

지능형 영상감시를 위한 원거리 사람 인식 테스트 데이터베이스

문기영*, 유장희*

요 약

원거리 사람 인식은 영상 분석의 주요 연구 분야로서, 많은 연구자들이 CCTV 영상에서 사람 추적이나 행위를 분석하여 얼굴인식, 보안 상황 추출 등을 연구하고 있다. 이러한 연구와 관련하여 지능형 영상 감시 알고리즘의 성능을 측정하고 평가하기 위한 테스트 데이터베이스가 있어야 한다. 본 논문은 국외에서 소개된 지능형 영상감시를 위한 원거리 사람 인식용 영상 데이터베이스를 분석하고, 국내 상황을 고려한 테스트용 영상 데이터베이스 조건을 간략하게 제안한다.

I. 서 론

원거리 사람 인식은 컴퓨터 비전 분야의 주요 연구 분야이며, 원거리 사람 인식, 로봇 비전, 지능형 영상감시 등 여러 응용 분야에서 활용될 수 있다. 원거리 사람 인식에서 주 인식 대상 객체는 사람이며 경우에 따라서는 사람과 관련 물건 등도 인식 객체에 포함되며, 이들 객체들은 설정된 보안 상황에 따라 휴먼의 행위(이벤트)가 검출된다. 이러한 휴먼 추적 및 행위 연구와 관련하여 알고리즘의 성능을 측정하고 평가하기 위해 다양한 데이터베이스가 존재한다.

본 논문은 기존에 주로 국외에서 구축된 비디오 영상 데이터베이스를 분석하고 국내 실정에 맞는 원거리 사람 인식용 데이터베이스 구축을 위한 고려사항을 기술한다.

II. 주요 원거리 사람 인식 데이터베이스

지능형 영상감시를 위한 원거리 사람 인식 데이터베이스 종류는 PETS나 AVSS, BEHAVED[9], ViSOR [11]와 같은 객체인식 관련 학회나 학교를 통해 제공되는 데이터베이스와 i-LIDS와 같이 정부기관이나 특정 단체에서 해당 기관의 요구사항을 중심으로 테스트와 기술 발전을 위해 제공하는 데이터베이스가 있다.

학회 제공 데이터베이스의 경우, 객체 및 휴먼 추적 연구와 관련하여 매년마다 많은 컨퍼런스와 워크숍이 개최되고 있으며, 각 알고리즘을 평가하기 위해 비디오 영상 데이터베이스를 제공해오고 있다. 대표적으로 IEEE Workshop on Performance Evaluation of Tracking an Surveillance (PETS), IEEE Conference on Advanced Video and Signal based Surveillance (AVSS), ViSOR(Video Surveillance Online Repository)는 학술대회나 학교 연구실에서 응용에 따라 각 알고리즘을 평가하기 위해 비디오 영상 데이터베이스를 제공하고 있다. 각 비디오 영상 데이터베이스는 실내의 환경에서의 사람 및 자동차의 추적, 배회 및 바닥에 방치된 가방 등의 이벤트 관찰 등의 목적을 가지고 구축되었다. 특히 PETS는 2000년부터 영국 리딩대학이 주축이 되어, 매해 데이터베이스를 발표하고 있다.

i-LIDS는 AVSS2007 학회에서 제공한 데이터베이스를 근간으로, 영국 내무부(Home Officer)에서 대테러 방지와 경찰력 보강을 목적으로 효율적인 비디오 분석(VA) 개발을 촉진하고 평가하기 위해 CAST(Centre for Applied Science and Technology)라는 조직을 통해 원거리 사람 인식 데이터베이스를 제공한다. I-LIDS는 영국 정부의 6가지 요구사항을 중심으로 6가지 시나리오 기반 비디오 영상 데이터베이스 세트로 구성되어 있다[10].

본 논문은 지식경제부 산업융합원천기술개발사업(10039149)으로 지원된 연구결과입니다.

* 한국전자통신연구원 사이버융합보안연구단 휴먼인식기술연구팀 (kymoon@etri.re.kr, jhy@etri.re.kr)

다음은 대표적인 원거리 사람 인식 데이터베이스인 PETS와 i-LIDS를 살펴본다.

2.1 PETS2000

지능형 영상분석을 위해 PETS 워크숍이 2000년 3월에 프랑스에서 처음으로 개최되었다[1]. 이 워크숍에서 PETS 워크숍 위원회 명의로 워크숍 참석자를 위해 데이터베이스가 공개되었다. 데이터베이스는 2가지 종류로, 첫 번째는 학습용으로, 두 번째 것은 테스트용이다.

- 목적 : 주차장 내의 움직임이 있는 자동차와 사람을 추적
- 영상 크기 : PAL (768×576 픽셀) - 25 frames / sec
- Camera Calibration을 위한 자료 포함
- 환경 : 실외 환경 (낮)



[그림 1] PETS2000 테스트 데이터베이스

2.2 PETS2001

2회 PETS2001에서는 실외에서 2대의 카메라로 사



[그림 2] PETS2001 카메라1 데이터베이스

람과 차량을 추적하는 데이터베이스를 공개하였다[2].

- 목적 : 실외 환경에서 자동차와 사람을 추적
- 영상 크기 : PAL (768×576 픽셀) - 25 frames / sec
- Camera Calibration을 위한 자료 포함
- 환경 : 실외 환경
- 특징 : 위치가 다른 두 대의 카메라로 동일 장면을 획득 (동기화 필요)



[그림 3] PETS2001 카메라2 데이터베이스

2.3 PETS2002



[그림 4] PETS2002 데이터베이스

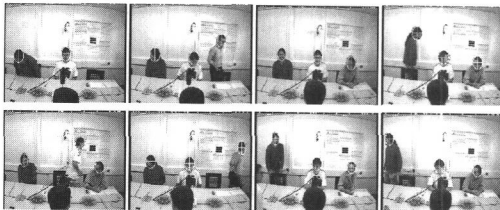
2002년 3번째 PETS는 덴마크에서 개최되었다. 워크숍에서는 상점 내부에 설치된 CCTV에서 상점 외부의 사람들을 찍은 데이터베이스가 공개되었다[3]. 연구자들은 이 데이터베이스를 참조하여 사람 추적 및 사람 카운팅 등의 연구 과제를 진행하였다.

- 목적 : shop 내부에 설치된 카메라로부터 shop window 밖의 사람을 추적
- 영상 크기 : 640×240 픽셀 (25 frames / sec)
- 환경 : 실내 환경

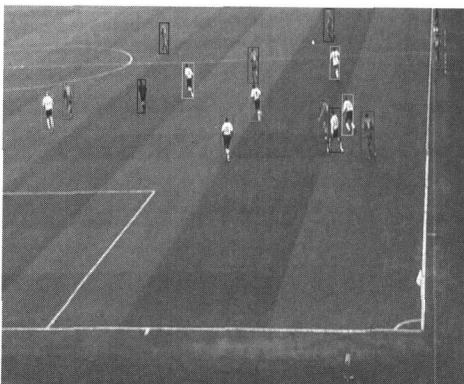
2.4 PETS2003

2003년 오스트리아에서 개최된 네 번째 PETS는 주로 얼굴인식에 관련된 연구결과가 많이 발표되었다[4]. PETS2003에서는 미팅 룸에서 사람들의 머리를 추적하고 인식하거나 사람의 행동이나 표정 등을 검출할 수 있는 데이터베이스가 공개되었다. 이외에 실외에서 움직이는 사람 추적을 위해 학회에서 논의는 되지 않았지만 축구장 영상도 공개되었다.

- 목적 : Interaction이 있는 사람의 얼굴 검출 및 사람 추적
- 영상 크기 : PAL (768×576 픽셀) - 25 frames / sec



(그림 5) PETS2003 데이터베이스



(그림 6) PETS2003 EXTRA 데이터베이스

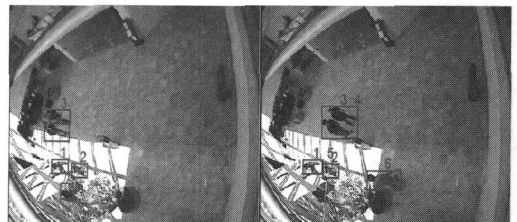
- Camera Calibration을 위한 자료 포함
- 환경 : 실내 및 실외 영상
- 특징 : 실내 meeting room에서의 영상은 근거리에서 촬영 (얼굴 검출이 주목적)

2.5 PETS2004

2004년 PETS2004는 CAVIAR(Context Aware Vision using Image-base Active Recognition) 컨소시엄이 개최를 하고, 다른 해와 달리 CAVIAR에서 데이터베이스를 제작하여, 2004년 데이터베이스를 CAVIAR 데이터베이스라고도 한다[5].

PETS2004 데이터베이스는 walking, browsing, collapse, leaving object, meeting, fighting 등 6가지 시나리오에 대해 총 28가지 비디오 클립을 만들어 공개하였다

- 목적 : 공공장소에서의 사람의 특정 행위를 관찰하기 위한 목적
- 영상 크기 : 384×288 픽셀 - 25 frames / sec
- 영상 왜곡이 있음.
- 환경 : 실내 영상 (INRIA 로비에서 촬영 - 2층 높이에서 1층에 있는 사람을 촬영)
- 특징 : 사람 추적과 함께 6가지 시나리오에 대한



(그림 7) PETS2004 데이터베이스

행위 관찰이 주목적임

- ✓ 시나리오 1 - 직선적으로 걷는 행위
- ✓ 시나리오 2 - 계시판을 둘러보는 행위
- ✓ 시나리오 3 - 의자에 앉는 행위
- ✓ 시나리오 4 - 물건을 버리는 행위
- ✓ 시나리오 5 - 마주치는 그룹의 사람
- ✓ 시나리오 6 - 싸우는 행위

- ✓ 사람 추적과 함께 사람이 들고 있는 가방을 분류하고 가방이 사람과 떨어져서 놓여지는 행위를 관찰하는 것이 주목적임.
- ✓ 다수의 카메라(4대)에서 촬영된 동일 시점에서 다양한 View의 영상을 제공

2.6 PETS2006

2006년 PETS2006은 미국 뉴욕에서 개최되었다. PETS2006의 주 주제는 공공장소에서 투기된 수화물을 검출하는 것이다[6]. 데이터 종류는 모두 7가지로 주로 기차역에서 수하물 방치나 검출을 목적으로 데이터베이스가 제공되었다.

- 목적 : 실제 공공장소(기차역)의 영상에서 수화물을 들고 있는 사람을 추적
- 영상 크기 : 720×576 픽셀 - 25 frames / sec
- Camera Calibration을 위한 자료 포함
- 환경 : 실내 영상 (기차역에서 다수의 카메라로 다양한 View의 영상을 획득)
- 특징

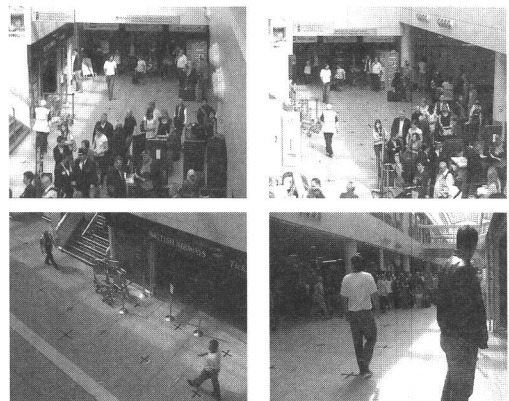


[그림 8] PETS2006 데이터베이스

2.7 PETS2007

PETS2007은 보안 및 범죄 상황 검출이 주요 주제였다[7]. 데이터베이스는 배회, 수화물 도난, 버려진 수화물 등 3가지 시나리오에 맞는 8가지 종류의 데이터베이스가 제공되었다.

- 목적 : 공공장소의 영상에서 3가지 이벤트(사람의 배회, 가방 도난, 놓여 있는 가방)를 검출
- 영상 크기 : 720×576 픽셀 - 25 frames / sec
- Camera Calibration을 위한 자료 포함
- 환경 : 실내 영상 (다수의 카메라로 다양한 View의 영상을 획득)
- 특징
 - ✓ 배회하는 사람이나 가방의 도난, 가방을 놓고 사람이 사라지는 행위 등을 관찰하는 것이 주목적임
 - ✓ 다수의 카메라(4대)에서 촬영된 동일 시점에서 다양한 View의 영상을 제공



[그림 9] PETS2007 데이터베이스

2.8 PETS2009

PETS2009에서는 군중 이미지 자동분석이 주 주제였다. 캠퍼스, 병원 등 공공장소에서 군중들의 행위, 분석이나 사람 추적을 통해 보안 이벤트를 감지하는 시나리오로 데이터베이스가 만들어졌다[8].

데이터베이스는 1개의 학습 데이터와 3개의 테스트 데이터베이스가 제공되었다. 3개의 테스트 데이터베이스는 사람 카운팅과 밀도 계산, 사람 추적, 사람 흐름 추적 등의 주제로 제작되었다.

- 목적 : 실외 환경(캠퍼스)의 다수의 군중으로부터 사람 추적 및 밀집도 조사
- 영상 크기 : 600×350 픽셀
- Camera Calibration을 위한 자료 포함
- 환경 : 실외 영상 (다수의 카메라로 다양한 View의 영상을 획득)
- 특징
 - ✓ 다수의 카메라(8대)에서 촬영된 다양한 View의 영상을 제공
 - ✓ 다양한 밀집도의 군중 영상 포함



(그림 10) PETS2009 데이터베이스

2.9 i-LIDS

i-LIDS 데이터베이스는 영국 내무부(Home Officer)에서 대테러 방지와 경찰력 보강을 목적으로 효율적인 비디오 분석(VA) 개발을 촉진하고 평가하기 위해 원거리 사람 인식 데이터베이스를 제공한다. i-LIDS는 6가지 시나리오의 영상 데이터베이스를 제공한다[10].

데이터베이스는 각 시나리오 별로 MJPEG 포맷으로 제공된다.

- 목적 : 공공장소에서 대테러 방지를 위한 지하철 영상에서의 가방 추적과 부정주차 차량 검출
- 영상 크기 : 720×576 픽셀 - 25 frames / sec
- 시나리오 :
 1. 버려진 수하물(짐) 검출

2. 출입구 추적(감시)
3. 새로운 기술 분야 (근적외선, 중파 열상 이미지, 장파 열상 이미지)
4. 주차차 차량 검출
5. 제한구역 또는 완충지역(sterile zone) 모니터링
6. 멀티카메라 트래킹

• 특징

가. 일반

- 1) 다른 조건(일별, 날씨, 백그라운드)에서 24시간 촬영
- 2) 지상실측정보(Ground Truth)는 각 비디오 파일로 제공한다. 지하철 영상

나. 지하철 영상

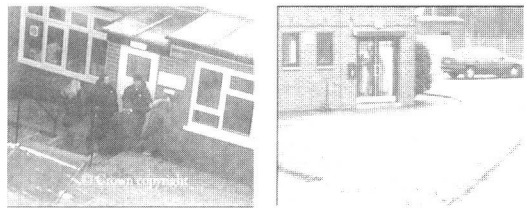
- 1) 방치되어 있는 가방을 검출하는 것이 목적
- 2) 3개의 영역(near, mid, far)을 나눔

다. 자동차 영상

- 1) 부정 주차 차량을 검출하는 것이 목적
- 2) 주차 영역과 주차 금지 영역을 나눔.
- 3) 주간 야간 영상이 제공됨
- 4) 실외 영상으로 카메라의 흔들림이 나타남



(그림 11) i-LIDS 시나리오 1 버려진 수하물 검출



(그림 12) i-LIDS 시나리오 2 출입구 감시



(그림 13) i-LIDS 시나리오 3 근적외선



(그림 14) i-LIDS 시나리오 4 주차 차량 검출



(그림 15) i-LIDS 시나리오 5 제한구역 모니터

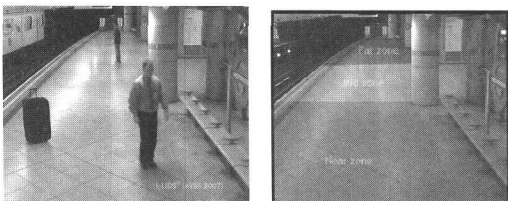


(그림 16) i-LIDS 시나리오 6 다중카메라 추적

2.10 지상실측정보 (Ground Truth) 이해

원거리 사람 인식 데이터베이스는 영상을 해석하고 검증하기 위해 지상실측정보를 XML이나 text 파일 형식으로 정보를 제공한다. 원거리 사람 인식 데이터베이스에서 지상실측정보는 각 비디오 영상이 가지고 있는 이벤트와 영상에서 표현된 객체들의 위치나 크기, 위치, 상태 정보를 제공한다.

그림 17과 같이 이벤트를 지상실측정보로 제공하는 i-LIDS 데이터베이스의 버려진 수화물 영상을 예를 들어 보면[10]:



(그림 17) 수화물 검출 시나리오

(표 1) XML스키마

Filename - Name of video file (inc .qtl suffix)
 Stage - <1|2> Camera view (1=Platform, 2=Walkway)
 Duration - <hh:mm:ss> Length of the clip
 AlarmEvents - Number of alarm events in the clip
 PeriodOfDay - <Low|Medium|High> Low=6-10pm;
 Medium=10am-2pm;
 High=3-7pm
 Alarms -
 Alarm -
 StartTime - <hh:mm:ss> Time elapsed in clip
 when alarm condition met
 AlarmDescription - <Abandoned object>
 AlarmDuration - <hh:mm:ss> Length of time
 for which the alarm
 condition remains valid
 ObjectZone - <Near|Mid|Far> Zone of object
 within detection area
 ObjectDescription - <Bottle|Drinks can|Family
 suitcase|Newspaper|Paper|Rucksack|Sports
 Bag> Type of object causing
 alarm
 SuspectDressCode - <Smart|Casual> Dress
 code of object owner

(표 2) 작성예

```
<?xml version="1.0" ?>
lidsLibraryIndex>
<Library>
<scenario>Abandoned Baggage</scenario>
<dataset>AVSS 2007</dataset>
<libversion>1.0</libversion>
<clip>
<filename>Easy.mov</filename>
<Stage>1</Stage>
<Duration>00:03:38</Duration>
<AlarmEvents>1</AlarmEvents>
<PeriodOfDay>Low</PeriodOfDay>
<Alarms>
<Alarm>
<StartTime>00:03:00</StartTime>
<AlarmDescription>Abandoned object </AlarmDescription>
<AlarmDuration>00:00:12</AlarmDuration>
<ObjectZone>Near</ObjectZone>
<ObjectDescription>Family suitcase </ObjectDescription>
<SuspectDressCode>Smart</SuspectDressCode>
</Alarm>
</Alarms>
</clip>
</Library>
</lidsLibraryIndex>
```

해설 : 3분 38초의 Easy.move 파일에서 버려진 객체를 비디오에서 3분에 12초간 알람을 알리고, 버려진 객체는 가족용 슈트케이스임.

III. 국내 원거리 사람 인식 데이터베이스 요건

앞장에서 대표적인 휴먼 및 객체인식 데이터베이스를 살펴보았다. 이들 데이터베이스들은 비교적 오랜 기간 동안 연구와 피드백을 통해 데이터베이스를 제작하고 개선해 왔다. 국내의 많은 연구자들도 이들 데이터베이스를 이용하여 연구 및 개발을 수행하고 있다. 그러나 이들 데이터베이스는 국내의 환경이나 필요 요구사항을 충분히 반영하지 못하여 연구 및 개발이나 비디오 분석 장비 도입 시 검증에 어려움이 있다.

국내 원거리 사람 인식 데이터베이스 구성에서 고려해야 할 주요 특성은 다음과 같다.

- 환경적 요인
 - 사계절(봄, 여름, 가을, 겨울)이 뚜렷하고, 눈과 비 등 자연환경의 변화가 많다
 - 지하시설이 많음(지하철, 지하상가, 지하주차장 등)
 - 밀집된 건물이 많아 채광이 다양함
- 대상 요인
 - 대테러 보다는 치안에 중점
 - 스킨존, 지하철역, 공항, 아파트단지, 지하주차장 등 어린이 보호 등 치안 수요가 많음
 - 비디오 추적과 출입통제 수요가 많음
- 향후 기술추이 요인
 - 3D 영상 기기의 발전으로 3D 영상 분석이 필요
 - 야간 등 특수지역을 위한 열상영상 및 적외선 영상 분석 필요.

V. 결 론

PETS 데이터베이스는 2000년부터 원거리 사람 인식 분야에서 체계적인 연구와 방향을 제시하였다. 또한,

i-LIDS 데이터베이스는 학술적인 연구를 넘어서 실제적인 사용자 요구사항을 반영하고, 데이터베이스의 규모도 실제 상황에 근접한 수준의 환경을 제공하였다.

본 논문에서는 지능형 영상감시를 위한 원거리 사람 인식 분야의 대표적인 2가지 데이터베이스에 대해 살펴보고, 국내의 실사용자들이 요구할 수 있는 데이터베이스들의 특성들을 제시하여, 국내에서도 국내 환경에 적합하고 국제적인 인정을 받는 지능형 영상감시를 위한 원거리 사람 인식용 데이터베이스 구축을 기대한다.

참고문헌

- [1] PET2000, <http://www.cvg.reading.ac.uk/PETS2001/pets2000.pdf>.
- [2] PETS 2001, <http://pets2001.visualsurveillance.org/>.
- [3] PETS 2002, <http://pets2002.visualsurveillance.org/>.
- [4] PETS-ICVS'2003, <http://pets2003.visualsurveillance.org/>.
- [5] PETS-ECCV'2004 - CAVIAR <http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/>.
- [6] PETS'2006, <http://www.cvg.cs.rdg.ac.uk/PETS2006/data.html>.
- [7] PETS'2007, <http://www.cvg.cs.rdg.ac.uk/PETS2007/data.html>.
- [8] PETS'2009, <http://www.cvg.cs.rdg.ac.uk/PETS2009/data.html>.
- [9] BEHAVE <http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/INTERACTIONS/>.
- [10] i-Lids, www.ilids.co.uk.
- [11] ViSOR, http://imagelab.ing.unimore.it/visor/video_categories.asp.

〈著者紹介〉



문기영 (Ki Young Moon)

정회원

1986년 2월: 경북대학교 전자공학과 졸업

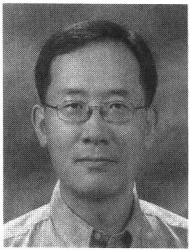
1989년 2월: 경북대학교 전자공학과 석사

2006년 2월: 충남대학교 컴퓨터학과 이학박사

1992년 1월~1994년 3월: 대우정보시스템 전임연구원

2010년 8월~2011년 8월: 미국 Univ. of California, Irvine 초빙연구원

1994년 3월~현재: 한국전자통신연구원 바이오인식기술연구팀장 역임/휴먼인식기술연구팀 책임연구원
<관심분야> 바이오인식, 지능형 영상보안, 정보보호



유장희 (Jang-Hee Yoo)

1988년: 한국외국어대학교 물리학과 졸업 (학사)

1990년: 한국외국어대학교 전산학과 졸업 (석사)

2004년: 영국 University of Southampton 전자 및 컴퓨터과학과 졸업 (박사)

1989년 11월~현재: 한국전자통신연구원 휴먼인식기술연구팀장

2007년 03월~현재: 과학기술연합대학원대학교(UST) 정보보호공학 전공 책임교수 (겸임)

<관심분야> 컴퓨터 비전, 생체인식, 지능형 영상보안, HCI 및 지능형 로봇 등