

# SUPPLEMENTARY MATERIAL: A SPATIO-TEMPORAL APPEARANCE REPRESENTATION FOR VIDEO-BASED PEDESTRIAN RE-IDENTIFICATION

Kan Liu<sup>†</sup>    Bingpeng Ma<sup>‡</sup>    Wei Zhang<sup>†</sup>    Rui Huang<sup>\*\*</sup>

<sup>†</sup>School of Control Science and Engineering, Shandong University, China

<sup>‡</sup>School of Computer and Control Engineering, University of Chinese Academy of Sciences, China

<sup>\*</sup>NEC Laboratories, China

sakuraxiafan@gmail.com, bpma@ucas.ac.cn, davidzhangsdu@gmail.com, huang-rui@nec.cn

We show some additional experimental results here to demonstrate the effectiveness of our spatio-temporal representation. In particular, six figures are shown. Each figure contains two approaches, (a) the FV3D approach and (b) the STFV3D approach. For FV3D, fixed-length (20 frames) fragments are segmented around the local maxima/minima of the FEP and a  $5 \times 2 \times 2$  regular grid is used to divide the video fragments into smaller cells to take into account the spatial and temporal alignment in a standard way. For STFV3D, walking cycles with variable lengths are extracted and  $6 \times 4$  body-action units are used for each walking cycle. For each approach, a pair of subsequences that is the best matched pair of subsequences from the query (top) and the gallery (bottom) using the Nearest Neighbor classifier are shown. The red boxes indicate the better match between the two approaches.

Figure 1 ~ Figure 4 are four examples in which the STFV3D approach finds the correct match while FV3D does not. In Figure 1, FV3D did not extract proper walking cycles, therefore it found a pair of subsequences from the query and gallery that are close both in appearance and in pace. STFV3D extracted walking cycles and found a pair of subsequences that are most similar in appearance, even though the lengths of the two subsequences are different. Similar observation can be made on the other three figures.

Figure 5 and Figure 6 are two examples in which FV3D finds the correct match while STFV3D does not. We consider these the failure cases of our spatio-temporal representation. As mentioned in the paper, the color difference between the two cameras causes the most trouble. But since we are comparing FV3D and STFV3D here, both of which are affected by the color difference, Figure 5 and Figure 6 mostly show the other important factors that may cause problems to our representation. First, the cluttered background causes inaccurate walking cycle extraction (Figure 5(b)); second, the viewpoint of the sequence is significantly different from our *sideview* assumption, which also causes inaccuracy of both spatial and temporal alignment(Figure 6(b)).

---

\*Corresponding author. This work was partially supported by NSFC 61203253, 61572465, 61573222, Shandong Province RF 2015ZDXX0801A02, BS2013DX023, and SRF for ROCS, SEM.

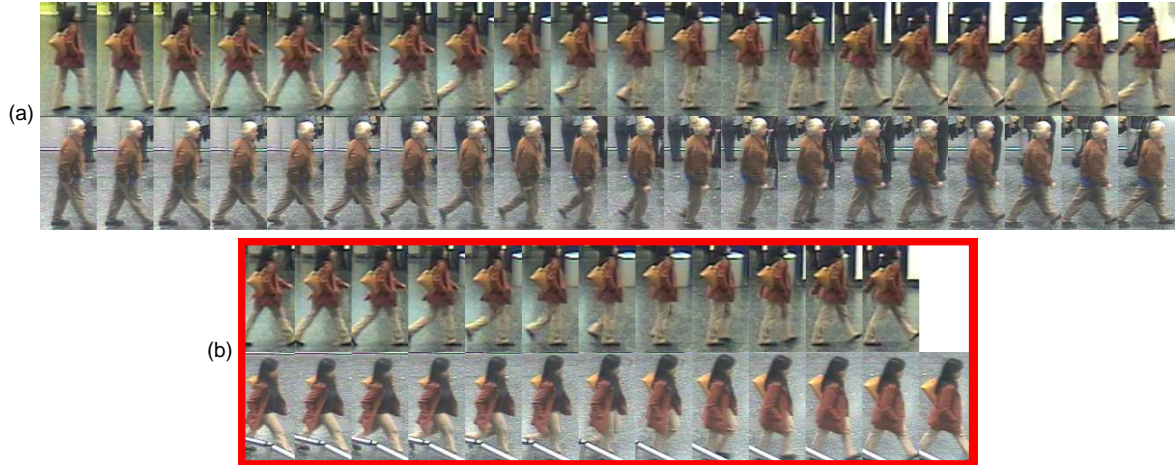


Figure 1. The example pairs of the image subsequences of the top-1 matching results. (a) the FV3D approach (top: query, bottom: gallery); (b) the STFV3D approach (top: query, bottom: gallery). Red box indicates the correct match.



Figure 2. The example pairs of the image subsequences of the top-1 matching results. (a) the FV3D approach (top: query, bottom: gallery); (b) the STFV3D approach (top: query, bottom: gallery). Red box indicates the correct match.



Figure 3. The example pairs of the image subsequences of the top-1 matching results. (a) the FV3D approach (top: query, bottom: gallery); (b) the STFV3D approach (top: query, bottom: gallery). Red box indicates the correct match.

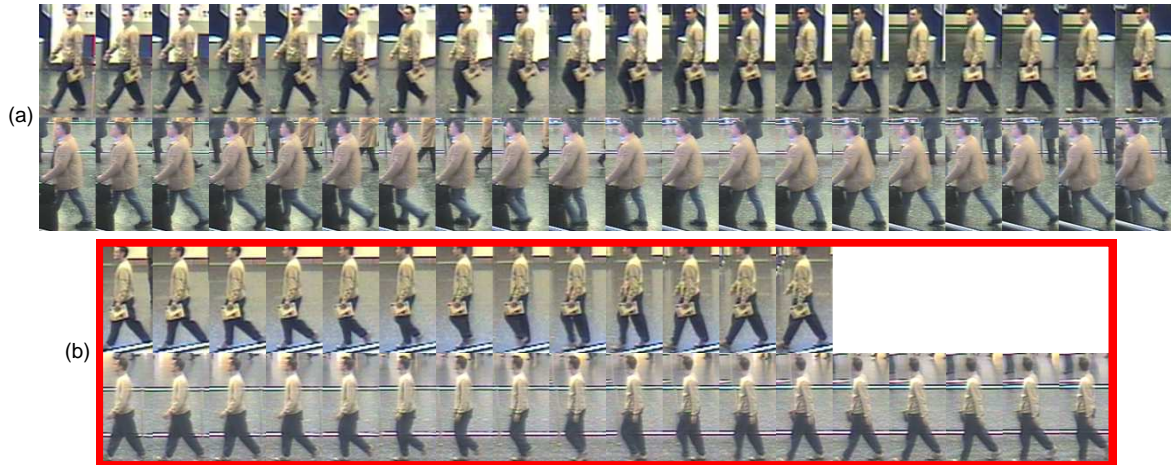


Figure 4. The example pairs of the image subsequences of the top-1 matching results. (a) the FV3D approach (top: query, bottom: gallery); (b) the STFV3D approach (top: query, bottom: gallery). Red box indicates the correct match.

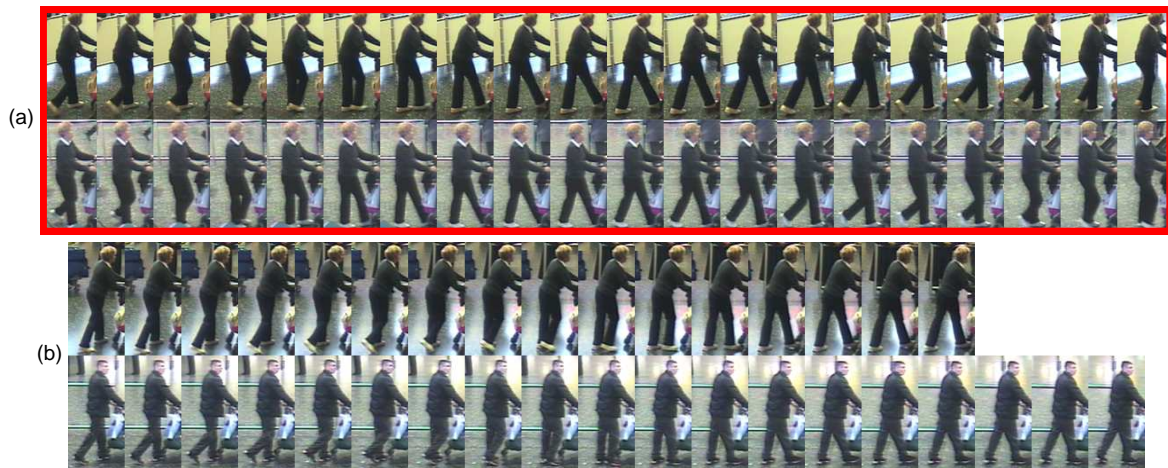


Figure 5. The example pairs of the image subsequences of the top-1 matching results. (a) the FV3D approach (top: query, bottom: gallery); (b) the STFV3D approach (top: query, bottom: gallery). Red box indicates the correct match.

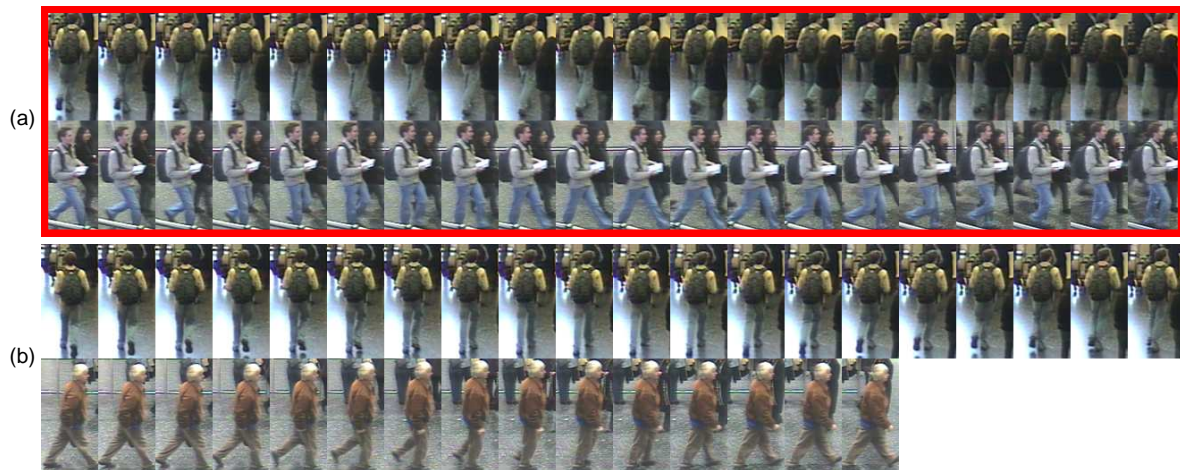


Figure 6. The example pairs of the image subsequences of the top-1 matching results. (a) the FV3D approach (top: query, bottom: gallery); (b) the STFV3D approach (top: query, bottom: gallery). Red box indicates the correct match.