하나의 비디오 입력을 위한 모습 기반법과 모델 사용법을 혼용한 사람 동작 추적법

박 지 헌[†] · 박 상 호[††] · J. K. Aggarwal[†††]

요 약

사람의 동작을 믿을 수 있게 따라가는 것은 감시용 비디오나 사람과 컴퓨터간의 사용자 인터페이스 개발에 있어서 필수적이다. 이 논문은 모습 기반법(appearance-based method)과 모델 사용법을 혼용하여 사람을 추적하는 새로운 방법에 관한 논문이다. 하나의 비디오 입력이 화소 단위 및 물체 단위로 처리된다. 화소 단위의 처리에 있어서 개별 화소색을 분류하는 훈련방법으로, 가우스 혼합 모델(Gaussian mixture model)을 사용하였다. 물체 단위의 처리에 있어서 사람 몸에 대한 삼차원 모델링을 하고, 모델 몸체를 투사면(projection plane)에 투사시켰다. 투사된 몸체와 배경을 제외한 영상과 계산 기하 방법을 사용하여, 화소보다 작은 단위로 겹쳐지는 면적을 계산하였다. 우리의 방법은 정방향 기구학(forward kinematics)을 사용하므로 역방향 기구학(inverse kinematics)을 사용하는 방법과 달리 계산 결함(singularity)을 갖지 않는다. 이 논문에서는 사람의 동작을 추적하기 위한 문제를 비선형 방정식 문제로 바꾸었다. 비선형 방정식의 비용 함수는 전경(foreground)의 영상 실루엣(silhouette)과 투사된 삼차원 모델 몸체의 실루엣의 겹쳐지는 면적이다. 화소 단위의 영상을 화소를 하나의 면적으로 계산함으로써, 겹쳐지는 면적에 대한 실수 단위의 계산은 계산 기하를 사용하였다. 이 논문의 방법은 다양한 사람 동작을 인식하기 위하여 사용되었다. 비디오에 나타나는 사람 동작 추적은 매우 우수하다.

Human Motion Tracking by Combining View-based and Model-based Methods for Monocular Video Sequences

Jihun Park[†] · Sangho Park[††] · J. K. Aggarwal[†††]

ABSTRACT

Reliable tracking of moving humans is essential to motion estimation, video surveillance and human-computer interface. This paper presents a new approach to human motion tracking that combines appearance-based and model-based techniques. Monocular color video is processed at both pixel level and object level. At the pixel level, a Gaussian mixture model is used to train and classify individual pixel colors. At the object level, a 3D human body model projected on a 2D image plane is used to fit the image data. Our method does not use inverse kinematics due to the singularity problem. While many others use stochastic sampling for model-based motion tracking, our method is purely dependent on nonlinear programming. We convert the human motion tracking problem into a nonlinear programming problem. A cost function for parameter optimization is used to estimate the degree of the overlapping between the foreground input image silhouette and a projected 3D model body silhouette. The overlapping is computed using computational geometry by converting a set of pixels from the image domain to a polygon in the real projection plane domain. Our method is used to recognize various human motions. Motion tracking results from video sequences are very encouraging.

키워드 : 사람 동작 추적(Human Motion Tracking), 모델 기반 및 모습 기반(Model-based and Appearance-based), 정방향 기구학(Forward Kinematics), 비선형 방정식 해법(Nonlinear Programming), 계산 기하(Computational geometry)

1. 서론 및 관련 연구

사람의 동작을 믿을 수 있게 따라가는 것은 감시용 비디오나 사람과 컴퓨터간의 사용자 인터페이스 개발에 있어서 필수적이다. 움직이는 사람과 같은 비강체(non-rigid) 대상을 추적하는 것은 컴퓨터 분석에 대해 몇 가지 어려움이 있다. 예를 들어, 사람 몸을 의미 있는 부분들로 나누는 것(segmentation), 가려짐(occlusion)을 처리하는 것, 각 신체 부분들을 비디오 상에서 추적하는 문제 등이 있다. 사람 신체를 추적하기 위해 제안된 방법들은 "모델에 기반한(model-based) 접근법"들과 "모습에 기반한(appearance-based) 접근법"들로 구분할 수 있다(자세한 내용을 위해서는 [1, 6]을 참고). 모델-기반 접근들은 기구학(kinematics)과 운동학(dynamics)으로 정의된 사전(a priori) 모델을 이용한다. 모습-기반 접근들은 사전 모델을 이용할 수 없는 경우, 경험적(heuristic) 가정들을 적용한다. 이 두 가지 접근들은 효율성을 높이기 위해 다양한 수준에서 통합될 수 있다[15]. 본 논문에서는 모습-기반 접근과 모델-기반 접근의 기법들을 통합하여 인

※ 이 논문은 홍익대학교 교수 연구년 기간(2002. 8~2003. 8)중 연구되었음.
† 정 회 원 : 홍익대학교 컴퓨터공학과 교수
†† 비 회 원 : University of Texas at Austin 대학원 전자 및 컴퓨터공학과
††† 비 회 원 : University of Texas at Austin 전자 및 컴퓨터공학과 교수
논문접수 : 2003년 5월 12일, 심사완료 : 2003년 8월 21일

간의 동작을 추적하는 새로운 접근법을 제시한다. 본 논문에서 제안된 시스템은 입력 영상을 화소 수준과 대상(object) 수준 모두에서 처리한다. 화소 수준에서는 가우스 혼합 모델(Gaussian mixture model)을 사용하여 개별 화소를 여러 개의 색상 집합들(color classes)로 분류하고, 명칭 완화(relaxation labeling)를 이용하여 그 색상 집합들을 동질적(coherent) 유사화소덩어리(blob)들로 묶는다. 대상(object) 수준에서는, 3차원 신체 모델을 2차원 영상 평면에 투사한 후 영상 자료와의 정합(fitting)을 계산한다. 화소 수준에서의 모습-기반 처리는 전경 실루엣을 제공하는데, 이는 대상 수준에서의 모델-기반 연산에 소요되는 부담을 효율적으로 감소시킨다.

모든 기구학(kinematics)을 사용하는 동작 추적 방법은 역방향 기구학(inverse kinematics)를 사용하였느냐 아니냐에 따라 분류될 수 있다. 만약 기구학 계산에서 역 기구학이 사용되지 않았으면, 우리는 정방향 기구학(forward kinematics) 방법이라고 부르고, 그렇지 않으면 역방향 기구학 방법이라고 부른다. 이 논문의 방법은 정방향 기구학을 사용한 방법이다. 반면 여기서 논의되는 관련된 다른 논문들[8, 12]은 역방향 기구학을 사용한 방법이다.

초기의 논문인 [12]에서는 미분된 역방향 기구학을 사용하여 영상의 겹치는 부분을 계산한 논문이다. 그러나 이 논문에서 사용된 미분 역방향 기구학은 로보틱스 [5]에서 많이 연구 개발된 것이다. [12]논문에서는 사람 모델을 전경 영상에 맞추기 위하여 이차원의 크기 조절이 가능한 길이 조정형 사람 모델을 사용하였다. 이차원에서 역기구학 문제를 풂으로써, 삼차원에서 풀 때 보다 많이 계산 불능 상태를 제거하였지만, 역기구학을 사용하였으므로 계산 불능 상태를 완전히 제거하지 못했다. Huang등의 저자들[8]은 [12]논문에서 사용된 역기구학 방법을 확장시켜 EM(expectation-maximization) 알고리즘을 사용하여 통계적 기법으로 관절체의 몸 형상을 결정하는 문제를 풀었다. Sidenbladh 등의 저자들[17]은 영상에 나타나는 사람의 동작을 따라가는 문제를 입력 영상이 주어졌을 때, 신체 모델 인자의 사후 확률(posterior probability)을 추정하기 위한 확률추론(probability inference) 문제로 바꾸었다

2. 화소 분류

2.1 색상 표현과 배경 제거

대부분의 카메라는 적, 녹, 청색 신호를 제공한다. 그러나 적, 녹, 청 색채 공간은 색상과 밝기에 대한 인간의 눈의 지각 모델을 위해 효율적이지 않다. 본 논문에서는 적, 녹, 청 색채 공간을 색상, 채도, 밝기(Hue, Saturation, Value) 색채 공간으로 변환하여 밝기(혹은 강도)를 색상으로부터 독립적으로 만든다.

각 낱장(frame)의 영상에 대해 배경 분리(background subtraction)를 함으로써, 전경(foreground) 영상 영역을 추출한다(자세한 과정은 [13]을 참조), 영상 좌표 (x, y)상의 화소 $v(x, y)$의 색상 분포는 하나의 가우스 분포(Gaussian)로 모델화 된다. k_b개($k_b=20$)의 낱장의 학습 영상을 이용하여, 각 화소 좌표 (x, y)에서 각 색채 경로(channel)에 대해 평균 $\mu(x, y)$과 표준 편차 $\sigma(x, y)$를 계산한다. 전경 분리(foreground segmentation)는 각 화소 $v(x, y)$에 대해 단순한 배경 모델(background model)을 이용하여 다음과 같이 계산된다. 즉, 각 영상 화소 좌표 (x, y)에서의 화소의 강도 변화를 가우스 배경 모델로부터의 마할라노비스(Mahalanobis) 거리 δ를 계산하여 평가한다.

$$\sigma = \frac{|v(x, y) - \mu(x, y)|}{\sigma(x, y)} \tag{1}$$

전경 영상 $F(x, y)$는 색상, 채도, 밝기의 세 거리 측정치, $\delta_H(x, y)$, $\delta_S(x, y)$, $\delta_V(x, y)$의 최대 값으로 정의된다.

$$F(x, y) = \max[\delta_H(x, y), \delta_S(x, y), \delta_V(x, y)] \tag{2}$$

F에 대해 역치(threshold)를 적용하여 이진 분리 영상(binary mask image)을 얻고, 작은 영역들로 이루어진 잡음 화소(noisy pixel)들을 제거하기 위한 사후 처리로서 변형연산(morphological operations)들이 가해진다.

2.2 색상 분포를 위한 가우스 혼합 모델(Gaussian mixture model)

색상, 채도, 밝기의 색채 공간에서 좌표 (x, y)상의 화소 값은 벡터 차원 d = 3을 갖는 무작위 변수(random variable) $v = [H, S, V]^t$로 표현된다. [13]의 방법에 따라, 전경 화소 v의 색상 분포 $P(v)$는 사전 확률 $P(\omega_r)$의 가중치를 갖는 C_0개의 혼합 가우스 분포의 합으로 다음과 같이 모델화 된다.

$$P(v) = \sum_{r=1}^{C_0} p(v \mid \omega_r) P(\omega_r) \tag{3}$$

이 때, r-번째 조건 확률은 하나의 가우스 분포로 다음과 같이 가정된다.

$$p(v \mid \omega_r) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left[-\frac{(v-\mu_r)^t \Sigma^{-1} (v-\mu_r)}{2}\right], \quad r = 1, \cdots, C_0 \tag{4}$$

각각의 가우스 분포 요소 (component)는 r-번째 색상 집합(color class) ω_r의 사전 확률 $P(\omega_r)$, 해당 화소의 색상 평균값 벡터 μ_r과 공변량 행렬(covariance matrix) Σ_r을 표현한다.

2.3 가우스 인자(Gaussian Parameters)의 학습(training)

입력 비디오 중 최초 η개의 낱개의 영상(image frame)을 학습 자료로 사용한 EM 알고리즘(EM algorithm)을 통해 가우스 인자들이 얻어진다. 가우스 인자의 초기화는 다음과 같이 실시된다. 모든 사전 확률은 동일한 것으로 가정한다.

평균 값은 화소의 가능한 값 범위 내에서 균일 분포(uniform distribution)로부터 무작위 추출된다. 공변량 행렬(covariance matrix)는 항원 행렬(identity matrix)로 가정된다. 학습 절차는 위에 언급된 인자들을 반복적으로 갱신(iteration)함으로써 달성된다(세부 사항은 [4]를 참조). 평균 값의 변화가 이전 반복 값의 1퍼센트 이내이거나 반복 횟수가 미리 지정한 한도를 초과하면 반복 갱신과정은 중지된다. 본 논문에서는 10개의 가우스 요소를 가지고 시작하며($C_0 = 10$), 유사한 가우스 분포들은 [10]의 방법에 따라 하나로 합쳐져 최종 C개의 가우스 요소만 남게 된다. 확립된 최종 C개의 가우스 요소들은 연속되는 입력 영상들을 C개의 색상 집합 중 하나로 분류하는데 사용된다.

2.4 개별 화소의 분류

개별 화소의 색상 분류는 최대 사후 확률에 의한 분류법(maximum a posteriori classifier)에 의해 이루어진다. 화소 색상에 대한 가우스 혼합 모델 G가 얻어지면, 연속되는 낱장의 입력 영상 내 화소가 각각의 가우스 요소에 속할 사후 확률이 계산된다. 주어진 화소 v에 대해 최대의 확률 값을 갖는 색상 집합이 해당 화소의 색상 명칭으로서 선택된다.

$$\omega_L = argmax_r [\log(P(\omega_r | v))], \quad 1 < r < C \tag{5}$$

2.5 명칭 완화(Relaxation Labeling)

화소 색상 분류는 서로 인접한 상이한 색상 명칭들의 임의적 형태(free-form)의 영역(blob)들을 생성한다. 연결-요소 분석(connected component analysis)을 통해 인접 영역들을 등록한다(register). 명칭 완화 기법(relaxation labeling) [13, 16]을 사용하여, 작은 영역들을 제거하고, 인접 영역들 간의 색상 유사 정도에 근거해서 균일한(coherent) 거대 영역(large blob)들을 다음과 같이 생성한다. 즉, 두 인접 영역 A_i과 A_j는 각 영역의 색상 특징 Φ간의 마할라노비스 거리(Mahalanobis distance) δ_Φ에 의해 다음과 같이 정의되는 색상 유사 정도를 만족할 경우 하나의 영역으로 통합된다.

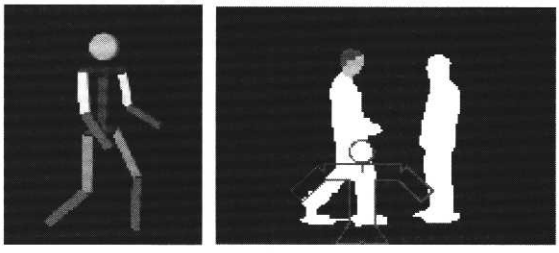
$$\delta_\Phi = (\Phi_i - \Phi_j)^T \Sigma_\Phi^{-1} (\Phi_i - \Phi_j) \tag{6}$$

$$\Phi = [\mu_H, \mu_S, \mu_V]^T \tag{7}$$

이때 Σ_Φ는 영상 내 모든 영역들의 색상에 대한 공변량 행렬(covariance matrix)이다. δ_Φ가 학습 자료에서 얻어진 역치 T_Φ 보다 작을 경우 영역 A_i와 A_j는 하나로 합쳐진다(세부 사항은 [13]을 참조). 명칭 완화과정의 결과, 일련의 영역들이 얻어지며, 이 영역들은 화소의 위치와 색상에 근거하여 전경 영상을 균일한 부분들로 분할한다. 이 과정은 전경의 사람들에 대한 초기 분할과 추적의 모습-기반 접근을 이룬다. 더욱 상세한 분할과 추적은 대상 수준의 처리에서 이루어지며, 대상 수준에서는 삼차원 인간 모델이 이차원 영상 평면에 투영되어 사용된다.

3. 사람 몸체 모델링 및 비용 함수

3.1 사람 몸체 모델링



(a) (b)

(그림 1) (a) 삼차원으로 모델링 된 몸체, (b) 전경 영상과 투사된 모델 몸체와의 겹친 상황

(그림 1)(a)에 보여진 것과 같이, 삼차원 몸체는 9개의 원통과 한 개의 구로 이루어져 있다(엉덩이와 어깨 부근의 두 개의 원통은 이해하기 쉽게 그려진 것이고, 실제 몸체 투사에는 사용되지 않는다). 이 삼차원 원통 및 구가 실수 공간인 2차원 투사면에 투사된다. 구는 사람의 머리를 나타내며, 머리 이외의 몸체 부분은 다양한 길이와 반경으로 이루어진 원통으로 구성된다. 이 논문에서 사용된 모델 몸체는 [17]논문에서 사용된 몸체와 유사하다. 보다 복잡한 형태로 몸체를 모델링 하는 것도 가능하겠으나, 이 논문에서 사용된 겹쳐지는 면적 계산 방법은 겹쳐지는 면적이 최적이 되도록 하는 것이므로, 대략적으로 몸체를 모델링 하여도 전경 영상의 사람 몸체를 잘 따라간다. 그러므로 복잡하게 잘 모델링한 몸체를 사용하여도 효과는 별로 크지않다. 현재 1 자유도 관절 9개를 사용하여 몸체를 모델링 하였으며, 몸체 이동에 관한 자유도 2개를 포함하면 총 자유도는 11이다. 또한 카메라 방향을 보정해 주기 위하여 바닥에 수직인 방향으로 회전이 가능하다. 11개의 자유도는 이 논문에서 사용한 변수 개수와 동일하다. 이 변수들은 비선형 방정식에 사용되는 비용함수의 제어 변수들이기도 하다. 모델 몸체의 각 부위는 계통적(hierarchical)으로 연결되어 있다. 이것은 엉덩이를 기준으로 하는 나무 구조로 되어있다. [8]논문에서 제안된 방법은 몸체 스스로 자신의 일부를 가리는 문제를 해결하지 못한다. 그리하여 몸체의 반을 무시하였다. 그러나 본 논문에서는 계산 기하를 사용하여 몸체의 가려지는 부분을 투사하여 합집합 연산을 하여 전체 면적을 계산하였다. 몸체 각부분을 독립적으로 투사하고, 투사된 몸체 전체에 대한 일종의 그림자(투사체)를 계산한다. 물론 원래의 몸체 배치에서 가려지는 것이 많을수록 정보의 손실은 많이 일어나나, 몸체가 서로 어떻게 가려지든 발생되는 일종의 그림자를 충실히 흉내내기 위하여 모델 몸체의 관절값을 결정하는 것이다. 몸체의 그림자를 계산하는 것과 유사한 이러한 방법으로 계산된 투사된 모델 몸체의 면적과 전경 영상의 면적을 계산하기 위하여 두 면적의 공통되는 영역을 교집합을 계산하여 구한다. 모델 몸의 형상을 좀 달리 하였을 때, 더 많은 공통되는 영역을 만들 수 있으면, 그 형상이 보

다 더 영상 상의 몸체를 잘 추적하는 방법이 될 것이다.

사람 몸체의 크기가 다양할 수 있으므로, 사람의 키와 비만 정도를 계산할 수 있어야 한다. 이러한 문제는 영상 상에서 사람 몸체에 대한 화소-밀도 지도(pixel density map)를 만들고, 화소의 개수를 셈으로써 키와 비만 정도를 예측 가능하다. 이 정보를 기반으로 하여, 사람의 크기를 추정하기 위한 비선형 문제를 풀면된다. 이 논문에서는 사람의 크기는 표준 체형을 따른다고 가정하였다. 초기 사람 몸체크기가 결정되면 몸체의 두께에 관련된 값들은 상수로 고정되고, 사람 형상에 관련된 값들만 변수가 된다. (그림 1)(b)는 최적으로 겹치는 형상을 찾기 위한 초기 상태를 나타낸다. 그림에서 보는바와 같이 초기의 관절 값들은 임의의 값이 주어진다. 사람 몸체 위치 및 관절 각도 등이 변수이므로, 이 값들을 바꿈으로써 최적으로 겹쳐지는 값을 알아낼 수 있다. 이 값이 원하는 최적으로 겹치는 형상이다.

3.2 정방향 기구학을 사용한 비용 함수

몸체의 투사는 형상보존(affine) 변환을 한다고 가정한다. 물체의 거리감이 포함되지 않는 직교 변환 카메라를 사용하는 경우에 해당되는 것이다. 이러한 사람 몸체의 동작을 추적하기 위한 본 논문의 방법은 몇 가지 가정을 한다. 첫째, 카메라의 주시 방향에 수직인 투사면과 병렬로 사람의 몸의 동작이 일어난다는 것이다. 둘째 비디오 입력이 하나이란 것이고, 셋째 입력은 옆면 위즈로만 들어온다는 가정이다. 물론 정확히 옆면일 필요는 없으나, 몸이 많이 돌아갈수록 동작 추적의 질은 떨어진다. 이러한 가정들을 기반으로 하여, 사람 신체 부위에 따른 거리감은 없다고 본 것이다. 이러한 경우는 길다란 복도를 다니는 사람들의 행동을 관찰하는 감시 카메라에 사용되어질 수 있다. 또한 최적의 형상을 구하는 문제를 푸는 것이므로, 사람 신체 부위에 대해 거리감이 고려되지 않더라도 문제가 되지 않는다.

관절 각도 값들과 몸체 변위 값들이 주어졌을 때, 정방향 기구학 함수 $h(\cdot)$는 모델 몸체의 삼차원 점들을 계산한다. 여기서 $\cdot$는 일반적인 변수이다. $\mathbf{P}$ 행렬은 앞에서 계산된 점들을 이차원 투사면에 투사시킨다. 투사된 점들로 이루어지는 다각형과 영상의 전경으로 이루어지는 다각형이 서로 비교된다. $g(\cdot)$ 함수는 영상의 전경 화소들을 입력으로 받아서 다각형(들)으로 바꾸는 함수이다. 입력 영상 $f(\cdot)$ 함수는 영상의 배경을 없앤다. 투사면은 실수의 이차원 공간으로 만들어지므로, 배경이 없어진 전경 영상의 실루엣을 $r(\cdot)$ 함수를 사용하여 다각형으로 바꾼다. 영상의 겹쳐지는 연산을 실수 공간에서 하는 이유는 비용 함수를 미분 가능하게 만들어 비선형 방정식 해법을 적용하여 해를 구하기 위해서이다.

$$c(I,\vec{\theta}) = -w_1 \text{ x } [a(r(f(I)) \cap (\bigcup_i g(\mathrm{P}\cdot \mathrm{h}_i(\vec{\theta},t))))]$$
$$+ w_2 \text{ x } \sum_{jk} (w(x,y) \text{ x } a(d(x,y) \cap (\bigcup_i g(\mathrm{P}\cdot \mathrm{h}_i(\vec{\theta},t)))))$$
$$+ w_3 \text{ x } (h_{hc}(\vec{\theta},t) - q(f(I)))^2 \quad (8)$$

식 (8)에 사용된 기호에 대해 보다 자세히 설명한다.

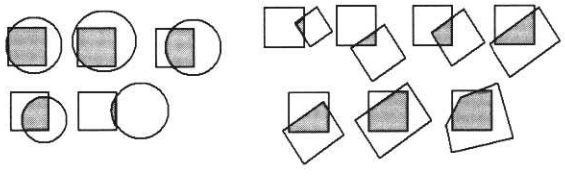
- $\mathbf{P}$: 직교형 투사 행렬(orthographic projection matrix)이며, 삼차원 몸체를 2차원 투사면으로 투사하기 위하여 사용한다.
- $h_i(\cdot)$: 관절 각도를 변수로 하는 i번째 몸체 부위에 대한 비선형 정방향 기구학 함수이다.
- $\vec{\theta}$: 관절 자유도 벡터이며, 여러 장의 영상을 고려할 경우, 시간의 함수가 된다.
- $g(\cdot)$: 이차원 점들을 받아서 다각형으로 변환하는 함수이다.
- $r(\cdot)$: 배경을 지운 전경 영상을 입력으로 받아서, 전경 부분을 다각형으로 바꾸는 함수이다.
- $f(\cdot)$: 원 영상이 입력되어졌을 때, 배경을 지운 전경 영상을 만드는 함수이다.
- I : 어떠한 처리도 되지 않은 원 영상이다.
- $I(x,y)$: 영상 위치 (x, y)에 있어서 회색 영상(grey image)의 화소 값을 나타낸다.
- $d(x,y)$: 이것은 화소거리지도(distance map) 영상 위치 (x, y)에 있는 화소 하나에 해당하는 다각형으로, 면적의 크기는 1이다.
- t : 시간을 나타낸다.
- $w_d(x,y)$: 영상 위치 (x, y)에 있는 배경 영상의 화소거리지도 값을 나타내는 스칼라 값이다.
- $\cap$: 이것은 두 개의 다각형을 입력으로 받아 교집합 연산인 중첩되는 부분을 계산하는 연산자이다.
- $\cup$: 이것은 두 개의 다각형을 입력으로 받아 합집합 영역을 계산하는 연산자이다.
- $a(\cdot)$: 다각형을 입력으로 받아, 면적을 계산하는 함수이다.
- $c(\cdot,\vec{\theta})$: 입력으로 원 영상과 모델 몸체의 관절각도 값들을 받아서 출력으로 비선형 방정식 의 비용함수를 계산한다.
- w_i : w_i, $i=1,2,3$는 가중치 값들이다.
- $q(\cdot)$: 전경 영상을 입력으로 받아 영상의 머리 부분의 중심을 계산한다.
- $h_{hc}(\cdot)$: 정방향 기구학을 사용하여 모델 머리의 중심을 계산하는 함수이다.

이상의 연산에서 입력이 관절 각도 벡터인 $\vec{\theta}$이므로, 이 연산은 순전히 정방향 기구학 계산방법이며, 정방향 기구학 계산에서는 계산 불능 상황이 결코 발생하지 않는다. 정방향 기구학 방정식을 유도하는데 있어서, 각 몸체 부위에 지역 좌표계(local coordinate)를 설정해 두고 계통적 구조의 나무 형태의 몸을 만들었다. 관절은 관절 회전을 제약하기 위한 각도를 설정하였다. 비선형 방정식 해법을 사용하므로 관절 회전 제약을 설정하는 것은 매우 간단하다. 이러한 점은 영상의 화소 처리 기법에 의존한 방법에 비하여 모델을 사용하여 사람의 동작을 추적하는 방법의 장점이기도 하다. 머리

부위에 관한 정보는 영상에 대한 전처리 과정에서 찾아낼 수 있다. 그러나 머리 부위에 대한 정보가 필수적인 것은 아니다. 식 (8)의 $w_3 = 0$으로 두고, 머리 부위에 대한 정보를 사용하지 않고 추적할 경우, 미세한 몸 떨림 현상을 감지할 수 있다. 본 논문에서 제시된 방법은 최적으로 맞는 형상을 찾는 방법이므로, 어느 정도의 영상 전처리에서의 오류가 있어도 형상을 잘 찾아낸다. 그러나 이 방법의 약점중의 하나는 영상 내에 사람이 몇 명이 존재하는지 쉽게 알아내지 못할 수 있다. 특히 영상내의 사람 크기가 매우 작을 때 이러하다.

(그림 1)(b)에 나타내진 것과 마찬가지로, 모델 몸체를 추적하기 위한 초기 형상은 임의의 값이 주어진다. 초기에 모델 몸체가 투사된 다각형과 영상의 전경으로 만들어지는 다각형의 교집합이 없을 경우라도, 본 논문에서는 배경에 대하여 화소거리 지도를 만들어 사용하므로, 쉽게 이동하고, 관절을 회전하여 최적의 형상 및 위치를 찾아낸다. 본 논문에서 사용된 비용 함수는 식 (8)에 나타나져 있으며, 삼차원 모델이 투사된 다각형과 영상의 전경이 최적으로 맞을 경우, 비용 함수 값은 최소가 된다. 만일 두 사람이 서로의 몸에 의해 몸이 가려질 경우, 각 각 개인에 대한 화소거리지도를 만들어 동작을 추적하여야 하고, 서로 껴안는 등 몸통이 완전히 붙어있는 경우를 제외하고는 동작 추적을 잘한다. 몸통이 붙어 있다가 다시 떨어질 경우, 다시 추적을 계속할 수 있다. 몸통이 붙어있는 상황에서는 누가 누구인지 분간이 힘들므로, 몸통이 붙어있는 두(혹은 여러) 사람 전체를 하나의 투사체로 보고 동작 추적을 하게 된다. 이 경우, 계산된 관절 각도는 부정확해 진다.

3.3 비용 함수를 위한 계산 기하



(그림 2) 화소와 모델 투사체가 겹쳐지는 12가지 경우

여러 사람이 겹쳐지거나, 혹은 한 사람이라도, 일부 신체가 다른 부분을 가리는 경우가 항상 발생한다. 이러한 경우를 충분히 고려하여, 본 논문에서는 각각의 신체부위를 투사면에 투사하고, 투사된 신체부위를 집합 연산을 사용하여 더한다. 더해진 신체부위와 영상의 전경과 비교를 하게되며, 잘 맞지 않을 경우, 모델의 관절 각도의 회전 혹은 모델 몸체 기준점의 이동을 통하여 최적으로 맞는 형상을 찾아내게 된다. 영상의 전경은 화소 단위로 되어있으므로, 가장자리가 수직 혹은 수평선으로만 구성되어있다. 또한 기본 단위는 화소의 크기이다. 이러한 전경 영상으로 이루어진 다각형과, 삼차원 모델 몸체를 이차원 투사면에 투사시켜 찾아낸 다각형과 공통되는 교집합을 계산하는 연산을 수행하게되는데, 투사된 다각형은 실수 공간의 값을 가진다. 그리하여 결과적으로 만들어지는 공통된 부분의 다각형 면적은 실수 값이며, 이러한 값은 미분 값을 계산할 수 있으므로, 비선형 방정식의 해법을 사용하여 최적의 면적(전경 영상과 공통되는 부분은 최대이나, 모델 몸체가 배경 영상으로 최소로 나가게 되는 형상)을 가지는 형상을 찾을 수 있다. 이러한 과정에서 몸체 각 부위를 더하여 전체 모델 몸체를 구하는데 필요한 합집합 연산, 또한 모델 몸체와 전경 영상과의 공통된 영역을 계산하기 위한 교집합 연산은 Weiler-Atherton[7]의 다각형에 대한 교집합을 계산하는 알고리즘과 이를 변형한 합집합을 계산하는 알고리즘을 사용하여 계산한다. 여기서 본 논문에서는 공통되는 부분을 계산할 때, 전경 영상 전체에 대하여 계산 기하 연산을 한 것이 아니고, 영상 화소 하나씩 분리하여 연산하였다. 이렇게 분리하여 연산한 이유는, 투사된 모델 몸체가 전경 영상에 머물도록 하기 위하여, 배경으로 나간 부분을 계산하여, 비용 함수에 반영하여 억제 시켜야 한다. 배경 함수는 화소거리지도를 만들어 사용하므로, 각 화소별로 화소거리지도 값이 다르다. 이러한 이유로 계산 기하 연산은 화소별로 처리되었다. 또 다른 특이한 점은 머리 부위는 투사될 경우, 원이 된다는 점이다. 즉 이것은 다각형이 아니므로, 알고리즘에서 원호를 선과 같이 처리해 주어야 한다. 이러한 기하 계산 결과로 발생하는 물체는 다각형이되, 일부는 원호로 구성되어있고, 많은 경우 다각형 내부에 구멍이 존재한다. 계산 결과 발생하는 불규칙 형태의 다각형에 대하여 삼각형 및 원호로 분리하는 작업(triangulation)을 거쳐 이들에 대한 면적을 계산한다. 이러한 결과는 본 기하 계산이 잘 동작하는 것을 보여준다.

(그림 2)에서는 모델의 머리가 투사면에 투사된 형태(원)와 전경 영상 화소간의 교차되는 모든 가능한 경우 5가지와, 모델 몸체가 투사면에 투사된 형태(임의의 다각형)와 전경 영상 화소간의 교차되는 모든 가능한 경우 7가지 등 총 12가지 경우를 나타낸다. 정사각형은 화소를 나타내며, 교차 계산은 한 화소를 실수 공간의 다각형으로 간주하여 교차되는 영역을 계산하는데, 교차 계산에는 Weiler-Atherton의 다각형 교집합 계산 알고리즘이 사용된다. 다음 계산 과정은 이렇게 계산된 모든 교차 영역들을 각 화소별로 합집합을 구하는 것이다. 이 경우 Weiler-Atherton의 다각형 교집합 계산 알고리즘을 수정한 합집합 계산 알고리즘[7]을 사용한다.

4. 기구학과 계산 불능(Kinematics and Singularity)

여기서는 역방향 기구학을 사용할 경우 계산 불능(singularity) 상황이 발생할 수 있음을 설명하고자 한다. 정방향 기구학은 관절각이 결정될 경우, 몸의 각 부위의 위치 혹은 방향을 계산하는 과정이다. 역방향 기구학은 몸의 각 부위의 위치가 결정되었을 경우, 이러한 값을 가지게 하는 관절 값들을 알아내는 것이다. 즉 정방향 기구학 방법은 입력 값이 주어졌을 때, 결과 값이 어떻게 나오는지 단순히 관찰하는 것이고, 역방향 기구학은 이러한 값을 가지게 하는 관절 값

이 무엇인지 알아내기 위하여 연립 방정식을 푸는 것이라고 생각하면 된다. 당연히 방정식을 풀 때 해가 없는 경우가 발생할 수도 있으며, 이러한 경우 계산 불능 상황이다. 정방향 기구학에서는 계산 불능 상황이 발생할 경우가 절대로 없다. 정방향 기구학은 방정식을 푸는 형태가 아니고, 단순히 함수값을 계산하는 형태이기 때문이다. 이것을 수식을 이용하여 자세히 설명하면 다음과 같다. 기구학은 미분을 한번 취하거나, 혹은 두 번 취하여 사용하기도 한다. 영상은 앞의 영상과 비교하여 차이점을 가지고 계산을 하기도 하는데, 이러한 차이가 미분과 연관되기 때문이다. 그러나 미분을 취하여 계산하는 방법이 항상 좋은 것만은 아니다. 영상 처리의 오차로 인해서 앞 뒤 영상간의 차이가 부정확할 수 있기 때문이다. 관련 연구에서 논의된 많은 논문들이 역방향 기구학을 사용하고 있으며, 이들 대부분은 역방향 기구학을 한 번 미분한 속도 레벨에서 수식을 유도하여 사용한다. 이 경우 영상간의 차이가 주어졌을 때 역방향 기구학 식을 사용하여 관절의 속도를 알아낼 수 있다. $\vec{\theta}$은 관절각도 벡터이고, $\overrightarrow{f(\vec{\theta})}$은 정방향 기구학 식이다. 정방향 기구학 식을 미분하면 정방향 미분 기구학식인 $\dot{\vec{f}} = \frac{\partial \vec{f}}{\partial \vec{\theta}} \dot{\vec{\theta}}$ 식을 얻을 수 있다. $\frac{\partial \vec{f}}{\partial \vec{\theta}}$은 자코비언(Jacobian) 행렬이라고 한다. 이 행렬의 의미는 관절의 변화가 몸체 부위(손 끝 등)에 미치는 영향이다. 여기서 영상의 차이값은 $\dot{\vec{f}}$이고, 이 값이 주어졌을 때 $\dot{\vec{\theta}}$ 값을 알고자 하는 것이다. 이 값을 알면 현재의 관절 값 $\vec{\theta}$에 시간에 따른 차이값을 더하면 다음 순간의 $\vec{\theta}$을 알 수 있다. 그러나 $\dot{\vec{\theta}}$을 알기위하여 $\dot{\vec{\theta}} = \left[\frac{\partial \vec{f}}{\partial \vec{\theta}}\right]^{-1} \dot{\vec{f}}$을 계산하여야 하며, $\frac{\partial \vec{f}}{\partial \vec{\theta}}$는 정방형(square) 행렬이 아닐 수 있다. 이 경우 가상의 역행렬(pseudo inverse matrix)을 구해야 하기도 하며, 어떤 경우에는 $\left[\frac{\partial \vec{f}}{\partial \vec{\theta}}\right]^{-1}$을 아예 구하지도 못한다. 이 경우 계산 불능 경우에 빠진다. 이 문제를 해결하기 위하여 singular value decomposition과 같은 방법을 사용하기도 하나, 계산이 부정확하다. 이 것은 엄청나게 많은 해들 중에서 하나의 해를 선택했다는 것을 의미한다. 선택된 해가 최선이라는 보장이 있는 것도 아니다. 정방향 기구학을 사용할 경우, 역 자코비언 행렬을 구할 필요가 없다. 그러므로 계산 불능에 빠질 염려도 없다. 이 논문에서 사용하는 정방향 기구학은 $\vec{\theta}$ 값이 제어 변수이다. 이 값이 결정되면 $\overrightarrow{f(\vec{\theta})}$을 구한다. 계산된 $\overrightarrow{f(\vec{\theta})}$은 전경 영상과 얼마나 맞는지 비용함수에서 계산되어진다. 보다 더 잘 맞는 $\vec{\theta}$ 값이 결정되면, 그 값으로 바꾼다. 계산이 끝나면, 각 영상에 대해 $\vec{\theta}$을 알아낼 수 있다. 여러 장의 영상을 합치면 시간의 개념이 포함되고, $\vec{\theta}$들을 보간하면 시간의 함수인 $\vec{\theta}(t)$을 구할 수 있다. 이 들을 한 번, 두 번 미분하면 관절 각 속도, 각 가속도 함수를 알 수 있다. 보간된 $\vec{\theta}(t)$ 값은 컴퓨터 그래픽스의 동작 캡처 등의 목적으로 사용될 수 있다.

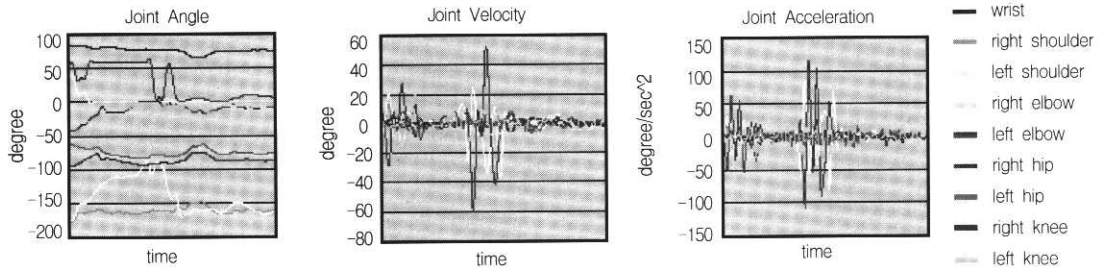
5. 계산 결과 및 결론

본 논문에서 제시한 방법에 대한 타당성을 제시하기 위하여 여러 가지 사람 동작 영상에 대해 동작 추적을 하였다. 걷기, 악수, 차기, 한 사람이 다른 사람 밀기, 손으로 가리키기를 하였으며, 모든 비디오에는 고정된 배경에 두 사람이 서로 동작한다. 밀기, 악수 동작에는 실제 신체 접촉도 있다. (그림 3)에는 왼쪽의 사람이 오른 쪽의 사람을 밀어낸다. 모든 동작은 카메라가 바라보고 있는 방향과 수직인 면에서 대부분 일어난다. 그림에서 빨간 색은 모든 모델 몸체를 기하학적으로 합한 결과를 보여주는 것이다. 다각형과 비슷하게 생긴 몸체의 내부에 구멍이 난 것도 볼 수 있다. 이것은 본 논문의 계산 기하 연산이 잘 동작한다는 것을 보여준다. 변수의 개수를 줄이기 위하여 한 번에 비디오에 나타나는 한 사람씩만 동작 추적하였다. 나타나는 사람들의 몸체끼리 극심한 간섭이 없는 한 동작 추적은 잘되며, 카메라의 각도가 약간 비틀려져 있어도 상관없이 추적할 수 있다. 나타나는 사람들 몸체끼리 간섭이 있으면 동작 추적의 질이 떨어지기도 하는데, 그 예가 (그림 3)의 세 번째 영상에 나타나 있다. 여기서 전경을 처리할 때, 사람 구분을 하지 않았으므로, 왼쪽 사람의 팔이 오른쪽 사람의 얼굴을 가리고 있다. 즉 비용 함수는 모델 몸체가 영상의 전경을 최대한 많이 가리되, 배경을 침입하는 것은 최소화하는 전략이다.

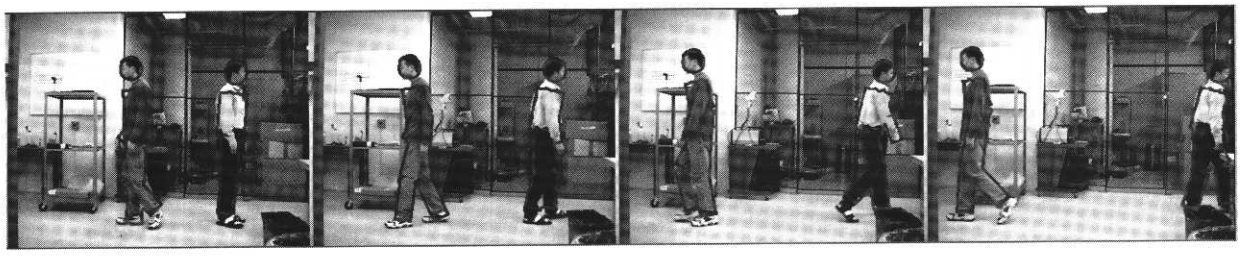
(그림 4)에서는 (그림 3)에 해당하는 비디오 동작에 대한 그래프를 나타낸다. 한 영상 프레임에서 모델의 형상을 가지는 관절 각을 계산한 후, 각 관절각도 값들을 Natural Cubic Spline 함수를 사용하여 C_2 연속으로 보간하였다. 보간한 함수에서 속도 및 가속도 값을 유추할 수 있다. 그래프를 관찰해 보면, 왼쪽 사람의 경우, 모델 몸체의 왼쪽 팔만 영상의 팔 움직임을 따라간다. 이것은 하나의 영상 입력만 사용하여 삼차원 동작을 따라가는 경우에 발생하는 정보 부족으로 인한 것이다. 이러한 점을 보완하기 위해서는 최소한 두 개의 비디오 입력이 필요하다. (그림 5), (그림 6), (그림 7)에서는 걷기, 악수, 손으로 가리키기 동작의 추적을 보여준다.

본 논문에서는 "모습에 기반한(appearance-based) 접근법"과 "모델에 기반한(model-based) 접근법"을 결합한 새로운 사람 동작 추적법을 제시하였다. 이 방법들의 결합은 화소 및 화소를 처리한 물체 단위에서 일어난다. 모습 기반법(appearance-based method)은 화소 수준에서는 전경 영상에서 나타나는 사람을 인식하기 위하여 가우스 혼합 모델(Gaussian mixture model)을 사용하여 개별 화소를 여러개의 색상 집합들(color classes)로 분류하고, 명칭 완화(relaxation labeling)을 이용하여 그 색상 집합들을 동질적(coherent) 유사화소덩어리(blob)들로 묶는다. 배경 제거가 되고, 유사화소뭉치들로 화소들이 분류된 후, 3차원 신체 모델을 2차

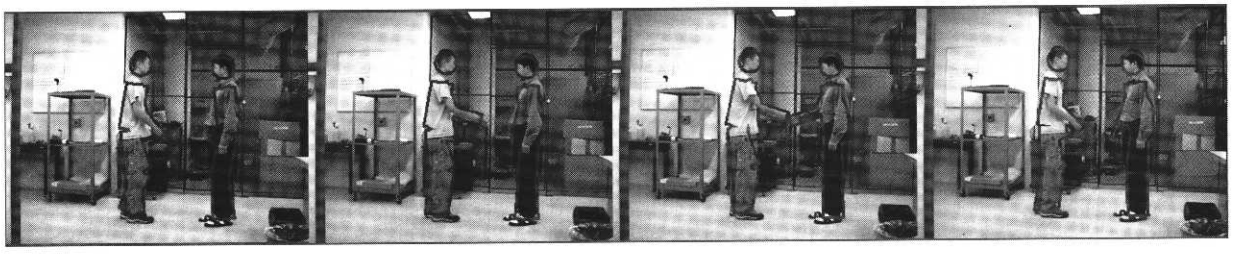
(그림 3) 밀기



(그림 4) (그림 3) 동작에 대한 관절 각도, 속도, 가속도 그래프



(그림 5) 걷기 : 헤어짐



(그림 6) 악수

(그림 7) 가리키기

원 영상 평면에 투사한 후 영상 자료와의 정합(fitting)을 계산한다. 화소 수준에서의 모습-기반 처리는 전경 실루엣을 제공하는데, 이는 대상 수준에서의 모델-기반 연산에 소요되는 부담을 효율적으로 감소시킨다. 다만 이 과정이 필수적인 것은 아니나, 이 과정을 거치면 보다 정교한 동작 추적을 할 수 있었다. 3차원 신체 모델 투사체를 사용하여 최적으로 정합되는 형상을 찾기 위하여 비선형 방정식 기법을 사용하였다. 모델을 사용하여 동작을 추적할 경우의 이

점은 단순한 모습 기반법에 의존할 때와 비교할 때, 동작 추적이 안정적이라는 점과, 동작에 대한 구체적인 수치 값을 알아낼 수 있다는 점이다. 동작의 추적에 있어서 정방향 기구학만을 사용하여, 역방향 기구학을 사용할 때 방정식 해를 구할 때 발생하는 계산 불능 가능성을 완전히 제거하였다. 다만 역방향 기구학에 비해 계산 시간은 좀 더 소요될 수도 있으나, 비선형 방정식은 어떤 관절을 움직여야 최적의 정합을 찾을 수 있는지 비용 함수 미분을 통하여 정확하게 알려준다. 계산 기하를 사용하여 다각형 및 유사 형태의 물체에 대한 기하학 적인 연산을 함으로써, 자신의 몸이나, 다름 물체에 의한 가려지는 현상은 신경 쓸 필요가 없다. 비용 함수에서는 있는 그대로 상황을 투사시켜, 현재 가진 정보에 대해 국부적으로 최적의 정합을 찾기 때문이다.

참 고 문 헌

[1] J. K. Aggarwal and Q. Cai, "Human motion analysis : a review," Computer Vision and Image Understanding, Vol. 73, No.3, pp.295-304, 1999.

[2] H. Asada and J. Slotin, Robot Analysis and Control, John Wiley and Sons, New York, NY, 1985.

[3] J. Craig, Introduction to Robotics Mechanics and Control, Addison-Wesley, Reading, MA, 1986.

[4] R. O. Duda, P. Hart and E. Stork, Pattern Classification, chapter Unsupervised Learning and Clustering, Wiley, New York, 2 edition, pp.517-583, 2001.

[5] R. Freeman and D. Tesar, "Dynamic Modeling of Serial and Parallel Mechanisms/Robotic Systems : Part I-Methodology," in Trends and Developments in Mechanisms, Machines and Robotics, 20th Biennial Mechanisms Conference, 1988.

[6] D. Gavrila, "The visual analysis of human movement : a survey," Computer Vision and Image Understanding, Vol. 73, No.1, pp.82-98, 1999.

[7] F. Hill, Computer Graphics, Macmillan, 1990.

[8] Y. Huang and T. S. Huang, "Model-based human body tracking," in International Conference on Pattern Recognition, 2002.

[9] S. X. Ju, M. J. Black and Y. Yacoob, "Cardboard people : A parameterized model of articulated motion," in International Conference on Automatic Face and Gesture Recognition, Killington, Vermont, pp.38-44, 1996.

[10] S. Khan and M. Shah, "Tracking people in presence of occlusion," in Asian Conference on Computer Vision, Taipei, Taiwan, 2000.

[11] L. Lasdon and A. Waren, GRG2 User's Guide, 1989.

[12] D. Morris and J. Rehg, "Singularity analysis for articulated object tracking," in Computer Vision and Pattern Recognition, 1998.

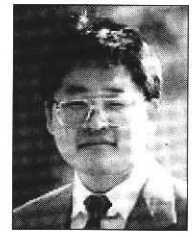
[13] S. Park and J. K. Aggarwal, "Segmentation and tracking of interacting human body parts under occlusion and shadowing," in IEEE Workshop on Motion and Video Computing, Orlando, FL, pp.105-111, 2002.

[14] W. Press, B. Flannery, S. Teukolsky and W. Vetterling, Numerical Recipes, Cambridge University Press, Cambridge, England, 1986.

[15] R. Rosales and S. Sclaro, "Inferring body pose without tracking body parts," in Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, pp.721-727, 2000.

[16] L. Salgado, N. Garcia, J. Menedez and E. Rendon, "Efficient image segmentation for region-based motion estimation and compensation," IEEE Trans. Circuits and Systems for Video Technology, Vol.10, No.7, pp.1029-1039, 2000.

[17] H. Sidenbladh, M. J. Black and David J. Fleet, "Stochastic tracking of 3d human gures using 2d image motion," in ECCV (2), pp.702-718, 2000.



박 지 헌

e-mail : jhpark@hongik.ac.kr
1983년 서울대학교 기계설계학과(학사)
1985년 한국과학기술원 전산학과(석사)
1990년 University of Texas at Austin 전산학과(석사)
1994년 University of Texas at Austin 전산학과(박사)
1983년~1986년 한국전자통신연구원
1986년~1993년 부산외국어대학 컴퓨터공학과 조교수
1994년~현재 홍익대학교 컴퓨터공학과 부교수
관심분야 : 컴퓨터 그래픽스, 컴퓨터 비젼

박 상 호

e-mail : sh.park@mail.utexas.edu
1985년 연세대 공대 전자공학과(학사)
1995년 연세대 문과대 심리학과(석사)
1998년 University of Texas at Austin 심리학과(석사)
2003년~현재 University of Texas at Austin 전자 및 컴퓨터 공학과(박사과정 재학)
1987년~1988년 한국통신
관심분야 : 컴퓨터 비젼, 패턴 인식, 동물·인간 시각

J. K. Aggarwal

e-mail : aggarwaljk@mail.utexas.edu
1964년 University of Illinois at Urbana-Champaign 전산학과(박사)
1987년~1989년 Chairman of the IEEE Computer Society Technical Committee on Pattern Analysis and Machine Intelligence
1992년~1994년 President of the International Association for Pattern Recognition
1964년~현재 University of Texas at Austin 전자 및 컴퓨터 공학과 교수
1976년~현재 A Fellow of IEEE
1998년~현재 A Fellow of IAPR
관심분야 : 컴퓨터 비젼, 패턴 인식