

부가적인 감정 정보를 활용한 라이프로그의 이미지 선택 방법

범정현, 염상길, 황지영, 추현승

성균관대학교 소프트웨어대학

e-mail : {bumjh, sanggil12, jjwhang, choo}@skku.edu

Lifelog Image Selection using additional emotion information

Junghyun Bum, Sanggil Yeom, Joyce Jiyoung Whang, and Hyunseung Choo

College of Software

Sungkyunkwan University

요약

라이프로그는 개인의 일상생활에 대한 디지털 기록이다. 웨어러블 마켓의 성장으로 최근 라이프로그에 대한 연구가 활발해지고 있다. 라이프로그는 비구조화된 방대한 양의 데이터를 끊임없이 생성한다. 따라서 이를 요약하고 시각화하는 기법이 필요하다. 본 논문에서는 라이프로그 이미지들로부터 비주얼 다이어리를 자동으로 생성하기 위해 키프레임 이미지를 선택하는 새로운 방법을 제시한다. 이미지의 선명도와 같은 품질 측면외에 부가적인 감정정보 추출한 뒤 분류의 대상으로 포함시킨다. 감정 정보를 활용함으로써 시각화의 다양성을 높일 수 있음을 보여준다.

1. 서론

라이프로그는 개인의 일상생활에 대한 디지털 기록이며, 라이프로그은 인간의 삶을 디지털의 형태로 자동으로 기록하는 과정이다. 라이프로그(Lifelog)라는 용어는 1945년 7월 미국국가과학기술연구소 소장인 Vannevar Bush가 Atlantic Monthly 지에 “As We May Think”라는 기고문에서 처음 언급하였다[1]. 최근 센서 및 웨어러블 마켓의 성장, 스마트폰의 고사양화 및 다기능화 등으로 보다 쉽게 라이프로그를 수집할 수 있게 되면서 그 연구가 활발히 진행되고 있다.

웨어러블 카메라는 비주얼 라이프로그깅을 위해 필수적인 장비이다. 시장조사기관인 Tractica의 2016.2Q 보고서에 따르면 웨어러블 카메라 시장은 2015년 약 \$10 억에서 2021년 \$34 억까지 성장할 것으로 기대한다[2]. 웨어러블 카메라는 HTR(High Temporal Resolution)과 LTR(Low Temporal Resolution) 두 종류로 구분되어는데 상대적으로 짧은 순간을 고화질의 동영상을 기록하는 GoPro, Mecam, Looxcie 및 Google Glass 등의 HTR 장비와 긴 타임프레임 동안 사진 데이터를 획득하는 Narrative Clip, SenseCam과 같은 LTR 장비가 있다.

라이프로그의 사진 데이터는 하루에 2,000 개 이상의 이미지(연간 800,000 이미지)가 생성된다. 비구조화된 방대한 양의 데이터 스트림으로부터 사용자가 효율적으로 정확하게 연관된 콘텐츠를 검색하고 추출할 수 있는 방법이 필요하다. 즉, 정보를 추상화하고 요약할 수 있는 기법이 적용되어야 할 것이다.

본 논문에서는 라이프로그 이미지들로부터 비주얼 다이어리를 자동으로 생성하기 위해 키프레임 이미지를 선택하는 새로운 방법을 제안한다.

본 논문의 구성은 다음과 같다. 논문의 2장에서 관련 연구를 카테고리별로 소개하며 3장에서는 제안 시스템을 소개한다. 4장에서는 결론과 향후 연구에 대해서 서술한다.

2. 관련연구

라이프로그와 관련하여 유의미한 데이터 및 유용한 이미지 수집, 이미지 세그멘트화, 시각적 요약 및 콘텐츠 기반 검색 및 추출 등의 연구가 다양하게 수행되어 왔다. 기존의 연구는 특별히 고안된 웨어러블(Wearable) 카메라를 착용하고 의도하지 않은 형태의 라이프로그를 수집하고 해석하는 방법의 연구가 주로 수행되었으나, 최근에는 보다 적극적인 형태로 라이프로그가 일상을 HTR 카메라로 촬영하면서 이를 시각화하는 기술이 발전하고 있다. 또한 최근 스마트폰이 고성능, 다기능화되면서 스마트폰을 활용한 라이프로그 방안에 대한 연구도 활발해지고 있다.

2.1 웨어러블 카메라 기반 라이프로그 연구

Microsoft의 Gordon Bell[2]은 SenseCam이라는 목에 거는 형태의 소형 카메라와 녹음기를 이용한 장치를 개발하여 사용자가 이를 휴대하고 다니면서 경험정보를 수집하면 PC 기반의 MyLifeBits라는 응용 프로그램에서 사용자의 경험정보를 관리하는 방식의 연구를 수행하였다. Chowdhury, S. et al.[3]은 Authgrapher라는 라이프로그 장치로부터 수집된 데

이타에서 품질 문제를 해결하고 시공간적 다양성을 보여도록 주도록 요약하였다. 라이프로깅 데이터 추상화를 위해 시각적 측면에서 시간적-공간적 클러스터링 개념으로 확장한 것이다. Dobbins, C. et al.[4]은 클라우드 저장소, 기계 학습 기법을 적용하여 HDM(Human Digital Memory)를 생성하는 DigMem 시스템을 제안하였다. 키워드 기반의 검색이 아니라 콘텐트가 확률론적으로 얼마나 더 근접한지 보여주는 알고리즘을 사용하여 시맨틱-기반 검색 방법을 사용하였다.

2.2 모바일 데이터 기반 라이프로그 연구

Keum-Sung Hwang et al.[5]은 모바일로부터 수집한 GPS, 통화, 사진, MP3 등의 데이터를 통계적으로 분석하여 상황을 추론하고 GPS 정보를 이용하여 지도에 시각화하였다.

Ao Guo et al.[6]은 스마트폰 기반으로 개인의 데이터를 관리하고 분석하는 시스템을 제안하였다. 특히 감정을 분석하는 부분이 존재하며 분석방법은 심박수를 이용하는 것이다. Wu et al.[7]은 실제 생활에서 사용하는 모바일 디바이스로부터 센싱정보를 수집할 수 있는 모바일 센싱 플랫폼인 MobiSens를 제시하였다.

2.3 라이프로그 시각화

Lijuan Zhou et al.[8]은 모바일 기기로부터 수집한 GPS, WiFi, 가속센서, 사진 등의 라이프로그 데이터를 타임라인 형태로 시각화하였다. Hopfgartner, F., et al. [9]은 라이프로그 시스템을 개발하는데 있어서 가이드라인 템플릿을 제안한다. 비주얼라이제이션은 직관적이고 논리적이며 이해되기 쉬워야 하는데 1) overview를 먼저 보여주고 2) 중요 이벤트를 검색할 수 있는 방식으로 제공되어야 한다. Ai GOMI 등[10]은 장소, 시간, 사람을 기반으로 개인적인 사진을 계층적, 지리적, 시간적 형태로 시각화하였다. 클러스터링 알고리즘을 적용하여 사진들을 분류하고 다양한 뷰를 제공한다.

기존의 연구가 라이프로그를 시각화하는데 있어서 이미지 품질의 우수성과 다양성에 초점을 맞추어 사용자의 감정 요소를 고려하여 핵심 이미지를 선정하는 방안의 연구가 필요하다.

3. 제안 Approach

본 논문에서는 라이프로그 시각화를 위해 사용자의 감정 요소를 고려한 이미지 선정 기법을 소개한다. 먼저 노이지 라이프로그 이미지들(blurred image)을 제거하고 key moments를 생성하기 위해 적용된 클러스터링 기법을 소개한다.

3.1 노이지 이미지 제거

비의도적으로 랜덤하게 자동 생성된 라이프로그 이미지는 초점이 맞지 않거나 흐릿한 이미지의 연속인 경우가 발생한다. 첨도가 높은 고품질의 이미지를 선택하기 위해 Tong et al.[11]이 제안한 Haar wavelet

transform 을 사용한다. 에지 검출을 위해 원본 이미지를 Level-3 Haar Wavelet Transform 을 수행한다. 2x2, 4x4, 8x8 각 윈도우 사이즈에 에지 맵을 계산하고 최대값 (local maxima, $E_{max,i}$, $i=1,2,3$)을 찾는다. $E_{max,i}$ 는 에지의 강도를 나타내며 threshold(약 0.35) 보다 크면 에지 포인트이다. 먼저 Roof-Structure 와 Gstep-Structure 에지를 구한다. 만약 에지 포인트가 Roof-Structure 와 Gstep-Structure 이며 $E_{max,i}$ 이 threshold 보다 작은 값을 갖는다면 blurred image 라고 할 수 있다. Blur Score ($0 \leq B \leq 1$)는 Roof-Structure 와 Gstep-Structure 에지 포인트 중 blurred image 의 비율로 계산한다.

노이지 이미지를 제거하기 위해 $B > 0.3$ 보다 큰 이미지를 제거하고 실험 데이터 중 약 75% 정도의 정상 품질의 이미지를 포함시킨다.

3.2 이미지에 감정 정보 태깅

노이지 이미지가 제거된 이미지 셋에 대해 감정 값을 태깅한다. Project Oxford 는 마이크로소프트에서 개발한 감정분석 툴이며 사용자의 표정을 각각 개체의 위상차로 구분하여 감정을 추출한다. 이미지에서 사용자의 얼굴을 인식한 뒤 얼굴에서 눈, 코, 입 등의 객체를 분리한다. 이후 객체의 높낮이를 구분하여 사용자의 표정을 예측하고 Emotion Value ($0 \leq EV[i] \leq 1$, $i=0,2, \dots, 7$) 를 출력하게 되며 이 값을 사진 정보에 태깅한다. 만약 하나의 이미지에 두개 이상의 Emotion 이 검출된다면 다양성을 높일 수 있도록 각 항목 중 최대값을 선택한다.

3.3 클러스터링을 통한 이미지 선택

라이프로그의 Key 모멘트를 선정하기 위한 방법으로 유사한 이미지들로 이미지그룹을 만든 후 그룹 내에서 Blur Score 가 가장 낮은 값을 선택하는 방법을 적용한다. 이미지 분류하기 위해 Oliva et al.[12] 이 제안한 GIST feature 를 채택한다. GIST feature 는 이미지 파일의 특성을 유지하면서 스케일을 줄여 주어 클러스터링에 드는 소요시간을 줄여준다.

제안하는 알고리즘은 다음과 같다.

Step 1: 각 이미지를 512-D GIST features 로 변환한다.

Step 2: 이미지 Emotion Value 태그 정보($0 \leq EV[i] \leq 1$, $i=0,2, \dots, 7$)를 가져와 520-D vector 포맷을 생성한다.

Step 3: Expectation-maximization(EM) 알고리즘을 적용하여 클러스터링한다.

Step 4: 각 이미지 그룹(BIN)에서 가장 낮은 Blur Score 값을 갖는 이미지를 Keyframe 으로 선택한다.

4. 실험 및 결과

4.1 실험 데이터셋

제안한 이미지 선택 알고리즘의 유용성을 실험하기 위해 이미지 IMAGELAB 의 공개된 데이터셋 EGO_GROUP¹을 활용하였다. EGO_GORUP 은 outdoor, party, laboratory, coffee break 네가지 상황에서 획득한 약

¹ <http://imagelab.ing.unimore.it/files/EGO-GROUP.zip>

3,083 개의 이미지들로 구성되어 있다.

4.2 노이지 이미지 제거 단계

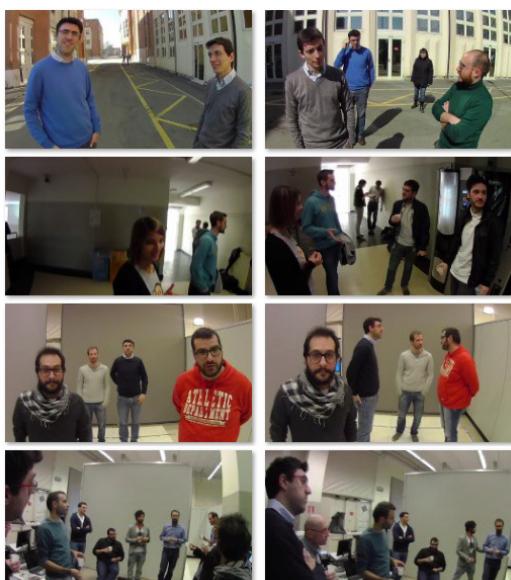
먼저 Haar wavelet transform 을 이용하여 Blur Score 를 계산한다. 여기서는 총 3,083 개의 이미지 중 Blur Score 가 0.3 미만인 이미지들만 2,042 개를 선택하였다. 라이프로그 이미지들은 초점이 맞지 않거나 선명하지 않은 중복된 이미지들이 다수 포함되어 threshold 를 0.35 이하로 설정한다.

4.3 감정정보 태깅 단계

감정정보 태깅단계에서는 마이크로소프트 Cognitive Services 인 Emotion API 을 적용하여 각 이미지들에서 감정정보를 추출한다. 이미지에서 얼굴이 차지하는 비중이 작거나 흐린 경우 안면을 인식하기 어렵다. 실험데이터셋에서 감정정보가 추출된 이미지는 약 52 % 였다.

4.4 이미지 선택 단계

다음은 이미지들을 512-D GIST features 로 변환하기 위해 GIST-global-Image-Descriptor Tool²을 활용한다. 생성된 512-D vector 는 앞서 추출한 감정정보와 결합하여 520-D vector 로 생성된다. 2042 x 512 vector 파일을 EM Clustering (클러스터수 8)로 수행한다. 8 개의 각 BIN 에 속한 이미지들중에서 Blur Score 가 가장 작은



(그림 1) 선정된 라이프로그 이미지

값을 갖는 이미지를 선택하였다. 선택된 이미지들은 그림 1 과 같다.

5. 결론 및 향후 연구

비주얼 라이프로깅은 장시간동안 일상생활을 캡처한 이미지들로 구성된다. 캡처된 사진들은 사람들이 어떻게 생활하는지와 관련된 상당한 Knowledge minig 기회를 제공한다. 따라서 헬스케어, 시큐리티, 레저

및 Quantified self 와 같은 새로운 응용분야로의 확장 가능성이 높다. 그러나 방대한 양의 이미지들은 사용자가 원하는 정보를 탐색하거나 핵심 정보만 요약하기 어렵게 만든다. 라이프로그 이미지들은 사용자의 목적에 따라 요약하는 방식이 달라질 수 있다. 치매와 같은 질병으로 보조적 기억수단으로 사용하는 경우 기억을 강화시키는 역할을 할 수도 있다. 따라서 다양한 시각화 기법이 연구되어야 할 것이다. 본 논문에서는 부가적인 감정 정보를 포함하여 클러스터링 함으로써 다양성을 증가시켰다. 향후 웨어러블 카메라를 통한 실험과 그에 따른 Key 이미지 분류 및 관련 시스템에 대한 연구를 진행할 것이다. 또한 선택된 이미지가 라이프로그를 얼마나 잘 대표하는지 User Study 를 진행할 필요가 있다.

ACKNOWLEDGEMENT

본 논문은 기초연구사업(NRF-2010-0020210)과 Grand ICT 연구센터지원사업 (IITP-2016-R6812-16-0001)의 연구결과로 수행되었음.

참고문헌

- [1] Bush, V. and A. W. M. Think, "As we may think," *The atlantic monthly*, 176(1): 101-108, 1945
- [2] Bell, C. G. and J. Gemmell, "Total recall: How the e-memory revolution will change everything," 2009
- [3] Chowdhury, S., P. J. McParlane, M. S. Ferdous and J. Jose, "My day in review: Visually summarising noisy lifelog data," ICMR, ACM, 2015.
- [4] Dobbins, C., Merabti, M., Fergus, P., & Llewellyn-Jones, D, "Creating human digital memories with the aid of pervasive mobile devices," *Pervasive and Mobile Computing* 12 160-178, 2014
- [5] Hwang, Keum-Sung, and Sung-Bae Cho. "A lifelog browser for visualization and search of mobile everyday-life," *Mobile Information Systems* 10.3: 243-258, 2014
- [6] Guo, Ao, and Jianhua Ma. "A Smartphone-based System for Personal Data Management and Personality Analysis," CIT/IUCC/DASC/PICOM, IEEE ,2015
- [7] Wu, Pang, Jiang Zhu, and Joy Ying Zhang. "Mobisens: A versatile mobile sensing platform for real-world applications," *Mobile Networks and Applications* 18.1: 60-80, 2013
- [8] Zhou, L. M., C. Gurrin, C. Yang and Z. Qiu, "From lifelog to diary: a timeline view for memory reminiscence," Irish HCI conference, 2013
- [9] Hopfgartner, F., Y. Yang, L. M. Zhou and C. Gurrin, "User interaction templates for the design of lifelogging systems," *Semantic Models for Adaptive Interactive Systems*, 187-204, 2013
- [10] Gomi, Ai, and Takayuki Itoh. "A personal photograph browser for life log analysis based on location, time, and person," ACM, 2011
- [11] Tong, H., M. Li, H. Zhang and C. Zhang, "Blur detection for digital images using wavelet transform," ICME'04, Vol. 1, IEEE, 2004
- [12] Oliva, Aude, and Antonio Torralba. "Building the gist of a scene: The role of global image features in recognition," *Progress in brain research* 155: 23-36, 2006

² <https://github.com/nrupatunga/GIST-global-Image-Descriptor>