체의 움직임 정보를 이용하여 사건 (event) 및 하이라이트 (highlight)에 대한 인식을 수행한다. 이를 통해 사용자의 질의에 효과적으로 대응하여, 그에 해당하는 부분을 제공할 수 있게 된다. Fig. 8는 본 장에서 제시한 비디오 사건 인식 시스템을 나타낸다.

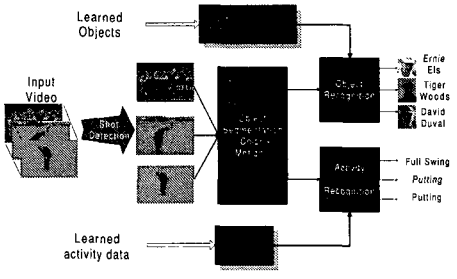


Fig. 8 Overview for recognition of events in video

비디오는 여러 개의 장면으로 구성되어 있기 때문에 장면 전환 검출 (shot boundary detection)을 통하여 모든 장면을 분리해야 한다. 각 장면에 대해 객체 분할 (object segmentation)을 적용하여 추출된 객체에 대해 행위 인식 (activity recognition) 및 객체 인식 (object recognition)을 수행하여 장면에서 일어나고 있는 사건을 분석한다. 본 장에서는 장면 전환 검출 기법과 행위 인식 방법을 소개하고 이를 골프 경기에 적용한 결과를 보인다.

5.1 장면 전환 검출 (shot detection)

비디오 상에서 장면들은 컷 (cut), 페이드 (fade) 및 디잘브 (dissolve) 등과 같은 편집 효과에 의해 연결되어 있다. 영상이 가지고 있는 카메라 또는 객체의 움직임 및 노이즈로 인해 장면의 경계를 정확히 추출하는 것이 어렵다. 움직임 (motion)에 의한 두 영상간의 변화와 편집 효과에 의한 변화의 모호성을 해결하기 위해 [42]에서 제안한 방법을 사용한다. 편집 효과가 가지고 있는 구조적 정보 (structural information)와 두 프레임 간에 칼라 및 밝기 변화 (frame difference)에 대한 확률 분포를 이용하여 장면의 경계를 효과적으로 찾아낸다.

Fig. 9는 한 골프 비디오 세그먼트에 대한 프레임 변화 (FD)와 이에 저대역(low-pass) 필터를 적용한 결과(FFD)를 보여준다. FD에서는 카메라 및 객체의 움직임으로 인해 편집효과를 판별하기가 어렵지만, FFD에서는 편집 효과의 구조적 특징 (컷의 경우 사각형 모양, 디잘브의 경우 삼각형 모양)이 분명히 드러남을 볼 수 있다. 본 장면 전환 검출 기법은 90% 이상의 높은 정확도를 보였다.

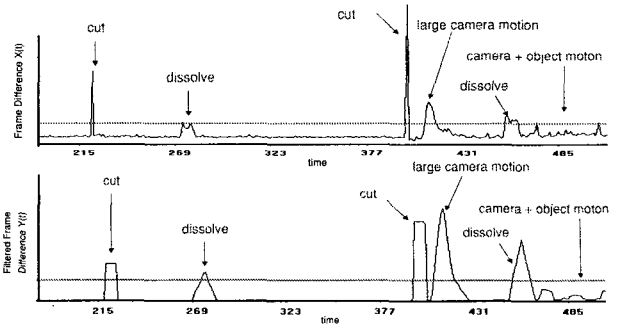


Fig. 9 Detection of the camera editing effect for a golf video using Frame Difference (upper) and Filtered Frame Difference (lower).

5.2 행위 인식 (activity recognition)

비디오 데이터는 문자 (text), 음성 (voice), 객체 (object), 움직임 (motion) 및 사건 (event), 분위기 (semantic mood)등 수많은 정보를 가지고 있다. 무엇보다도 비디오는 정지 영상에서 좀처럼 볼 수 없는 동적인 사건이 존재한다. 동영상에 담겨진 객체나 카메라의 움직임은 하이라이트나 사건을 구성하는데 결정적인 역할을 한다. 특히 스포츠 경기에는 다양한 사건 (event)이 존재하는데, 사람들은 축구 경기의 득점 장면이나 골프경기의 퍼팅 장면과 같

은 하이라이트만을 원하는 경우가 있다. 따라서 사용자가 원하는 장면이 담긴 비디오를 추출해 내기 위해서 객체의 행위 분석 및 인식이 필요하다.

행위(activity) 인식은 여러 문제점을 가지고 있어 다음과 같은 연구주제를 만들어 낸다. 첫째, 장면검출(shot detection)을 통하여 얻어진 각 장면은 다중 행위(multiple activities)와 서로 다른 장면의 길이를 가지고 있다. 둘째, 같은 행위라 할지라도 조금씩 다른 방법으로 수행자에 따라 행위가 다르게 나타나고 그 행위시간 역시 달라진다. 셋째, 행위의 시작과 끝을 알아내기가 쉽지 않다. 넷째, 행위 속도가 비 선형적이다. 마지막으로 움직임 궤적(motion trajectory)은 카메라의 관측 위치와 객체까지의 거리에 따라 변한다. 본 시스템은 위의 문제점을 해결하기 위해 다음 그림과 같은 행위인식 방법을 제안한다.

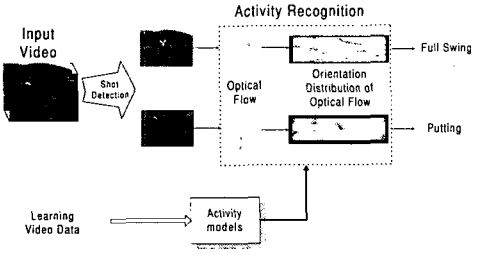


Fig. 10 Activity recognition of golf swings

행위(activity)는 각 시간에서 측정된 관측치의 시간상의 궤적 (temporal trajectory)이라고 정의할 수 있다. 그 관측치는 움직임의 광류 (optical flow)에 기초를 두고 있는데, 광류의 방향 분포 (orientation distribution)와 크기 (magnitude)를 가지고 궤적을 만들어 낸다. 이 궤적이 골프의 스윙을 분류하여 인식하는 데 사용된다.

각 행위의 궤적을 만들어 내기 위한 특징 량인 광류의 방향 분포 (Orientation Distribution of Optical Flow, 이하 ODOF)는 다음과 같이 정의된다.

$$A_i[\theta_l] = \sum_{\{k|\theta=\theta_l\}} (m_i[k])^2 \qquad (11)$$

여기서, $m_i[k]$는 화소(pixel) k에서 광류의 크기 (normalized amplitude), θ_l은 양자화된 방향 (quantized angle)을 의미한다. 즉, ODOF는 각 방향 (orientation angle)에서 광류 크기 (magnitude)의 제곱의 합으로 정의된다. 제곱을 취한 이유는 각 특정한 행위에서 주된 움직임의 영향을 극대화시키기 위함이다. ODOF는 시간에 걸쳐 각 행위에 해당하는 특정한 궤적 패턴을 보이게 된다. 이 패턴은 장면이나 행위의 지속 시간에 따라 크게 변하지 않고, 정규화 된 크기를 사용하기 때문에 카메라와 객체사이의 거리 의존도가 작다. 그렇지만 같은 동작일지라도 관찰자의 위치에 따라 다른 ODOF를 보이기 때문에, 서로 다른 클래스로 분류한다. 다시 말하면, 관찰자의 시점도 인식이 가능하다. 다음 그림은 골프 스윙과 관찰 지점에 따른 ODOF의 궤적을 보인 것이다. X축은 각도 (0~360도)를 의미하고 Y축은 시간(frame number)을 의미한다. 관찰자의 시점은 선수의 앞 (front), 뒤 (behind), 그리고 목표지점(target)이다.

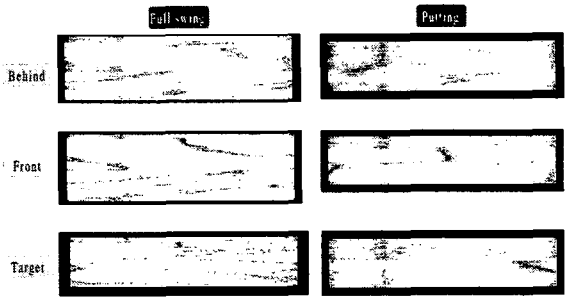


Fig. 11 Trajectory patterns of ODOF for six golf swing activities: full swing and putting from three camera directions.

Table 1은 골프 스윙에 대한 인식 결과를 보여준다. 각 표에서 대각선에 해당하는 부분이 행위 및 관찰자의 시점이 모두 정확히 인식된 결과이다. 관찰자의 시점을 고려하지 않는다면 풀 스윙 (full swing)의 경우 10/11, 즉 91%의 성공률을 보인다.

Table 1 Activity detection results for golf swings

Full Swing	Behind	Front	Target
Behind	4		
Front	3	2	1
Target			1

Putting	Behind	Front	Target
Behind			1
Front	1	2	1
Target			

6. 가스계량기의 숫자인식

가스 계량기의 보정에 있어서 보다 신속하고 정확한 보정을 위해 성능 검사 장치의 자동화가 필요하다. 현재 컴퓨터 비전을 이용해서 가스 계량기의 숫자 인식 시스템을 개발하기 위한 많은 노력이 있지만, 정밀도와 자동화 측면에서 만족스럽지 못한 상태이다. 따라서 기존의 가스 계량기를 그대로 사용하면서 그 정보를 자동으로 신속, 정확하게 처리할 수 있는 가스 계량기의 숫자 인식 시스템이 필요하다. 본 연구에서는 영상합성 방법과 신경 회로망을 이용한 새로운 숫자 인식 방법을 제안하고, 실제 현장에서 사용되는 가스 계량기에 대해서 실시간으로 숫자를 인식한 결과를 제시하고자 한다.

6.1 숫자 인식 방법

가스 계량기의 숫자들은 마지막 숫자를 제외하고는 불연속적으로 변하기 때문에, 일반적으로 숫자 인식에 많이 사용되는 신경 회로망을 사용하여도 큰 무리가 없다. 신경 회로망을 이용한 숫자 인식을 위하여 다층퍼셉트론(multi-layer perceptrons) 회로망(network)을 사용하였고, 망의 학습을 위해 에러역전파(error back propargation) 방법을 이용하였다 [43].

가스 계량기의 숫자 인식 시스템의 성능을 떨어뜨리는 가장 큰 요인은 마지막 숫자가 연속적으로 회전하는데 있다. 지금까지 가스 계량기의 숫자 인식을 위해서 신경 회로망을 주로 사용하였는데 이러한 방법만으로는 마지막 숫자를 정확하게 인식하기는 불가능하다. 따라서 보다 신뢰도가 높은 마지막 숫자의 인식을 위해서 영상합성(mosaicing) 방법을 이용한다. 영상합성 방법은 연속된 입력영상 사이의 움직임을 추정하여 파노라마 영상을 획득한 후 저장한다. 저장된 완전한 숫자와 실시간으로 입력되는 가스 계량기의 숫자를 템플릿 매칭을 이용하여 서로 부분 비교하여 인식한다. 이렇게 제안된 방법을 사용하면 가스 계량기의 마지막 숫자에 대해서 소수점 단위의 정확한 값을 인식할 수 있다.

6.2 숫자 인식 결과

실제 시스템을 구현하는데 있어서 주변의 조명 변화는 가스 계량기의 숫자 인식에 커다란 제약 조건이 된다. 이를 해결하기 위해서 매번 입력 영상을 받을 때 각 숫자에 대해서 반복최적문턱치(iterative optimal threshold) 방법을 적용하여, 각 숫자를 인식하기 이전에 이진화한다. 이러한 전처리 과정을 거침으로 해서 제안된 시스템은 조명이 변하는 환경에서도 강건한 숫자 인식이 가능해 진다. Fig. 12는 영상합성 방법을 이용하여 얻은 가스 계량기의 마지막 숫자의 파노라마 영상이다. 직선은 각 숫자의 중심을 나타내며, 그 옆의 숫자는 템플레이트 매칭에 의하여 인식된 숫자를 나타내고 있다.

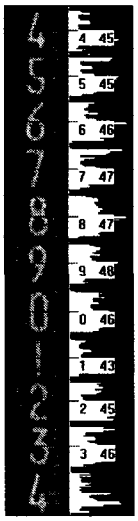


Fig. 12 Panorama Image

구성된 파노라마 영상을 이용하여 3시간 동안 실시간으로 얻은 영상에 대한 실험결과, 마지막 숫자의 인식률은 Table 2 와 같다.

Table 2 Recognition rate of the last digit.

총 횟수	인식 횟수	오인식 횟수	인식률
10623	10611	12	99.89%

위의 결과로부터 제안한 방법이 가장 어려움이 많았던 가스 계량기의 마지막 숫자 인식에 아주 효과적임을 확인할 수 있다.

최종적으로 가스 계량기의 모든 숫자를 변화하는 조명 하에서 인식한 결과의 한 예가 Fig. 13에 나타나 있다. 마지막 숫자의 경우 보여진 바와 같

이 4와 5 사이에 걸쳐 있어서 기존의 템플레이트 매칭이나 신경회로망 방법으로 인식이 어렵지만, 제안된 방법은 영상합성을 이용하여서 정확하게 숫자인식을 하고 있음을 알 수 있다. 특히, 조명 변화가 존재하는 경우에도 제안된 방법은 비교적 강인하게 작동하고 있음을 알 수 있다.

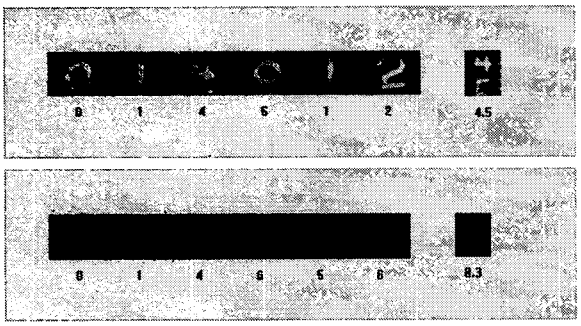


Fig. 13 Snap shots of recognition results for input images taken under bright illumination (upper) and dark illumination (lower).

구현된 시스템은 1초 주기로 입력 영상을 샘플링하면서 가스 계량기의 숫자를 변화하는 조명 하에서 0.1 미만의 오차를 가지고 99.9% 이상의 인식률로 가스 계량기의 숫자를 인식할 수 있다. 이러한 성능은 기존에 이용되던 가스 계량기의 성능 검사 시스템에 비해 탁월한 것으로 현장에서 유용하게 사용될 수 있을 것으로 기대된다.

7. 결론

머신비전 기술은 지금까지 주로 생산공정의 자동화와 측정 등에 많이 응용되어 왔다. 이러한 경향은 최근 인터넷의 급속한 보급으로 인하여 여러 가지 다양한 분야에 응용되기 시작하였다.

최근 영상인식 분야의 연구에서 많은 주목을 받는 중요한 주제는 투사(projection)에 의하여 발생하는 영상의 모호성(ambiguity) 문제의 해결, 조명이 변화하거나 복잡한 환경에서도 작동할 수 있는 머신비전 기술, 많은 양의 영상데이타를 효과적으로 다루기 위한 머신비전 기술의 개발 등이다. 본 특집 논문에서는 머신비전의 이러한 최근 연구 동향을 본 실험실에서 수행한 연구를 중심으로 살펴보았다. 물론, 본 특집에서 다루지 못한 많은 중요한 영상인식에 관련된 다른 결과와 연구들이 있음을 밝히는 바이다.

후 기

본 특집논문이 있기 까지는 저자들 이외에도 로보틱스및컴퓨터비전 실험실의 모든 구성원들의 도움이 매우 컸음을 밝히며, 본 실험실에서 함께 생활했던 졸업생들께도 감사를 드린다. 또 한, 본 연구의 수행을 위한 한국과학재단, HWRS-ERC, 삼성종합기술원, 서울도시가스(주)의 재정적 후원에도 감사를 드린다.

참고문헌

1. K. Ikeuchi and T. Kanade, "Automatic Generation of Object Recognition Programs," Proceedings of the IEEE, 76(8), pp. 1016-1035, 1988.
2. R. Katsuri and R. C. Jain, "Computer Vision: Advnaces and Applications," IEEE Computer Society Press, Las Alamitos, California, 1991.
3. D. Lowe, "Three-dimensional Object Recognition from Single Two-dimensional images," Artificial Intelligence, 31(3), pp. 355-395, 1987.
4. R. Jain, R. Katsuri, B. G. Schunck, "Machine Vision,", McGraw Hill, 1995.
5. L. O'Gorman and R. Katsuri, "Document Image Analysis," IEEE Computer Society Press, Los Alamitos, California, 1995.
6. F. F. Sabins, "Remote Sensing: Principles and Interpretations," W. H. Freeman, New York, 1987.
7. S. J. McKenna and S. Gong, "Tracking Faces," Proc. International conference on Face and Gesture Recognition, pp. 271-276, 1996.
8. S. H. Han and I. S. Kweon, "Shot detection combining Bayesian and structural information," SPIE: Storage and Retrieval for Media Databases 2001, San Jose, CA, January, 2001.
9. O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?," Proceedings of the 2nd ECCV, Santa Margherita, Italy, G.Sandini, Ed., Springer-Verlag, pp. 563-578, May, 1992.

10. D. Lee and I. Kweon, "A Novel Stereo Camera System by a Biprism," IEEE Trans. on Robotics and Automation, 16(5), pp. 528~541, 2000.
11. J. Weber and J. Malik, "Robust Computation of Optical Flow in a Multi-scale Differential Framework," International Journal of Computer Vision, 14(1), pp. 67-81, 1995.
12. J. Ha and I. Kweon, "Robust Direct Motion Estimation consdiering Depth Discontinuity," Pattern Recognition Letters, 11(21), pp. 999~1011, 2000.
13. Y. Moon and I. Kweon, "Hybrid Hierarchical Motion Estimation under Illumination Variations," Electronics Letters, 36(6), pp. 518~520, 2000.
14. J. Kim, C. Kim, Y. Seo, and I. Kweon, "Color Indexing using Chromatic Invariant," Pattern Recognition, 20(1), 2000.
15. B. A. Wandell, "Color Appearance: the Effects of Illumination and Spatial Patterns," Proc. Nat. Acad. Sci., USA, Vol. 90, pp. 1494-1501, 1993.
16. K. Yoon, C. Kim, Y. Seo and I. Kweon, "Moving Object Segmentation based on Human Visual Sensitivity," Proceedings of IEEE Int. Workshop on Biologically Motivated Computer Vision, pp. 62~72, 2000.
17. D. Forsyth et al., "Invariant Descriptors for 3-D Object Recognition and Pose," IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(10), pp. 971-991, 1991.
18. K. Roh and I. Kweon, "3D Object Recognition using a New Invariant Relationship by Single View," Pattern Recognition, 33(1), 2001.
19. K. Roh and I. Kweon, "2D Object Recognition using Invariant Contour Descriptor and Projective Refinement," Pattern Recognition, 31(4), 1998.
20. G. Jang and I. Kweon, "Robust Real-time Face Tracking using Adapative Color Model," Proceedings of International Symposium on Mechatronics and Intelligent Mechanical System, pp. 72~77, 2000.
21. C. A. Rothwell, D. A. Forsyth, A. Zisserman, and J. L. Mundy, "Extracting Projective Structure from Single Perspective Views of 3D Point Sets," Proceedings of the 4th ICCV, Berlin, Germany, pp. 573-582, May, 1993.
22. Y. Zhu, L. D. Seneviratne and S. W. E. Earles, "A New Structure of Invariant for 3D Point Sets from a Single View," Proceedings 12th ICRA, Nagoya, Japan, pp. 1726-1731, 1995.
23. Y. Zhu, L. D. Seneviratne and S. W. E. Earles, "Three Dimensional Object Recognition Using Invariants," Proceedings 12th IROS, Pittsburgh, USA, pp. 354-359, 1995.
24. J. L. Mundy and A. Zisserman, Eds., "Geometric Invariance in Computer Vision," MIT Press, Cambridge, Massachusetts, USA, 1992.
25. M. H. Brill and A. B. Barrett, "Closed-Form Extension of the An-harmonic Ratio to N-Space", Computer Vision Graphics Image Process, Vol. 23, pp. 92- 98, 1983.
26. J. B. Burns, R. S. Weiss, and E. M. Riseman, "The Non-existence of General-case View-invariants," Geometric Invariance in Computer Vision, J. L. Mundy and A. Zisserman, editors, MIT Press, Cambridge, Massachusetts, USA, 1992.
27. L. Quan, "Invariants of Six Points from 3 Uncalibrated Images," Proceedings of the 4th ECCV, Stocholm, Sweden, pp. 459-470, 1994.
28. L. Quan, "Invariants of Six Points and Projective Reconstruction From Three Uncalibated Images," IEEE Transactions on PAMI, Vol. 17, No. 1, pp. 34-46, January, 1995.
29. E. B. Barrett and P. M. Payton, General Methods for Determining Projective Invariants in Imagery, CVGIP: Image Understanding, Vol. 53, No. 1, January, pp. 46-65, 1991.
30. C. R. Wren, A. Azarbayejani, T. Darrell, A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1997, Vol. 19, No. 7, pp. 780-785
31. C. Wang and M. S. Brandstein "A Hybrid Real-Time Face Tracking System," Proc. Acoustics, Speech, and Signal Processing, Volume 6, 1998.
32. G. D. Hager and P. N. Belhumeur, "Real-time tracking of image regions with changes in

geometry and illumination," CVPR, pp. 403-410, 1996.

33. M. La Cascia, S. Sclaroff, V. Athitsos, "Fast, Reliable Head Tracking under Varying Illumination-An Approach Based on Registration of Texture-Mapped 3D Models," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 22, No. 4, April 2000.
34. J. Yang, A. Waibel, "A Real-Time Face Tracke r," Proceeding of WACV, pp. 142-147, 1996.
35. M. J. Jones, J. M. Rehg, "Statistical Color Models with Application to Skin Detection," Proc. CVPR, pp. 274-280, 1999.
36. R. J. Qian, M. I. Sezan, K. E. Matthews, "A Robust Real-Time Face Tracking Algorithm," ICIP, Vol. 1, pp. 131 ?135, 1998.
37. P. Fieguth, D. Terzopoulos, "Color-based Tracking Of Heads And Other Mobile Objects at Video Frame Rates," Proceedings of CVPR, 1997.
38. M. Isard, A. Blake, "CONDENSATION-conditional density propagation for visual trackin g," International Journal of Computer Vision, 1998.
39. M. Isard, A. Blake, "Contour tracking by stochastic propagation of conditional density," Proceedings of ECCV, pp. 343-356, 1996.
40. G. Healey, "Using Color for Geometry-insensitive Segmentation," J. Opt. Soc. Am. A, 6(6), pp. 920-937, 1989.
41. T. Uchiyame and A. M. Arbib, "Color Image Segmentation Using Competitive Learning," IEEE Transactions on Pattern Analysis and Machine Intelligent, Vol. 16, No. 12, pp. 1197-1206, December 1994.
42. S. H. Han, I. S. Kweon, C. Y. Kim, Y. S. Seo, "Bayesian Shot Detection using Structural Weighting," International Workshop on Machine Vision Applications, Japan, November, 2000.
43. S. Haykin, "Neural Networks," Prentice Hall, 1999.