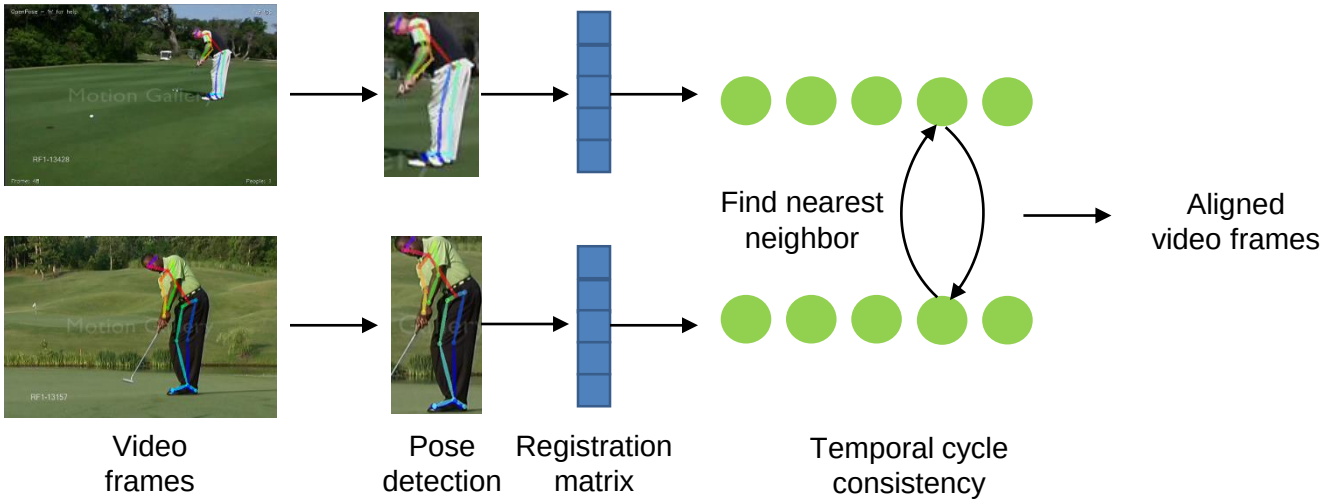
Person Video Alignment with Human Pose Registration

Yu ZHANG, Zihua WANG, Jingyang ZHOU, Siya MI

Frontiers of Computer Science, DOI: 10.1007/s11704-022-1347-7

Problems & Ideas

- Problems of conventional video alignment approaches:
  - difficulty in distinguishing from similar frames for alignment without context information.
  - person videos contain abundant pose information which has not been taken full advantage of.
- Ideas: Use pose registration to combine temporal and spatial information to dig the correspondences of two videos.



First, we detect the person bounding box and the joint points. Then we calculate the registration matrix and feature similarity of the standard video frames. By comparing the similarity between frames, the corresponding frames are determined so that two videos can be aligned without fine-grained annotations.

Main Contributions

- Contributions:
  - We introduce a method that uses both temporal and spatial information to dig the correspondences.
  - To solve the problem of disordered alignment, we introduce a sequential-controlled loss to improve alignment efficiency.
  - We use the feature points of human pose for pose registration, where the rotation and deformation information will provide characteristics of human actions in different phases.

Method	Dataset	Kendall's Tau
Supervised Learning	Penn	0.558
TCC	Penn	0.735
Pose-only	Penn	0.110
Without-registration	Penn	0.722
Our method	Penn	**0.740**
Supervised Learning	UCF101	0.141
TCC	UCF101	0.235
Without-registration	UCF101	0.246
Our method	UCF101	**0.251**
Without-registration	HMDB51	0.191
Our method	HMDB51	**0.193**

Video alignment performance on Penn action, UCF101, and HMDB51. Kendall's Tau is used to measure the performance of the alignment.