Person Identification with Convolutional Neural Networks

Kang Zheng

Person identification aims at matching persons across images or videos captured by different cameras, without requiring the presence of persons' faces. It is an important problem in computer vision community, and has many important real-world applications, such as person search, security surveillance and no-checkout stores. However, this problem is very challenging due to various factors, such as illumination variation, view changes, human pose deformation, and occlusion. Traditional approaches generally focus on hand-crafting features and/or learning distance metrics for matching to tackle these challenges. With Convolutional Neural Networks (CNNs), feature extraction and metric learning can be combined in a unified framework.

In this work, we study two important sub-problems of person identification: cross-view person identification and visible-thermal person re-identification. Cross-view person identification aims to match persons from temporally synchronized videos taken by wearable cameras. Visible-thermal person re-identification aims to match persons between images taken by visible cameras under normal illumination condition and thermal cameras under poor illumination condition such as during night time.

For cross-view person identification, we focus on addressing the challenge of view changes between cameras. Since the videos are taken by wearable cameras, the underlying 3D motion pattern of the same person should be consistent and thus can be used for effective matching. In light of this, we propose to extract view-invariant motion features to match persons. Specifically, we propose a CNN-based triplet network to learn view-invariant features by establishing correspondences between 3D human MoCap data and the projected 2D optical flow data. After training, the triplet network is used to extract view-invariant features from 2D optical flows of videos for matching persons. We collect three datasets for evaluation. The experimental results demonstrate the effectiveness of this method.

For visible-thermal person re-identification, we focus on the challenge of domain discrepancy between visible images and thermal images. We propose to address this issue at a class level with a CNN-based two-stream network. Specifically, our idea is to learn a center for features of each person in each domain (visible and thermal domains), using a new relaxed center loss. Instead of imposing constraints between pairs of samples, we enforce the centers of the same person in visible and thermal domains to be close, and the centers of different persons to be distant. We also enforce the feature vector from the center of one person to another in visible feature space to be similar to that in thermal feature space. Using this network, we can learn domain-independent features for visible-thermal person re-identification. Experiments on two public datasets demonstrate the effectiveness of this method.