Joint Flow: Temporal Flow Fields for Multi Person Pose Tracking

Andreas Doering, Umar Iqbal, Juergen Gall
Computer Vision Group, University of Bonn

Motivation:
- Problem: Multi-Person pose estimation and tracking in videos
- Contributions:
  - Task-specific novel representation (Temporal Flow Fields) for person association
  - Propose an online tracking approach that operates on two (consecutive) frames

Introduction

1. Pose Estimation
2. Pose Association

Given sequence of frames

Temporal Flow Fields

- Aggregated Temporal Flow Field
  - Number of unit vectors at pixel
  - \( p_i \)
  - \( A(p_i) \)
  - \( p = (x, y) \)

Generation of Temporal Flow Fields

- Given a set of joints of joint class \( j \) of person \( k \) in frames \( I_{t-1} \) and \( I_t \):
  - Define a temporal edge of length \( \lambda_j = \sqrt{(x_{j,t-1} - x_{j,t})^2 + (y_{j,t-1} - y_{j,t})^2} \)
  - Restrict temporal edge to pixels close to the joint motion by parameter \( \sigma \)
  - Set of pixels along temporal edge:
    - Temporal Flow Fields for each joint class \( j \) of person \( k \):
      - \( \psi_{j,k}^t(p) = \frac{1}{\sum_{A(p_i)}} \sum_{A(p_i)} p_i \)
      - \( p = (x, y) \)

CNN Architecture

- a) Siamese Pose Estimation Network (1):
  - Trained on COCO + PoseTrack
  - 12 instead of 10 VGG layers
  - Improved pose configuration
  - 8 instead of 10 stages
- b) Temporal Model:
  - Trained on FlowTrack
  - Takes spatial features

Inference Overview

- Spatial Inference (1):
  - Bottom-up pose estimation approach:
    - Generate feature maps for neck and hip
  - Temporal Model:
    - Predict Temporal Flow Field

Optical Flow:
- Alternative formulation for temporal potentials:
  - \( \mathbf{f}^{p}_{j,k} = \mathbf{f}^{p}_{j,t-1} - \mathbf{f}^{p}_{j,t} \)

Evaluation

Different Edge Configurations

Comparison to Baselines

Comparison to State-of-the-art

References

[3] B. Xiao, H. Wu and Y. Wei, Simple Baselines for Human Pose Estimation and Tracking, ECCV 2018

doering@iai.uni-bonn.de