

비디오 감시 및 분석알고리즘: 현황과 기술

한준희 (포항공과대학교)

I. 서론

여러 가지 국내의 범죄 및 국제적인 테러사건 등으로 범죄방지 및 범죄자의 적발, 테러방지, 안전 및 보안, 증거용 등으로 CCTV가 매우 중요한 역할을 담당하게 된다는 것을 많은 사람들이 인식하게 되었다. 그러나 많은 감시용 카메라가 설치되어도 일일이 사람이 감시 할 수는 없다. 연구 결과에 의하면, 한사람이 두 대의 화면을 감시할 때 22분정도 지나면 96%의 주의력이 떨어지는 것으로 알려져 있다. 현재 대부분의 경우에 사건이 발생한 후 녹화된 내용을 검색하여 활용하는 방법으로 이용되고 있다. 감시 카메라시스템의 설치대수는 빠른 속도로 증가하고 있어 사람이 모든 모니터를 보고 감시할 수는 없다. 따라서 사건이 발생한 즉시 이를 보안요원에게 알려주는 지능형 비디오 감시시스템이 필요하게 되었다. 여기에서 지능형이라 함은 영상 내에서 변화의 유무나 센서의 문제점 등을 검출하는 단순한 내용이 아니라, 영상분석, 컴퓨터 비전, 패턴인식 등의 기술을 적용하여 영상에 대한 상위수준의 정보를 추출할 수 있는 시스템을 의미한다. 산업체에서는 일반적으로 이러한 비디오 분석기술을

“Video Analytics” 라고 하고, 내용은 컴퓨터 비전 기술이 사용된다. 비디오 분석알고리즘을 사용하여 사건을 탐색할 수 있으면, 감시자가 모니터를 보고 있을 필요가 없으며, 어떠한 사건이 발생하였을 때 즉시 자동으로 관련 시스템에 통보하여줄 수 있다. 또한 비디오로부터 필요한 정보를 추출하여 텍스트 형태로 바꾸어 전송하거나 저장하여 상위레벨의 정보 추출에 응용할 수 있으며, 자동으로 통계적 자료를 만들 수 있다.

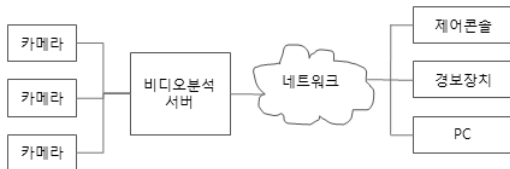
본 고에서는 비디오 감시와 관련된 산업계의 방향, 세부적인 기술적 연구문제, 그리고 비디오 분석알고리즘의 평가에 대한 내용을 다룬다.

II. 비디오감시시스템 구성

비디오 감시 시스템의 구성은 카메라의 종류, 비디오 분석의 기능 및 서비스 형태에 따라 다를 수 있다. 가장 간단한 경우에 카메라와 DVR/NVR, 모니터들로 구성하여 녹화하고, 모니터를 통하여 감시하는 것이다. 비디오 분석기능이 있는 서버에 카메라가 연결되어 있는 경우에는 비디오의 내용을 분석하여 특정한 사건이 탐색되



〈그림 1〉 간단한 비디오 감시시스템



〈그림 2〉 비디오분석 서버가 있는 경우 감시시스템의 예

면 휴대전화 등의 경보시스템에 경보를 줄 수 있다. 제어 콘솔을 통하여 영상을 보거나 카메라를 제어할 수 있다. <그림 1>과 <그림 2>는 이러한 시스템을 간략하게 보여준다.

비디오 감시 시스템의 구성은 판매회사의 제품의 특성에 따라 다르다. 몇 가지 예를 들어보면 IBM의 도시지역 감시시스템^[1]은 행동분석 SSE (Smart Surveillance Engine), 자동차번호인식 SSE, 얼굴인식 SSE들이 비디오를 분석하여 대상물체의 색깔, 크기, 위치 등의 특성정보를 계산

하여 데이터베이스에 저장하고, 사건질의 서비스를 통하여 사건의 탐색, 정보검색, 경보시스템 응용에 적용하며, 비디오 데이터도 검색할 수 있게 하는 구조로 구성되었다. Cisco사의 시스템에서는 이동 순찰차량에 의한 감시시스템의 구조로 무선을 통한 시스템을 구성할 수 있으며, SightLogix사는 광범위한 야외에 신속하게 설치할 수 있는 제품을 보이고 있다.

III. 산업체 동향

1. 지능형 비디오 감시 관련 산업

우리나라를 포함하여 세계적으로 매우 많은 회사들에서 비디오 감시 관련 제품을 생산, 판매하고 있다. 국내의 관련보고서^[2]에 국내관련 업체가 조사되어 있으나, 본고에서는 IPVideo-Market.info에서 비디오 분석 또는 지능형 비디오 감시와 관련된 제품을 다루는 주요 회사라고 명시하는 몇 개 회사의 제품 및 특성을 <표 1>에 요약하였다. 이 표는 각 회사의 자료를 근거로 작

〈표 1〉 Video Analytics(V.A)와 관련된 주요회사 요약

업체명	국 가	주요 특성	비 고 (각 회사의 자료에 의함)
3VR	미국	얼굴인식, 번호인식, 사람계수	한국경찰청 얼굴인식 실험
AgentVi	이스라엘	비디오분석 S/W	i-Lids 인증
Aimetis	캐나다	IP 비디오감시, V.A S/W	야외감시, 일반 매장적용
BRS Labs	미국	비디오 분석 S/W	적응형 학습알고리즘
Cernium	미국	서버에 의한 V.A	공항에 관련된 V.A
Geutebrueck	독일	DVR, 카메라, V.A 등	다양한 일반 감시적용
ioimage	이스라엘	분석기능을 카메라에 내장	고성능, 고가 지능형 카메라, 한국 자동차 부품회사에 설치
NICE systems	이스라엘	네트워크를 통한 분석알고리즘 등	회사, 공항, 대중교통시설, 공공장소 등의 응용
ObjectVideo	미국	임베디드시스템	알고리즘을 카메라나 인코더에 설치하여 적용
SightLogix	미국	무선시스템, 신속하게 야외설치가능	야외적용, GPS 연동, 거대시설감시
via:sys	독일	비디오 경보시스템	저가 S/W, 가정용 응용

성한 것이며, 주요 특성은 그 회사의 다양한 기능이나 제품 중 일부를 나타낸다. 비디오 분석알고리즘은 다양한 사건(침입탐지, 사람 수 계수, 불법주차, 추적 등등)의 탐색을 하는 알고리즘으로 되어 있다. 그러나 이러한 분석기능들 중 일부는 현재까지는 상용제품으로서 만족할 만큼의 성능을 내지 못하는 것으로 알려져 있다.

대부분의 보안이나 범죄예방을 위한 CCTV 운영은 녹화하여 사후에 검색하는 방식이고, 실시간 감시가 필요한 교통시설, 카지노, 공장 등에서는 모니터링 및 녹화 방식으로 사용되고 있다.

2. 시스템의 발전 방향

초기의 CCTV를 이용한 비디오 감시시스템은 사람의 시각범위를 연장시키는 역할과 아날로그 방식을 이용하여 VCR에 녹화하여 사후에 검증하는 방식으로 운영되었으나, 최근에는 H/W, 컴퓨터 망, 알고리즘, 센서 등의 발달로 디지털 감시카메라, IP 카메라, 고해상도 및 고감도, 새로운 영상압축방식 및 임베디드 알고리즘의 적용 등 다양한 진화를 이루고 있다. 전 세계적으로 보아 현재 설치된 감시용 카메라는 아날로그 카메라가 주종을 이루고 있으나, 앞으로 5~7년 이내에는 디지털 카메라가 주종을 이루고, 아날로그 카메라는 자취를 감추게 될 것으로 전망된다. 카메라는 아날로그 → 디지털/IP camera로 변화하고, 100만 화소 이상의 고해상도 카메라, WDR (Wide Dynamic Range) 카메라의 사용이 증가될 것으로 보이며, H.264 코딩기술이 사용되는 추세이다. 단순하게 기록하는 기능에서 비디오 분석알고리즘이 임베디드 형태로 카메라에 내장되거나(Smart camera) 비디오 서버에서 운영 되도록 하고 다양한 기능이 추가되고 있다. 또한

영상코딩	MJPEG,MPEG2 -> H.264
카메라	VGA,D1->Mega Pixel, WDR, Smart Camera
연결방식	Closed Circuit-> Network, Wireless
운용방식	Point Operation-> Networked operation
알고리즘	Simple motion -> High level Event Detection (Analytics)

<그림 3> 비디오 감시 시스템의 변화추이

몇 대의 카메라가 연계되어 감시하는 것을 넘어 광범위한 지역에 많은 수의 감시카메라를 연결하여 상위레벨의 감시정보를 추출하는 시스템이 모색된다. 종합하면 장래에는 유/무선 네트워크를 통한 고해상도, 고감도 IP 스마트 카메라를 이용한 광범위한 감시시스템의 구축이 가능해질 것이다. <그림 3>은 관련 산업의 이러한 변화를 요약하여 보여준다.

3. 비디오 감시 마켓

ABI리서치[2009.4.1]에 의하면, 2008년도의 세계적인 불황으로 비디오 감시 시장의 확장속도가 둔화 되었음에도 불구하고 2008년도와 2009년도 사이에 10%의 성장을 보였으며 이 성장속도는 증가 될 것으로 전망되고 있다. 그러나 IMS리서치는 2009년 9월 보고에서 CCTV와 비디오 감시 장비 부분에서 세계시장은 2009년 3%의 성장을 보일 것으로 전망하였다. 그러나 일본을 제외한 아시아에서의 수익의 증가는 13%를 넘을 것으로 전망하고 있다. 동기관의 2009년 3월 보고서는 특히 중국의 비디오 감시 장비 시장의 성장이 두드러질 것으로 보고 있다. 2007에서 2012까지 시장은 연평균 23%성장하여 30억불 이상이 될 것으로 전망하

고 있다.

IP카메라, 비디오 서버, NVR, Network Video Surveillance S/W는 아직 비디오 감시관련 총합에서 수익이 10%정도 밖에 미치지 못하나 2007년에서 2012년까지는 이 부분의 연평균성장률이 55.6%로 될 것으로 전망하고 있다.

IPVideoMarket.info에 의하면 ABI사에서는 비디오 감시분야의 시장은 2006년 135억불에서 2014년까지 410억불까지 성장 할 것으로 전망하고 있다고 보도하고 있다. 많은 분석가들은 IP/Network 카메라의 성장은 추후 5년 동안 평균 30~40%의 성장을 보일 것으로 전망하고 있다. IMS리서치는 메가픽셀(Mega pixel) 카메라의 성장은 2009~ 12까지 연간 100%의 성장을 할 것으로 보고 있다. 그러나 아직 전체 수로 보아서 메가픽셀 카메라의 수는 적은 편이며, 2012년까지 메가픽셀 카메라가 판매의 6%에 이를 것으로 보고 있으며, IP 카메라는 세계적인 경기 침체와 중국에서의 설치 비율이 낮아 당초 예상보다 보급률이 낮은 것으로 보고 있다.

비디오 분석알고리즘(Video Analytics) 관련 분야는, 순전한 소프트웨어만으로 판매되는 것을 계산하기 어렵기 때문에, 정확한 시장전망과 규모를 추측하기 어려운 것으로 보인다. 이러한 분석 S/W는 카메라나 다른 H/W시스템에 넣어 판매할 수 있기 때문이다. 상위레벨 비디오 분석은 ioimage가 주도를 하고 있다고 본다. 전반적으로 보아 아직 Video 분석 알고리즘 분야의 상용화 적용은 제한이 있는 것으로 판단된다.

4. 비디오 감시 적용 분야

비디오 감시는 보안 및 범죄, 도난 등을 방지

탐색할 수 있는 다양한곳에 적용되고 있다. 특히, 상용시스템이 실제적으로 적용되는 곳은 정부 및 지자체, 은행, 카지노, 매장, 공항 및 교통시설, 주거지역, 교육시설 등이다.

IV. 비디오분석 관련문제의 학술적연구

비디오 분석알고리즘에 대해서는 아직 학술적으로 해결해야할 문제가 많다. 이 단원에서는 주요 세부문제의 접근방법을 요약한다.

1. 물체 탐색/배경 모델링(object detection, background modeling)

비디오 감시에서 가장 먼저 수행되어야 하는 것은 감시대상 영상에서 어떠한 변화가 일어났는지 알아내는 것이다. 이것은 일반적으로 전경/배경(Foreground/Background)분리 문제로 볼 수 있으며, 간단한 문제로 생각 될 수 있으나 비디오 감시에서 가장 중요하고 또한 실제적으로 쉽지 않은 문제이다. 물체를 찾을 때 물체를 인식하지 않고 화소레벨에서 결정하여야 하기 때문이다. 일반적으로 하나의 화소가 배경에 속하는지 또는 움직이는 물체에 해당하는지를 알아내기 위하여 조건부 확률을 이용할 수 있다. 즉 카메라로부터 관측되는 화소의 값을 y 라 하고 X 를 배경이나 전경의 값을 가질 수 있는 확률변수라 할 때, $P(X=x|y)$ 는 관측된 화소 값이 알려져 있을 때 의 전경 또는 배경이 될 조건부 확률을 나타낸다. 배경지역에 대한 화소 값의 확률분포를 구하고 이를 이용하여 한 화소에 대한 배경, 또는 전경 여부를 결정하는 방법이 적용된다. 이

러한 방법으로 많이 쓰이는 방법은 배경을 다수의 가우시안 함수로 모델링하는 방법(Mixture of Gaussian(MoG)이다^[3].

이 방법에서는 일정시간 동안의 화소의 값, $\{x_1, x_2, \dots, x_t\}$ 의 분포를 다음의 함수로 가정한다.

$$P(x) = \sum_{i=1}^K \omega_i * G(x, \mu_i, \Sigma_i), \text{ 여기서 } G(\cdot) \text{는}$$

다변수 가우시안 분포(multi-variate Gaussian distribution)를 나타내며 ω_i 는 $\sum_{i=1}^K \omega_i = 1$ 을 만족 시키는 i 번째 가우시안 함수의 가중치를 나타낸다. 이러한 분포를 계산하기 위해서 정해진 갯수 K 에 대하여 기대치 최대화(Expectation maximization[EM]) 방법을 적용하여 가중치 및 각각의 가우시안 파라미터를 계산한다. 이 계산은 배경의 모든 화소에 대하여 각각 계산된다. 계산 시간을 줄이기 위하여 공분산 행렬을 간단한 모델로 사용하고 가우시안 함수의 수도 제한한다.

서서히 변화되는 배경에 대한 적응을 시키기 위하여 배경의 분포는 시간에 따라 변화될 수 있도록 만들어 진다. MoG방법은 확률을 몇 개의 파라미터 값의 조합으로 나타내기 때문에 효율적이라는 장점이 있는 반면, 일반적인 형태의 분포를 나타내는 데에는 제약이 있다. 좀 더 일반적인 확률 분포를 나타내기 위하여 Parzen 윈도우 확률 분포추정 방식과 같은 개념으로 Kernel을 이용하는 방법도 제시되었다.

배경모델링(배경에 대한 화소의 확률분포)을 시간에 따라 적응되도록 할 때의 문제는 움직이던 물체가 정지하면 배경으로 흡수될 수 있다는 것이다. 이러한 문제가 발생하지 않도록 하는 방법으로 다양한 동적 층(Dynamic layer)의 구성

으로 움직이는 물체를 탐색하고 추적하는 방법이 제시되었다^[4]. 이 방법에서는 시간 t 에서의 동적층 A_t 를 이전단계의 모양정보 Φ_t (shape prior), 움직임 모델 Θ_t , 그리고 층의 관측모양 A_t 의 세 가지 합 $A_t = (\Phi_t, \Theta_t, A_t)$ 으로 나타내었다. 비디오로부터 물체를 추출하고 추적하는 것은 영상 I_0, \dots, I_t 이 주어졌을 때 마코프(Markov) 성질과 베이스 룰을 적용하여 최대 사후 확률(Maximum a posteriori probability [MAP])을 계산하는 방법으로 문제를 정의하였다. 따라서 물체를 찾고 추적하는 문제는 다음 수식을 시간에 따라 계산하는 문제로 만들었다.

$$\begin{aligned} & \max_{A_t} \arg P(A_t | I_t, \dots, I_0, A_{t-1}, \dots, A_0) \\ & = \max_{A_t} \arg P(I_t | A_t, I_{t-1}, A_{t-1}) P(A_t | I_{t-1}, A_{t-1}) \end{aligned}$$

이 문제를 풀기 위한 방법으로 EM알고리즘을 사용한다.

2. 물체 추적 (Tracking)

비디오 감시에서 대상물체를 탐색한 다음 단계는 대상물체가 어떻게 움직였는지를 추적하는 것이다. 물체 추적은 각 영상프레임에서 관측된 정보를 이용하여 각 프레임 마다 물체의 상태(위치, 속도 등)를 추정하는 문제이다. 일반적으로 감시대상 영역에 여러 개의 물체가 나타나거나, 각각의 물체의 특징이 뚜렷하게 구별되지 않기 때문에 물체 추적 알고리즘의 중요성이 대두된다.

많은 경우, 추적하고자 하는 물체의 상태를 확률변수라고 놓고, 확률론적 추적 방법을 적용한다. 이러한 방법의 대표적인 것이 칼만필터링 방법이다. 칼만필터링 방법에서는 상태변수(위치, 속도 등)의 분포가 정규분포라고 가정하는 제한

점이 있다. 이러한 제약을 완화한 방법의 하나로 입자필터(Particle Filtering)방법이 제안 되었다^[5]. 여기에서 추적대상의 위치, 속도는 상태 벡터 X 로 표시되고, 시간 t 에 비디오로부터 관측된 정보(히스토그램 등 영상으로부터 추출한 정보)를 Z_t 라 할 때 조건부 확률분포 $P(X_t|Z_1, \dots, Z_t)$ 을 계산하는 문제로 되며, 이 수식을 관측데이터의 독립성과 베이스 룰을 적용하여 이전 단계의 정보로부터 추정하는 방법을 사용한다. 여기에서는 확률분포를 가우시안 함수처럼 매개변수를 이용하는 방법이 아니라 입자의 가중치와 위치로부터 추정한다. 이러한 방법에서는 초기의 상태를 어떻게 자동으로 설정할 것인가가 하나의 문제가 될 수 있다.

물체의 상태 값을 계산하는 방법으로 상태를 확률변수라고 하였을 때 그 변수의 분포를 추정하는 방법으로 커널을 이용하는 형태로 모델링하고, 이 함수의 경사도(gradient) 값을 수식적으로 계산하여 상태분포의 평균값이 각 프레임마다 변화하는 것을 추정하는 방법(mean shift)도 이용되고 있다^[6]. 이 방법을 적용하는 추적에서는 물체의 움직임이 커널의 크기를 벗어나면 추적할 수 없는 문제점이 있다.

사람의 모양의 특성을 이용하여 추적하는 방법들도 제시되었다. 사람에 대한 *보편적인* 모양을 추출하여 추적하는 방법도 연구되었고, 팔, 다리 몸체 등으로 구성되는 사람의 *모델*을 영상으로부터 구성하고 이러한 것을 이용하여 추출 및 추적에 이용하는 방법도 연구되었다^[7]. 이러한 방법에서는 먼저 모델에 따른 물체를 추출하고, 그 다음, 모델의 모양에 따라 추적하는 문제를 다룬다.

광범위한 지역을 카메라로 감시하거나, 문제의 특성상 여러 대의 카메라가 다수의 물체를 추

적하는 경우에는 새로운 문제가 대두된다. 각각의 카메라가 관측하는 영역이 겹칠 경우도 있고, 그렇지 않을 경우도 있다. 특정 카메라에서 특정 물체가 보이지 않는 경우도 발생한다. 감시 대상 지역이 겹치는 경우 카메라 간의 정보 교환문제(Camera hand off)를 적용할 수 있으나, 서로 감시영역이 겹치지 않는 경우에 연속적인 추적을 위해서는 물체의 상태벡터 및 모양의 특징을 이용할 수 있다. [8]은 카메라들이 넓은 지역을 감시하며 감시영역의 중첩이 없을 때 연속적으로 추적하는 문제를 카메라로부터 관측된 정보와 물체의 궤적과 연관을 짓는 방법으로 해결하고자 하였다. 각각의 카메라에서 관측되는 위치를 하나의 노드로 보았을 때 물체가 다른 지역에서 관측 가능하다는 것은 한 노드에서 다른 노드로의 변이가 가능하다는 것과 같다. 하나의 궤적(ω_K) 이 K 개의 독립적인 궤적구간(T_1, \dots, T_K) 으로 구성된다고 보면, 그 궤적의 확률은 $P(\omega_K) = \prod_{k=1}^K P(T_k)$ 로 계산되며, $O = \{o_j; j=1, \dots, N\}$ 를 모든 카메라로부터 얻은 관측정보로 나타내면, 최적의 궤적을 계산하는 것은 다음의 조건부 확률을 계산하는 문제로 된다.

$$P(\omega_K|O) \propto \prod_{k=1}^K P(T_k)P(O|\omega_K). \text{ 이 문제를 풀기}$$

위하여 MCMC(Markov Chain Monte Carlo) 방법이 적용되었다.

카메라간의 겹침이 있는 경우 [9]는 각각의 카메라에서 관측된 궤적을 호모그래피를 이용하여 공동의 가상 평면지역에 투사하고, 궤적을 정규화 된 특성벡터로 만들고, 이 특성벡터를 이용하여 궤적사이의 정합을 실시한다. 궤적사이의 정합이 이루어지면 결국 카메라간에 하나의 통일된 궤적으로 결합될 수 있어 연속적인 추적 결

과가 계산된다.

물체 추적에서 최근의 경향은 다수의 카메라를 사용한 광범위한 지역에서의 추적문제와 더불어, 감시지역에서 많은 수의 대상물(사람)이 움직이는 경우 가려짐 현상이 발생하고 개개의 물체를 추적하기 어렵기 때문에, 다수의 카메라를 동시에 사용하는 방법들도 모색된다. 이러한 방법은 복잡한 지역에서 많은 사람을 개별적으로 추적하는 문제 (예를 들어 축구경기에서 개개의 선수를 추적하는)의 해결 방법으로도 연구되고 있다.

3. 활동/행동/사건 추출

감시를 위한 비디오 분석알고리즘에서 정해진 사건 또는 사람의 행동을 탐색하여 발생 즉시 그 정보를 제공하도록 하여야 한다. 이러한 기능을 수행하도록 하기 위해서는 물체 추출, 추적 정보를 이용하고, 탐지할 사건을 추출할 수 있는 모듈을 만들어야 한다. 금지구역 침입, 역방향 주행, 주차위반 등의 사건은 실제 응용문제에서 많이 다루어진다.

이러한 문제보다 상위수준의 사건은 다양한 추론 모델을 이용하여 추출된다. 사건은 영상의 특정지역을 침입하여 발생하는 문제, 추적대상이 비정상적인 궤적을 남기는 경우의 문제, 특정한 행동의 시퀀스로부터 유추할 수 있는 상위레벨의 사건 등 다양하다.

추론을 이용하지 않는 방법으로 감시대상 영역에서 전경/배경 분리를 한 후, 전경에 해당하는 지역에 대한 시공간영역에서 화소사이의 동시발생행렬(co-occurrence matrix)을 학습하고, 이 모양으로부터 비정상 사건을 검출하는 방법도 발표되었다(ICCV2009, P. Jodoin등). 이

방법은 대상지역의 물체가 하나의 그룹으로 형성되었을 때 적용할 수 있는 방법이며, 물체의 인식이나 추적과 같은 알고리즘이 사용되지 않아도 되는 장점이 있으나, 특정한 사건을 탐색하는 것이 아니라 비정상상태가 일어났는지의 여부만 판단할 수 있다는 제약이 있다.

사람이 구부리는 동작, 손으로 가리키는 동작과 같은 특정한 기본적인 행동(action)을 탐색하는 방법들도 연구되고 있다. [10]은 영상으로부터 MHI(Motion History Image Feature)를 계산하고, 영상의 모양에 대한 정보로 전경영상과 HOG(Histogram of Oriented Gradients feature)를 종합하여 특정한 동작에 대한 특성으로 사용하였다. 학습영상은 수동으로 취득하였고 분류를 위해서는 모사담금질기법(simulated annealing)을 적용한 SVM(support vector machine)을 사용하였다.

위의 예에서보다는 복잡한 행동을 인식하는 방법으로 M. Ryou(ICCV2009)는 시공간상의 특성을 계산하여 특정한 행동을 포함하는 비디오와의 비교를 통하여 사건을 인식하는 방법을 제시하였다.

가장 이상적인 사건의 탐색은 비디오로부터 유추된 추상적인 사건의 탐색이다. 예를 들어 두 사람이 만나 물건을 전달하는 사건, 한사람이 물건을 들고 와서 놓고 가는 사건, 서로 싸우는 사건, 물건을 훔쳐서 나가는 사건 등이다. 이러한 사건을 인식하기 위해서는 하위레벨에서 대상물체의 추출 및 부분별 추적, 부분별 물체의 공간적 상호관계, 물체의 시간적 상호관계 등을 종합적으로 판단하여 사건을 추출할 수 있는 모델로 설정하여야 한다. 이러한 경우 사용될 수 있는 가장 보편적인 방법은 HMM(Hidden Markov Model)이다. 이 방법에서는 사건을 이루기 위한

세부조건 및 상태(state)가 있고 이러한 조건들이 만족되면 사건에 대한 상태가 새로운 상태로 전이하여 가도록 하고, 특정한 상위레벨의 사건으로 전이되면 그 사건이 발생하는 것으로 볼 수 있다. 각 상태의 전이는 HMM 모델을 구성할 때 학습하여 놓은 상태전이 확률과, 비디오로부터 들어오는 관측정보로 결정할 수 있다. 이러한 모델을 적용할 경우의 장점은 복잡한 상위레벨의 사건을 찾을 수 있다는 것이며, 제약점은 비디오로부터 기본 동작을 추출하는데 대한 어려움과, 각각의 HMM 모델을 설정하고 학습하는데 많은 노력이 든다는 것이다.

4. 물체 계수(Counting)

대상물체의 수를 세는 것은 사람들의 수계산, 군중의 밀도계산, 상품 전시에서의 고객들의 반응정도, 안전을 위한 사람들의 안내 및 통제 등에 적용될 수 있다. 실제 상용 비디오 분석알고리즘에서도 계수기술을 중요한 항목의 하나로 두고 있다. 제한된 공간에서의 사람 수 계수는 출입구를 중심으로 들어온 사람 숫자, 나간사람 숫자를 세는 방법으로 적용 될 수 있다. 출입구가 제한되지 않은 열린 공간에서의 사람 수 계수 문제는 가려짐, 중복 등으로 문제가 어려워진다. 다수의 카메라를 이용하는 방법, 사람의 모양을 특정한 형태로 가정하고 이러한 모양이 대상지역에 투사 되었을 때의 정보를 이용하여 찾는 방법, 사람의 모양(머리와 어깨모양 등)을 찾아 분리하는 방법, 사람의 정확한 계수보다 분포를 계산하는 방법 등이 모색 되었다.

W. Ge 등(CVPR2009)은 영상으로부터 사람의 수를 세기위하여 Marked Point Process [MPP]를 정의하였다. 이 모델에서는 물체(s_i)

를 위치(p_i)와 모양정보(Mark) (m_i) 로 나누고, 모양은 방향, 크기 등을 정의하는 외부 파라미터와, 특정한 모양을 나타내는 고유파라미터로 나누었다. 방향(서 있는 사람의 영상에서의 축 방향)은 학습영상으로부터 소실점을 계산하여 추정하고, 크기는 지표면이 평면이라는 가정과 사람의 키와 원근에 의한 변화를 감안하여 학습한다. 물체 모양에 대한 정보는 학습영상으로부터 계산된다. 영상에 존재하는 사람을 세는 것은 학습된 정보를 이용하여 현재 영상(동적 물체에 해당하는 지역)을 가장 잘 생성할 수 있는 물체의 위치와 모양을 계산해 내는 방법을 사용하고, 이 답을 얻기 위하여 MCMC 방법을 사용한다. 본 방법에서는 비디오의 동작정보를 활용하지 못하며, 탐색되는 사람의 모양이 서있는 경우로 제한되는 단점이 있다.

사람들이 개별적으로 분리되기 어려운 군중 영상에서는 위의 방법을 사용하기 어렵다. 또한 개별적인 사람의 모양에 대한 특징을 계산하기도 어려운 경우가 많다. 이러한 경우에는 전경/배경 분리를 마친 상태에서, 먼저 Homography를 적용하여 원근효과를 보정한 다음 전경에 해당하는 지역에 대하여, 에지나 방향 경사 값(orientation gradient) 등의 특성을 계산하여, 실제 사람의 수와의 관계함수를 도출하고, 그 함수를 실제 군중 수를 추정할 때 사용하는 방법도 적용된다. 이러한 방법에서는 사람의 수를 구체적으로 세는 것이 아니라 영상의 특성을 이용하여 추정치만 계산할 수 있다.

5. 그림자 (Shadow)

비디오로부터 전경과 배경을 분리하는 문제에서 그림자가 문제를 어렵게 만든다. 특히 움직이

는 물체와 함께 변화되는 그림자는 물체탐색의 정확성을 떨어뜨리고, 물체의 분리를 어렵게 하여 잘못된 결과를 줄 수 있다. [11]은 동적으로 변화되는 그림자를 본그림자(umbra)와 반그림자(penumbra)의 두 가지로 나누고, 비디오 감시환경에서 탐색하기 어려운 본그림자의 추출 및 제거방법을 제시하였다. 이들은 전경/배경 분리알고리즘에서 전경으로 분리된 지역을 그림자 지역과 실제 전경에 해당하는 지역으로 다시 분리하는 과정에서 추출된 지역의 텍스처, 경사도, 색도 등을 사용하였고, 그림자 지역과 움직이는 물체와의 인접성을 이용하였다.

[12]에는 움직이는 그림자 탐색 및 제거와 관련된 다양한 방법에 대한 조사가 되어 있다.

V. 알고리즘 평가 및 벤치마킹

본 단원에서는 알고리즘에 대한 평가 기준 및 벤치마킹을 위한 대표적인 활동에 대하여 살펴본다.

1. 평가기준

비디오 감시를 위한 시스템을 만들었을 때 어떠한 조건으로 알고리즘의 성능을 평가 할 것인가 하는 기준이 있어야 한다. 비디오 감시시스템(알고리즘)에서 항목과 평가 방법에 대한 조사는 [13, 14]등에 잘 기술 되어있다. 비디오 감시시스템의 주요 평가항목들은 다음과 같다.

- **물체 정합** : 시스템에서 탐색된 물체와 검증용 물체(Ground truth) 간의 차이를 평가.
- **영상분할** : 비디오에서 물체를 추출하였을 때 검증 데이터와의 차이평가. 여기에는 추출된

물체의 기하학적 충실도, 추출된 물체의 화소 수에 의한 정밀도(precision), 민감도(sensitivity), F-value ($F = \frac{precision * recall}{precision + recall}$) 등이 적용됨.

- **물체추출** : 추출물체의 수를 이용한 측정값, 물체의 정합을 이용하는 방법 등.
- **추출된 물체의 위치에 대한 평가** : 물체 圖心(centroid)의 상대적인 거리, 물체면적의 상대적인 비율.
- **물체 추적** : 추적된 물체의 궤적에 따른 평가, 물체를 추적하여 새로운 물체를 찾을 때까지의 지연시간에 대한 평가, 궤적의 불완전성에 대한 평가, 궤적의 추출비율, 궤적의 끊어짐 정도 등에 대한 평가지표를 사용.
- **사건추출(Event Detection)**: 정밀도, 민감도, F-value 등을 적용.

2. 벤치마킹

자동감시를 위한 다양한 벤치마크 비디오 자료가 마련되었고, 여러 알고리즘들이 앞서 언급한 방법에 의하여 평가되었다. 본 단원에서는 어떠한 문제들이 주요 벤치마킹 자료로 이용되는지 살펴본다.

PETS^[14] : (Performance Evaluation of Tracking and Surveillance) 2000년부터 비디오 감시에 관련된 문제를 평가 하고, 알고리즘 및 시스템을 평가하기 위한 비디오 데이터베이스를 제공해오고 있다. 각 비디오 데이터는 검증정보가 같이 제공 되었다. 연도별 벤치마킹 항목은 다음과 같다.

- 2000년 : 실외 영상에서 단일카메라로 사람 및 차량을 추적하는 문제
 - 2001년 : 실외 영상에서 두 대의 카메라(1대의 이동 카메라, 1대의 전(全)방향 카메라(Omnidirectional Camera)로 사람 및 차량을 추적하는 문제.
 - 2002년 : 실내 사람 추적 및 계수문제.
 - 2003년(PETS-VS) : 실외영상(축구장)에서 사람추적. 대상지역의 영상에서 축구 선수를 모든 프레임에서 개인별로 추적하는 문제.
 - 2003년(PETS-ICVS) : 실내에 사람의 개인별 구분(6명), 동작, 표정을 인식하는 문제. 10개의 동작, 4가지의 표정. 1대의 전역 카메라와 2대의 일반 카메라영상이 주어짐.
 - 2004년 : CAVIAR 프로젝트 비디오 데이터로, 혼자 걷기, 사람과 만남, 쓰러짐/기절함, 윈도우 쇼핑, 공공장소에 짐을 두고 가는 행위, 여러 사람이 같이 걸어가는 것, 만나서 같이 가는 것, 만나서 같이 가다 헤어지는 것, 두 사람이 싸우는 것 등의 문제.
 - 2005년 : 물 위에서의 물체 탐색 및 추적
 - 2006년 : 공공장소(지하철역)에서 물건을 두고 가는 시나리오로 7가지 다른 경우의 비디오를 세트로 구성되어있다. 문제는 물체를 두고 갔을 때 이를 탐색하고 경보를 주는 것이다.
 - 2007년 : 안보와 범죄와 관련된 사건을 탐색하는 것이다. 본 문제와 관련된 비디오 데이터는 한사람이 서성이는 사건, 한사람이 감시 대상지역에 가방을 들고 와서 서성이는 사건, 한 사람이 가방을 들고 와서 다른 사람에게 전달하고 둘 다 대상지역에서 벗어나는 사건, 네 사람 중 한사람이 배낭을 들고 와서 놓고 다른 사람이 가지고 가는 사건, 한사람이 배낭을 들고 와서 놓고 다른 사람이 이를 훔쳐가는 사건, 두 사람이 두개의 큰 가방을 가지고 와서 바닥에 놓고 행인에게 길을 안내하는 사이 한사람(도둑)이 몰래 훔쳐가는 사건, 한사람이 두개의 가방을 들고 와서 하나만 가져갔다 다시 와서 나머지를 가져가는 일, 그리고 한사람이 큰 가방을 들고 와서 바닥에 놓고 가버리는 일 등으로 구성되어 있다.
 - 2009년 : 군중(crowd)과 관련된 문제이다. 첫 번째 세트는 사람 수 계수와 군중의 밀도계산에 관련된 몇 가지 난이도의 문제, 두 번째 세트는 사람들 중 각각을 추적하는 문제를 여러 난이도에 따라 주어지며, 세 번째 세트는 영상 시퀀스에서 사람들의 흐름을 각각 판단하여 내는 문제이고, 마지막 문제는 앞에서 주어지는 비디오로부터 걷는 것, 달리는 것, 대피하는 것, 지역적으로 대피하는 것, 군중이 형성되는 것 등의 사건을 탐색하는 것이다.
- VACE** : (Video Analysis and Content Extraction) 비디오로부터 내용을 추출하고 사건을 인식하는 창의적인 알고리즘 개발 및 구현을 목표로 설정되었다. 1단계와 2단계의 프로그램을 통하여 비디오 내용분석에 많은 진전을 가져왔다. 미국의 표준기술연구소(NIST)의 지도로 수행되었다.
- TRECVID** : NIST에서 주관하는 TREC(Text REtrieval Conference)과 같은 형식을 비디오 정보추출 평가 (Video Retrieval Evaluation)에

연장시킨 것으로, TREC Video Retrieval Evaluation을 나타내는 말이다. 2003년부터 2008년까지 뉴스 등 다양한 비디오에 대한 정보 분석 알고리즘을 평가하였다. 이중 비디오 감시와 관련된 2008년도 감시상황의 사건추출(Surveillance event detection)에 대해서만 요약한다. 이 평가의 목적은 컴퓨터가 사건을 인식하도록 하는 컴퓨터 비전기술 발전을 도모하도록 하는데 있으며, 영국 내무부에서 마련한 100시간(5대의 카메라 x 2시간 x 10일) 분량의 런던 Gatwick 공항 감시용 비디오로부터 사건을 추출하는 것을 평가하였다. 감시 비디오에서의 사건(event)은 다음과 같이 정의 되었다: 뛰는 것, 휴대전화 통화, 물건 놓는 것, 사람 만나는 것, 헤어지는 것, 껴안는 것, 손으로 가리키는 것, 엘리베이터 잘못 들어가는 것, 사람들의 흐름에 거스르는 것, 그리고 사진 찍는 것. 여기에 정의된 문제들은 현재의 기술 수준에서 난이도가 높은 것들이다.

전반적인 벤치마킹 평가결과를 보면 달리는 것과 물건 놓은 것의 탐색은 다른 항목에 비하여 잘 추출된다고 볼 수 있으나 대부분의 사건은, 그들을 정확하게 인식하기 위해서 많은 연구가 필요한 상태이다.

기타 : VERAAE, CHIL ETISEO등 비디오 분석에 관련된 벤치마킹 프로젝트가 수행되었다. (PETS2006 참조)

VI. 결론

비디오 분석을 통한 지능형 비디오 감시 기술은 산업체 및 학계로 부터 많은 관심을 받고 연

구되고 있다. 카메라, 센서, 컴퓨터의 등 H/W의 발달에 따라 비디오 감시시스템이 많이 발전되었으나 아직 상용제품으로의 적용과 학계의 연구결과는 차이가 있다. 특히 복잡한 사건을 인식하는 문제는 많은 연구가 필요 할 것으로 보인다. 이러한 문제의 해결은 컴퓨터 비전, 패턴인식, 기계학습 등의 학문적 발달에 기대를 걸고 있다. 산업체에서는 우선 문제를 제한시켜 신뢰성 있게 적용할 수 있는 세부문제를 발굴하여 적용하고, 난이도가 높은 문제는 학계와 연계하여 지속적인 연구를 추진 할 필요가 있다고 본다.

참고문헌

- [1] A. Hampapur, et. al, Video Analytics in Urban Environments, Proc. AVSS 2009, pp.129-133.
- [2] 산은경제연구소 산업분석 2팀, DVR 산업의 최근 동향과 시사점, 2008. 3. 6.
- [3] C. Stauffer and W. Grimson, Learning Patterns of Activity Using Real-Time Tracking, IEEE Trans. PAMI, v.22, n.8, 2000, pp.747-757.
- [4] H.Tao, H. Sawhney, and R. Kumar, Object Tracking with Bayesian Estimation of Dynamic Layer Representations, IEEE Trans. PAMI, v.24, n.1, 2002, pp.75-89.
- [5] M. Isard and A. Blake, A Smoothing Filter for Condensation, ECCV '98, Vol. I, LNCS 1406, 1998, pp.767-781.
- [6] D. Comaniciu, V. Ramesh, Mean Shift And Optimal Prediction For Efficient Object Tracking Proc. ICIP, V.3, 2000. pp.70-73.
- [7] D. Ramanan, D. Forsyth, and A. Zisserman,

- Tracking People by Learning Their Appearance, IEEE Trans. PAMI, v.29, n.1, 2007, pp.65-81.
- [8] H. Kim, J. Romberg and W. Wolf, Multi-Camera Tracking on a Graph Using Markov Chain Monte Carlo, Proc. 3rd ICDS 2009.
- [9] N. Anjum, A. Cavallaro, Trajectory association and fusion across partially overlapping cameras, Proc. AVSS2009.
- [10] Y. Hu, et. al., Action Detection in Complex Scenes with Spatial and Temporal Ambiguities, Proc. 2009 ICCV.
- [11] I. Huerta et al., Detection and Removal of Chromatic Moving Shadows in Surveillance Scenarios, Proc. 2009 ICCV.
- [12] A. Prati et al., Detecting Moving Shadows: Algorithms and Evaluation, PAMI, v.25, n.7, 2003, pp.918-923.
- [13] A. Baumann, et al., A Review and Comparison of Measures for Automatic Video Surveillance Systems, EURASIP Journal on Image and Video Processing, 2008.
- [14] Proc. Workshops on Performance Evaluation of Tracking and Surveillance (PETS). 2000 ~2009.

저자소개



한 준 희

1979년 2월 한양대학교 전자공학과 공학사
 1981년 2월 서울대학교 전자공학 석사
 1988년 4월 미시간대학교, 앤아버, 공학박사
 1988년 10월 ~2009년 10월 현재 포항공과대학교
 컴퓨터공학과 교수

주관심 분야 : 컴퓨터 비전, 비디오 감시, 패턴인식