

파트 기반의 보행자 재식별 연구

노영민⁰, 최진영

서울대학교 전기정보공학부

자동차시스템공동연구소(ASRI) 인지지능연구실
{treeoflife, jychoi}@snu.ac.kr

요 약

본 연구에서는 서로 다른 카메라에서 탐지된 같은 사람을 찾는 것을 의미하는 보행자 재식별이라는 문제를 다룬다. 이 문제는 서로 다른 탐지 결과로부터 같은 사람을 구별하는 것은 장애물, 조명, 포즈나 서로 다른 카메라 뷰로부터 오는 보행자 외관의 극단적인 변화로 인해 컴퓨터 비전의 난제 중 하나이다. 본 논문에서 다루는 아이디어는 인간의 인지 기능이 전체적인 외관 외에도 각 신체부위의 특징들을 세밀하게 비교하는 것에서 착안되었다. 이를 위한 기초 연구로써 사람의 각 신체부위에 따른 성능 변화를 확인하여 사람들의 여러 파트정보들을 보다 효율적으로 사용할 수 있는 방법을 찾는 것이 본 논문의 목표이다. 우선 각 파트만을 이용하여 보행자를 재식별할 수 있는지 여부와 어떤 파트가 가장 좋은 특징으로 사용될 수 있는지를 실험을 통해 나타내었다. 실험에 사용한 데이터셋은 CUHK03을 사용하였고 각 파트들은 pose estimation 알고리즘을 사용하여 추출하였다.

1. 서론

서로 다른 시점의 두 카메라의 두 이미지에서 탐지한 여러 사람에 대해서 같은 사람(같은 ID) 인지를 추론하는 보행자 재식별(Person Re-identification) 문제는 컴퓨터 비전의 대표적 문제 중 하나이다. 재식별 문제는 다중 카메라 추적, 군중 밀도 감지, 다중 물체 추적 등의 감시환경(surveillance system)의 응용분야에서 사용되고 있다[1-3]. 최근에 들어서는 차량의 재식별(Vehicle ReID)[4] 등의 분야에서도 연구가 진행되고 있다.

보행자 재식별은 장애물, 조명, 포즈 등의 카메라 뷰의 차이에서 생기는 보행자 외관(appearance)의 극단적인 변화로 인해 그 성능을 올리는 데에 큰 어려움이 있다. 이러한 문제를 해결하기 위하여 보행자의 다양한 포즈와 조명 변화에 강인한 대표 특징(representative feature)을 추출하는 연구[5-7]가 진행되어 왔다. 또한 데이터셋 안에서 각 보행자의 ID를 클래스라고 할 때, 같은 클래스와 다른 클래스간의 차이를 멀어지도록 하는 distance metric을 정의하여 학습하는 방법의 연구들[8-10]이 있다.



그림 1. 파트를 사용하여 ID를 구별하는 예시

최근 컴퓨터 비전 분야에서는 딥러닝을 이용하여 representative feature를 추출하고 이를 사용해 각 객체들을 분류한 뒤 추론을 통해 의미있는 결과를 내는 연구가 진행되고 있다. ResNet[11](residual network)은 이미지 분류 분야에서 큰 획을 이룬 CNN(convolutional neural network) 으로서 깊은 layer를 쌓는데 성공하였고 여러 가지 challenge에서 우승한 network 구조이다. 보행자 재식별의 분야에서도 ResNet(50-layer)과 L2 loss만 사용하여 metric을 학습하여도 state of the art의 준하는 성능을 낼 수 있다.

본 연구에서는 일반적인 사람이 보행자를 식별하는 방법에 영감을 받았다. 일반적으로 서로 다른 사람들을 분별하는 것은 전체적인 모양이나 색깔로 구분할 수 있지만 비슷한 옷차림의 사람들을 구분하는 것이라면 더 이상 전체적인 모양을 보는 것이 큰 의미가 없다. 오히려 작은 차이점들을 통하여 분류하게 될 것이다. 예를 들면 그림 1-(a)와 같이 같은 흰색 상의 이지만 반팔인지 아닌지를 통해서, 그리고 그림 1-(b)와 같이 손목시계를 착용했는지 여부를 통해서 쉽게 다른 사람임을 구별할 수 있다.

보행자의 각 파트들(신체부위: 팔, 다리, 머리, 몸통 등)을 사용하는 연구들[12-14]이 최근에 활발하게 진행되고 있다. 하지만 여전히 파트정보를 적절하게 활용하지 못하거나 그 성능 향상의 정도도 미미한 수준이다. 본 연구에서는 보행자의 파트정보를 활용하는 연구의 기초연구로써, 각 파트만을 가지고도 각 사람을 구별하는 것이 가능한지 여부와 어떤 파트가 더 좋은 구별 성능을 가지는지에 대한 실험과 분석을 진행하였다.

2. 파트의 정의와 학습방법

2.1 파트의 정의

최근 각 사람의 신체부위를 인식하여 서로 연결하는 자세 추정 분야에서 눈에 띄는 성능결과의 발전이 이루어지고 있다[15-17], 본 연구에서 사용된 각 파트들은 자세 추정 알고리즘[16]의 결과를 기반으로 하여 추출하였다.

본 논문에서는 전체 사람 객체를 그림 2와 같이 10개의 파트로 나누었다. 각 파트의 정의는 표 1에 표시 하였다.

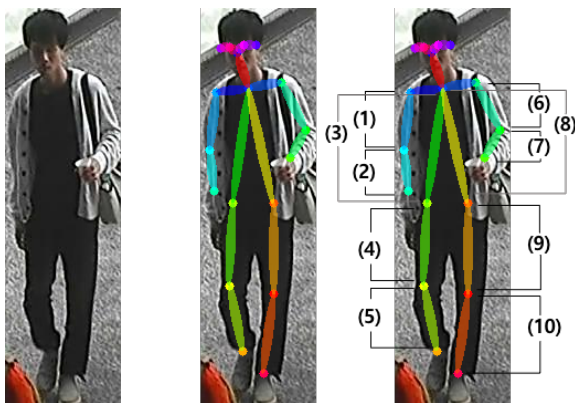


그림 2. pose estimation 결과와 파트의 정의

표 1. 각 파트의 정의

Parts	Form	To
(1)/(6)	right/left shoulder	right/left elbow
(2)/(7)	right/left elbow	right/left wrist
(3)/(8)	neck	right/left hip
(4)/(9)	right/left hip	right/left knee
(5)/(10)	right/left knee	right/left ankle

2.2 Architecture 와 Loss function

본 연구에서 특징 추출기(feature extractor)는 ResNet[11](18-layer)을 사용하였다. 각 파트의 이미지는 전체 보행자의 이미지에 비하여 약 1/20 이하 정도의 크기이므로 ResNet중에 가장 작은 18-layer를 사용하였다. 각 residual block은 CONV-BN-RELU-CONV-BN-RELU 순서로 되어있다 (CONV[convolution layer], BN[batch normalization], ReLU[rectified linear unit]).

Loss function은 수식 (1)과 같이 Triplet loss를 사용하였다.

$$L_{tri} = \sum_{a, p, n} [m + D_{a,p} - D_{a,n}]_+ \quad (1)$$

$D_{a,p}$ 는 기준이 되는 anchor 이미지와 같은 ID 사이의 distance이고 $D_{a,n}$ 는 기준 anchor 이미지와 다른 ID 사이의 distance이다. m 은 margin threshold 값으로 $D_{a,p}$ 과 $D_{a,n}$ 의 차이가 m 이상이 될 때만 loss가 발생한다. $[\]_+$ 는 $\max([\], 0)$ 을 의미한다.

2.3 Dataset

실험에 사용한 데이터셋은 CUHK03이다. CUHK03은 1360개의 ID에 대하여 13164장의 이미지로 구성되어있다. 일반적으로 200개의 ID의 이미지를 100개의 ID는 validation set로 구성하고 나머지 100개의 ID에 해당하는 이미지는 test set으로 사용한다. 본 연구는 기초 연구로서 validation set에서의 결과를 명시하였다. (물론, validation data는 training data와 겹치지 않는다.)

성능 측정은 재식별에 자주 사용되는 CMC(cum

ulative Matching Characteristics) curve를 사용하였다[18], 수치가 높을수록 더 좋은 성능이다.

3. 결과 및 분석

파트별 보행자 재식별 결과는 표 2와 같다. 제 6 파트(왼쪽 어깨부터 팔꿈치)에서 top-1 score 47.84%로 가장 좋은 성능을 얻을 수 있었다. top-10 score는 제 10 파트(왼쪽 무릎부터 발꿈치)에서 84.99%로 가장 높았다. 왼쪽에 위치한 파트들에서 분류 성능이 더 좋은 것을 확인 할 수 있었고 이는 데이터셋의 특징인 것으로 보인다. 제 6 파트의 성능은 기존에 보행자 재식별에서 딥러닝을 적용하지 않았을 때의 몸 전체 이미지를 특징으로 한 알고리즘[19]과 비슷한 수준인 것으로 나타났다.

4. 결론

본 연구에서는 일반적인 사람의 보행자 재식별 과정에 영감을 받은 연구로서 사람의 파트(신체부위)를 기반으로 하여 보행자 재식별 실험을 진행하였다. 특정 한 파트만 가지고도 보행자 재식별이 가능하다는 것을 실험을 통해 밝혔고 이는 딥러닝이 적용되지 않고 몸 전체를 가지고 보행자 재식별을 했을 때와 top-1성능은 비슷하거나 top-5나 top-10에서는 오히려 높게 나타났다. CUHK03 데이터 셋에서 가장 좋은 구별 성능을 내는 파트는 제 6파트(왼쪽 어깨부터 팔꿈치)이다.

표 2. 파트별 보행자 재식별 결과 (단위:%)

(Best: **bold**, Second: underlined)

part	top-1	top-5	top-10
1	22.15	49.27	63.59
2	26.64	59.72	74.06
3	27.26	46.18	61.27
4	29.49	61.14	76.16
5	25.32	66.42	<u>83.62</u>
6	47.84	71.84	81.91
7	31.27	60.72	73.28
8	23.79	53.57	70.27
9	31.12	63.20	80.89
10	34.01	<u>69.46</u>	84.99
kLFDA[19]	48.20	<u>59.34</u>	<u>66.38</u>

감사의글

이 논문은 2017년도 미래창조과학부 및 정보통신기술진흥센터의 정보통신, 방송연구개발사업[2014-0-00059, 예지형 시각 지능 원천 기술 개발]과 Brain Korea 21 플러스 사업의 지원을 받아 수행된 연구이다.

참고문헌

- [1] J. Berclaz, F. Fleuret, and P. Fua. Multi-camera tracking and atypical motion detection with behavioral maps. In ECCV, 2008
- [2] A. Chan and N. Vasconcelos. Bayesian poisson regression for crowd counting. In ICCV. 2009
- [3] X. Wang, K. Tieu, and W. Grimson. Correspondence-free multi-camera activity analysis and scene modeling. In CVPR, 2008.
- [4] Z. Wang, L. Tang, X. Liu, Z. Yao, S. Yi, J. Shao, and X. Wang. Orientation Invariant Feature Embedding and Spatial Temporal Regularization for Vehicle Re-identification. In CVPR 2017
- [5] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In CVPR, 2010.
- [6] B. Ma, Y. Su, and F. Jurie. Bicov: a novel image representation for person re-identification and face verification. In BMVC, 2012.
- [7] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian. Query-adaptive late fusion for image search and person re-identification. In CVPR, 2015.
- [8] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In CVPR, 2016.
- [9] C. Liu, C. C. Loy, S. Gong, and G. Wang. Pop: Person re-identification post-rank optimisation. In ICCV, 2013.
- [10] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, and Y. Tian. Unsupervised cross-dataset transfer learning for person re-identification. In CVPR, 2016.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep

- p residual learning for image recognition," arXiv:1512.03385, 2015.
- [12] L. Zhao, X. Li, J. Wang, and Y. Zhuang. Deeply-learned part-aligned representations for person re-identification. In ICCV, 2017.
 - [13] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian. Pose-driven deep convolutional model for person re-identification. In ICCV, 2017.
 - [14] Z. Haiyu, T. Maoqing, S. Jing, S. Shuyang, Y. Junjie, Y. Shuai, W. Xiaogang, and T. Xiaohou. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In CVPR, 2017.
 - [15] S. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In CVPR, 2016.
 - [16] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In CVPR, 2017.
 - [17] X. Chen and A. Yuille. Articulated pose estimation by a graphical model with image dependent pairwise relations. NIPS, 2014.
 - [18] L. Zheng, Y. Yang, and A. G. Hauptmann. Person re-identification: Past, present and future.
 - [19] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In ECCV, 2014.