베이지안 네트워크를 이용한 동영상 기반 라이프 로그의 분석 및 의미정보 추출

정태민, 조성배
연세대학교 컴퓨터과학과
realone@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

Context Extraction and Analysis of Video Life Log Using Bayesian Network

Tae-min Jung°, Sung-Bae Cho
Dept. of Computer Science, Yonsei university

요 약

최근 라이프 로그의 수집과 관리에 관련된 연구가 많이 진행 중에 있다. 또 핸드폰 카메라, 디지털 카메라, 캠코더 등의 발전으로 자신의 일상생활을 비디오로 저장하고, 인터넷을 통해 공유하는 사람도 증가하고 있다. 비디오 데이터는 많은 정보를 포함하고 있는 라이프 로그의 한 예로. 동영상의 촬영 및 수집이 활발해짐에 따라 동영상의 메타정보를 생성하고, 이를 이용해 동영상 검색과 관리에 이용하려는 연구들이 진행 중이다. 본 논문에서는 라이프 로그를 수집하고 수집된 동영상과 라이프 로그를 이용하여 의미정보를 추출하는 시스템을 제안한다. 의미정보란 사용자의 행동을 나타내는 정보로써 컴퓨터 사용, 식사, 집안일, 이동, 외출, 독서, 휴식, 일, 기타로 9가지의 의미정보를 추출한다.

제안하는 방법은 사용자로부터 GPS, 가속도센서, 캠코더를 이용해 실제 데이터를 수집하고, 전처리 과정을 통하여 특징을 추출한다. 이때 추출될 특징은 위치정보와 사용자의 상태정보 그리고 영상처리를 통한 RGB와 HSL 색공간의 요소와 MPEG-7의 EHD(Edge Histogram Descriptor). CLD(Color Layout Descriptor)이다. 추출된 특징으로부터 사람 행동과 같은 불안정한 상황에서 강점을 보이는 확률모델 네트워크인 베이지안 네트워크를 이용하여 의미정보를 추출한다. 제안하는 방법의 유용성을 보이기 위해 실제 데이터를 수집하고 추론하고 10-Fold Cross-validation을 이용하여 데이터를 검증한다.

1. 서론

최근 모바일 디바이스는 개인의 생활 로그를 저장할 수 있는 기기로 자리 잡았다. 모바일 기기를 통해 GPS, 핸드폰 사용 내역, 사진, 비디오 등 다양한 로그를 수집할 수 있게 되었고, 이를 관리하고 분석하는 연구들이 많이 진행 중이다[1,2].

라이프 로그 중 비디오 데이터는 영상과 음성을 뿐만 아니라 시간을 포함하는 3차원 로그이다. 비디오 데이터를 이용함으로써 개인의 생활을 더 효과적으로 분석할 수 있을 것이다.

Hewlett-Packard Company에서는 동영상을 검색하고 관리하는 PVM(Personal Video Manager) 시스템을 제안하였다[2]. 이 연구에서는 웹 기반 어노테이션 툴을 제공하여 동영상에 사용자에게 직접 레이블링 하도록 하고 특징을 추출하여 동영상을 검색/관리하였다. PVM에서는 라이프 로그(동영상)의 분석에 동영상에서의 특징 추출만을 이용하였으나 본 논문에서는 수집된 여러 라이프 로그를 이용하여 향상된 성능의 의미정보(컨텍스트)를 생성하는 방법을 제안한다. 제안하는 방법이 유용한지를 보이기 위하여 동영상 검색 시스템을 구현하고 사용성 평가를 통해 시스템의 유용함을 보인다.

2. 베이지안 네트워크를 이용한 의미정보 추출

베이지안 네트워크는 다양한 추론기능과 불확실성에 강인한 특징을 가지고 있어 모바일 기기의 불확실한 정보로부터 사용자의 행동을 추론하기에 적합한 모델이다. 그리고 동적 베이지안 네트워크는 이러한 베이지안 네트워크의 강점을 가지며 추가로 시간의 흐름을 고려하는 확률 모델로, 동영상 정보에서의 특징 추출에 적합하다[3, 9].

본 논문에서는 의미정보를 추론하기 위해 동적 베이지안 네트워클 이용한다. 의미정보란 사람에게 의미 있는 정보 즉, 사람의 행동이나 특별한 사건을 나타내는 정보를 말하는 것으로 라이프 로그 분석에 중요한 정보이다. 그림 1은 제안하는 방법의 시스템을 설명한다.

제안하는 시스템은 먼저 센서를 통해 데이터를 수집하

고, 이를 전처리를 통해 특징을 추출한다. 그리고 이 특징을 바탕으로 동적 베이지안 네트워크를 이용하여 의미정보를 추론한다. 최종적으로 추론된 의미정보를 이용하여 라이프 로그의 관리 및 검색에 이용하는 인터페이스를 제공한다.

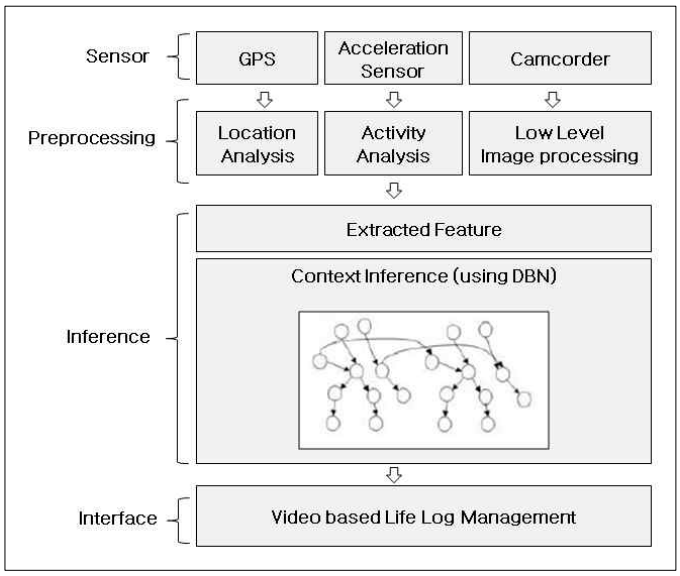


그림 1 제안하는 시스템 구상도

2.1 라이프 로그의 수집

본 논문에서 제안하는 시스템을 개발하기 위해 센서를 이용하여 데이터를 수집한다. 사용되는 센서의 종류는 GPS, 가속도 센서 그리고 캠코더이다. 수집 방법에는 특정한 제한상황을 두지 않고, 자유롭게 수집하도록 하였다. 비디오데이터는 사용자의 생활이 모두 기록되어지도록 가능한 모든 시간동안 촬영하였고, 가속도 센서는 x, y, z 방향의 3축 가속도 센서로 손목과 허리에 장착하여 사용자의 상태정보를 수집한다. GPS 센서는 항상 소지하고 사용자의 위치정보를 수집한다. 표1은 센서를 이용해 수집할수 있는 로그와 추출 가능한 특징을 나타낸다.

표1. 수집할 데이터와 특징 값

센서종류	수집로그	추출 가능한 특징 값
GPS 센서	위도, 경도	위치정보
가속도센서	움직임	안기, 서기, 뛰기, 걷기
캠코더	비디오 영상	RGB, HSL, Edge 정보

2.2 라이프 로그의 특징 값 추출

수집된 라이프 로그의 특징 값을 추출한다. GPS 데이터는 네이버맵 OPEN API를 이용하여 지도에 표기하고 이를 통해 레이블링 한다. 레이블링 할 목록은 “공항, 아파트, 버스정류장, 카페, 집, 마켓, 음식점, 거리, 영화관, 학교, 직장이다.

가속도 센서는 3축 가속도 센서로 허리와 손목에 각각 부착하여 룰을 이용하여 사용자의 상태 정보를 추출한다. 이때 추출될 수 있는 값은 ‘앉기, 서있기, 뛰기, 걷기’이다.

동영상 데이터는 전처리 과정으로 먼저 영상의 프레임을 추출하였다. 추출한 프레임을 바탕으로 영상처리를 실행한다. 추론될 특징은 MPEG-7의 EHD(Edge Histogram Descriptor)와 CLD(Color Layout Descriptor)와 RGB, HSL의 색 공간을 이용한다.
MPEG-7은 다양한 형태의 멀티미디어 정보에 대한 서술(묘사, 설명)을 표준화여 멀티미디어 콘텐트에 대한 서술 인터페이스를 제공한다[4].

EHD(Edge Histogram Descriptor)는 영상의 지역적 경계선의 분포를 나타낼수 있는 기술자이다. 추출된 프레임을 4 x 4 개수의 겹치지 않는 서브영상으로 나누어 각각의 경계선(0°(수직), 45°(대각선), 90°(수직), 135° (대각선), 비방향성)을 사용하여 영상의 공간적인 분포를 표현한다. 경계선을 각 3bit로 나타내어 4 x 4 x 5(종류) x 8(3bit)의 240빈을 가진 히스토그램을 생성한다. 본 논문에서는 경계선의 종류를 기준으로 히스토그램을 합한 값을 특징으로 추출하였다. 그림2는 경계선을 추출하는데 사용되는 5개의 경계선을 나타낸다.

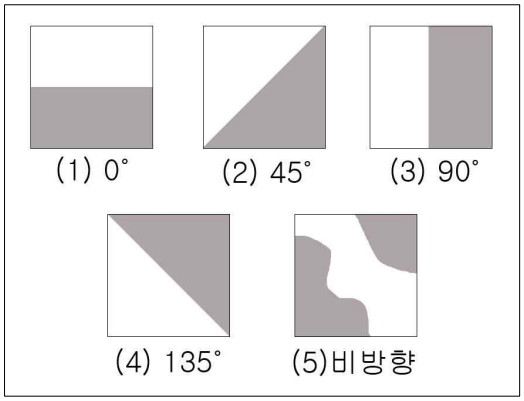


그림 2 EHD에 사용되는 방향성

CLD(Color Layout Descriptor)는 영상의 유사한 색상간의 히스토그램과 공간구조의 분포를 나타내는 MPEG-7에 정의된 기술자로, 구현이 간편하고 좋은 성능을 가진다[6]. 영상을 64(6 x 6)개의 겹치지 않은 서브영상으로 나누어 각 블록을 대표하는 색을 정한다. 색의 요소는 YCbCr 색공간의 값이고, 대푯값을 DCT 변환을 한 계수를 사용한다. 본 논문에서는 각각 6(Y), 3(Cb),

3(Cr)을 추출하였고, 각각에 대한 평균값을 사용하였다.

추출된 프레임을 RGB와 HSL 색공간의 이미지로 각각 변형시킨 후 평균값을 구하였다. 각각은 이미지에 포함된 색상과 채도 명도 정보를 가진다. 표2는 비디오 영상에서 추출된 특징을 나타낸다.

표 2 비디오 데이터의 추출된 특징

구분	특징	의미
RGB 색공간	Red	빨간색의 평균
	Green	초록색의 평균
	Blue	파랑색의 평균
HSL 색공간	Hue	색상 평균
	Saturation	채도 평균
	Lightness	명도 평균
MPEG-7 EHD	0°	수평선의 개수
	45°	45°도 대각선의 개수
	90°	세로선의 개수
	135°	135°도 대각선의 개수
	비방향	비방향성인 블럭의 개수
MPEG-7 CLD	Y1	1-3 값의 평균
	Y2	4-6 값의 평균
	Cb	Cb 값의 평균
	Cr	Cr 값의 평균

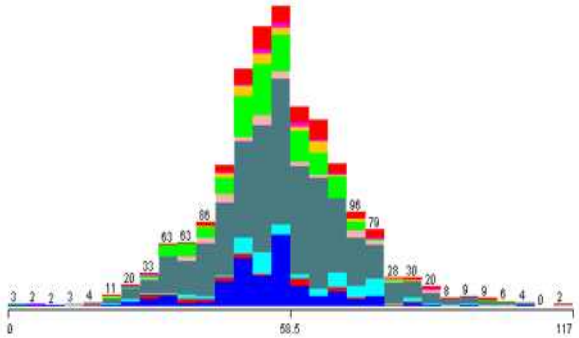
2.3 추출된 특징의 이산화

본 논문에서 추출된 특징 값의 대부분은 연속적인 수치이고, 그 값이 다양하게 분포 된다. 이를 베이지안 네트워크에서 사용하기 위해서는 이산화 과정이 필요하다. 이산화 방법으로는 평균값과 표준편차를 이용하였다. 4가지 등급("매우적음", "적음", "많음", "매우 많음") 다음 그림 3은 이산화 방법을 나타내고, 그림 4는 이를 통하여 이산화 한 결과를 나타낸다.

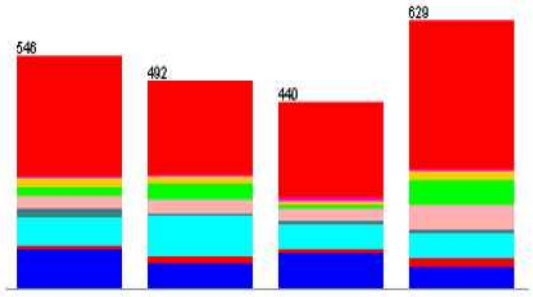
f_i //i번째 특징 값
$f_{i.val}$ //i번째 특징의 각 원소의 값
$f_{i.avg}$ //i번째 특징 값의 평균
$f_{i.stddev}$ //i번째 특징값의 표준편차

$IF\ f_{i.val} < f_{i.avg} - f_{i.stddev}\ THEN$
  매우적음
$ELSE\ IF\ f_{i.val} < f_{i.avg}\ TEHN$
  적음
$ELSE\ IF\ f_{i.val} < f_{i.avg} + f_{i.stddev}\ TEHN$
  많음
$ELSE$
  매우많음

그림 3 특징 이산화 수도코드



(가) 이산화 하기 전의 특징 (Hue:색상) 데이터 그래프



(나) 이산화 후의 특징 (Hue:색상) 데이터 그래프

그림 4 영상 특징(Hue:색상)의 이산화 그래프

2.4 베이지안 네트워크의 학습

베이지안 네트워크를 학습하기 위해 수집한 데이터를 ARFF(Attribute-Relation File Format)파일 형식으로 만들어 weka를 이용하여 베이지안 네트워크의 구조와 파라미터를 학습하였다. weka는 뉴질랜드의 Waikato 대학에서 데이터 마이닝의 여러 알고리듬을 사용하고 테스트할 수 있는 오픈 소스 툴로, 베이지안 네트워크의 학습 기능을 제공한다[8]. 본 논문에서는 weka에서 제공하는 베이지안 네트워크 학습 알고리듬 중 Hill Climbing를 이용하였다. 최대 부모 노드는 5개로 하였고, 데이터를 기반으로 구조와 파라미터 학습을 하였다. 그림 5는 학습된 베이지안 네트워크를 나타낸다.

3. 실험 및 결과

본 논문에서 제안한 시스템을 평가하기 위하여 20대 여대생으로부터 실제 데이터를 수집하고, 의미정보를 추출하였다. 센서에서 수집된 데이터와 각각의 데이터에 대한 행동도 레이블링 하였다. 모든 데이터에는 시간이 명시되어있어야 하고, 24시간 모두 수집하는 것을 원칙으로 하였다. 추론할 의미정보는 "컴퓨터, 식사, 집안일,

이동, 외출, 독서, 휴식, 일, 기타” 로 9가지 종류이다.

비디오 로그를 기준으로 하였고, 비디오 로그가 존재하지 않은 경우에는 다른 센서데이터를 삭제하였다. 총 2107개의 데이터가 수집되었으며, 분류된 행동의 수는 표 3과 같다.

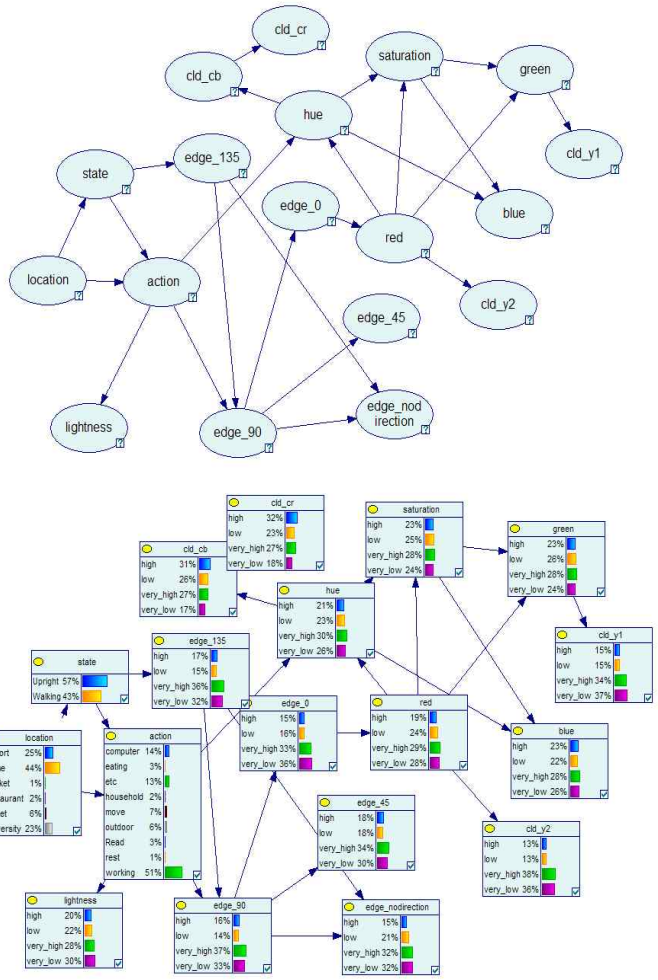


그림 5 학습한 베이지안 네트워크의 구조

표 3 수집된 행동의 수

행동	개수	행동	개수
컴퓨터	286	독서	67
식사	58	휴식	23
집안일	40	일	1071
이동	154	기타	279
외출	129	총합	2107

수집된 데이터에서 특징을 추출하고 추출된 특징의 개수를 조절하며 10-Fold Cross-validation 기법을 이용하여 테스트 하였다. 먼저 위치정보와 상태정보만을 실험을 하였다. 표 4는 실험 결과를 Confusion Matrix로 나타낸 것이다.

실험 결과는 올바르게 분류한 것이 69.77%로 다소 높게 나타났으나, 전체 분류의 50%를 가지고 있는 “일” 분류가 1071건 중 1건만이 잘못 판단되었다. 이것으로 “일”이라는 행동은 지역과 상태에 의존적임을 알 수 있었다. 그 외에 정확률과 재현율을 기준으로 보면 식사, 가사, 일, 기타의 행동에 영향을 미침을 알 수 있었다.

표 4 위치정보, 상태정보를 이용한 Confusion Matrix

a	b	c	d	e	f	g	h	i	
0	0	176	1	0	0	0	0	109	a = 컴퓨터
0	41	0	0	0	0	0	0	17	b = 식사
0	0	233	14	0	0	0	0	32	c = 기타
0	0	0	24	0	0	0	0	16	d = 집안일
0	12	0	17	58	23	0	0	44	e = 이동
0	0	0	0	0	44	0	0	85	f = 외출
0	0	0	1	0	0	0	0	66	g = 독서
0	0	0	1	0	0	0	0	22	h = 휴식
0	0	0	1	0	0	0	0	1070	i = 일

표 5 비디오 특징을 이용한 Confusion Matrix

a	b	c	d	e	f	g	h	
201	3	38	2	20	15	6	1	a = 컴퓨터
14	7	11	2	16	7	1	0	b = 식사
75	5	141	2	30	21	4	1	c = 기타
7	0	12	10	5	4	2	0	d = 집안일
32	4	33	4	49	28	4	0	e = 이동
20	3	20	2	17	61	6	0	f = 외출
14	0	19	0	13	12	9	0	g = 독서
12	1	4	0	3	3	0	0	h = 휴식

표 6 모든 정보를 이용한 Confusion Matrix

a	b	c	d	e	f	g	h	i	
157	0	65	0	1	4	4	0	55	a = 컴퓨터
2	37	0	0	4	2	2	0	11	b = 식사
68	0	173	7	2	3	3	0	23	c = 기타
1	0	5	14	6	0	1	0	13	d = 집안일
5	9	2	8	78	14	1	0	37	e = 이동
0	0	0	0	8	75	3	0	43	f = 외출
1	1	0	1	0	11	13	0	40	g = 독서
0	0	0	1	0	0	0	1	21	h = 휴식
24	1	10	1	1	34	9	0	991	i = 일

다음으로 영상 특징 값을 이용한 실험을 하였다. 앞의 결과 “일”의 행동은 지역정보와 상태정보에 의존적이고, 데이터 수가 50%를 영상에서 추출한 특징이 의미정보 추론에 영향을 주는지 알아보기 위해 ‘일’의 분류를 삭제하고 실험하였다. 다음 표 5은 실험 결과를 나타낸다. 실험 결과 인식률은 46%로 다소 낮게 나왔으나 위치정보와, 상태정보만을 이용한 것보다 컴퓨터와 외출 독서의 결과는 좋아진 것을 볼 수 있었다.

이번에는 모든 특징을 합하여 실험을 하였다. 이번에는 “일” 데이터를 삭제하지 않고, 이용하였다. 표6은 모든 특징값을 이용한 실험 결과를 나타낸다. 실험결과 74.04%의 인식률을 보였다.

4. 결론 및 향후연구

본 논문에서는 비디오 데이터를 포함한 라이프 로그를 수집하고, 이를 베이지안 네트워크를 이용하여 라이프 로그의 의미정보를 자동 생성하였다. 생성된 라이프 로그 의미정보는 검색 또는 사용자의 생활 요약에 활용될 수 있을 것이다.

향후 연구로는 가속도 센서와 GPS센서를 좀 더 활용하여 걷기, 뛰기, 머무름 등의 행동을 인식하고자 한다. 이러한 행동인식은 LBS(Lcoation Based Service)를 제공하는데 타당성을 부여해 줄 수 있을 것이다.

감사의 글

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2009-0083838)

참고문헌

[1] J. Gemmell, G. Bell, and R. Lueder, “MyLifeBits: Personal database for everything,” *Communications of ACM*, vol. 49, no. 1, pp. 88-95, 2006.

[2] Peng Wu and Pere Obrador , "Personal video manager: managing and mining home video collections, " *Proc of SPIE* vol. 5960, pp. 775-785, 2005.

[4] Thomas Sikora, "The MPEG-7 Visual Standard for Content Description. : An Overview," *IEEE Trans. On Circuits And Systems For Video Technology*, vol. 11, no. 6, 2001.

[5] Ankush Mittal, Sumit Gupta, "Automatic content-based retrieval and semantic classification of video content," *International Journal on Digital Libraries* vol. 6 no. 1, pp. 30-38, 2006.

[6] A. Buturovic, "MPEG-7 Color Structure Descriptor for visual information retrieval project VizIR", http://vizir.ima.tuwien.ac.at

[7] Chee Sun Won, Dong Kwon Park, "Efficient Use of MPEG-7 Edge Histogram Descriptor", *ETRI Journal*, Vol. 24, Number 1, 2002

[8] http://www.cs.waikato.ac.nz/~ml/weka/

[9] Zhang, Y., & Ji, Q. “Active and dynamic information fusion for multisensor systems with dynamic Bayesian networks,” *IEEE Trans. on Systems, Man, and Cybernetics*, Part B, pp.467-472.