ICIP 2017, BeiJing, China

Human-human interaction recognition based on spatial and motion trend feature (Paper ID #2688)

Bangli Liu, Haibin Cai, Xiaofei Ji, and Honghai Liu

Email: bangli.liu@port.ac.uk, honghai.liu@port.ac.uk

University of Portsmouth

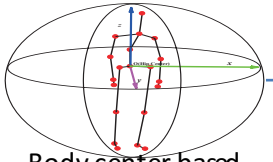
1. Abstract

Human-human interaction recognition has attracted increasing attention in recent years due to its wide applications in computer vision fields. However, it still remains a challenge due to mutual occlusion, various subject appearance or body size and complex context. In this paper, a novel feature descriptor based on spatial relationship and semantic motion trend similarity between body parts is proposed for human-human interaction recognition.

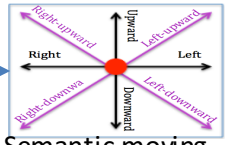
2. Feature Extraction

The moving direction of each joint is firstly quantified into several semantic words and then the motion trend similarity is captured by histogram intersection.

2.1 Motion trend feature



Body center based coordinate system



Semantic moving directions $\mathbf{V}_j$ in 3D space

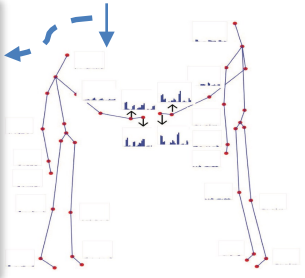
- Frame-direction:

$$\mathbf{v}_t^i = \{x_{p_t^i} - x_{p_{t-1}^i}, y_{p_t^i} - y_{p_{t-1}^i}, z_{p_t^i} - z_{p_{t-1}^i}\}$$

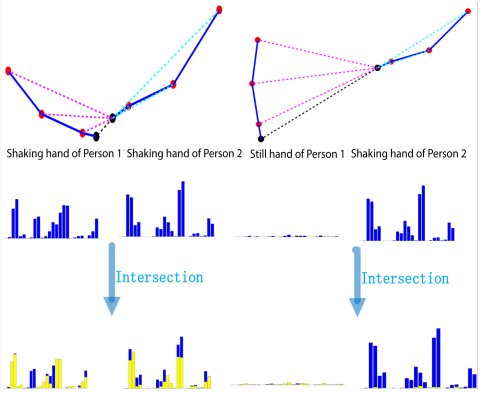
- Soft direction encoding depending on cosine similarity and displacement:

$$cos\theta_j^i(t) = \frac{\mathbf{v_j} \cdot \mathbf{v_t^i}}{\|\mathbf{v_t^i}\|\|\mathbf{v_j}\|}, j \in [1, m]$$

$$\begin{cases} bin_{first} = bin_{first} + Dis^i(t) \times cos\theta^i_{first}(t) \\ bin_{second} = bin_{second} + Dis^i(t) \times cos\theta^i_{second}(t) \end{cases}$$



2.2 Motion trend similarity



$$InterSectionJ\left(b_i^k, b_j^k\right) = min\left(|b_i^m|, |b_j^m|\right)$$

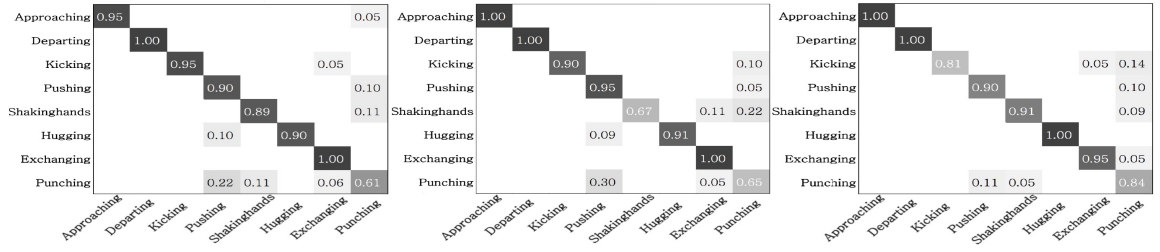
$$H = \left(\underbrace{1,...,1}_{H_1},\underbrace{0,...,0}_{N-H_1},\underbrace{1,...,1}_{H_2},\underbrace{0,...,0}_{N-H_2},...,\underbrace{1,...,1}_{H_m},\underbrace{0,...,0}_{N-H_m}\right)$$

$$K_{type} = \sum_{p=1}^{8}\sum_{q=1}^{8}\sum_{k=1}^{26} InterSectionJ\left(b_p^k, b_q^k\right) = \sum_{p=1}^{8}\sum_{q=1}^{8} K\left(H\left(p\right), H\left(q\right)\right)$$

Where,

$$K\left(H\left(i\right), H\left(j\right)\right) = \sum_{k=1}^{m} InterSectionJ\left(b_i^k, b_j^k\right) = H\left(i\right) \cdot H\left(j\right)$$

3. Experiment Results



(a) *Test One* (b) *Test Two* (c) *Cross Subjects Test*

Fig. 1 Confusion matrixes

Table 1: Recognition Accuracy (%) on SBU dataset.

State-of-the-art	Joint features [13]	80.3
	CFDM [9]	89.4
	[18]	86.9
Proposed Method	*Test One*	90.58
	Test Two	90.28
	Cross Subjects Test	92.50
	Average	**91.12**

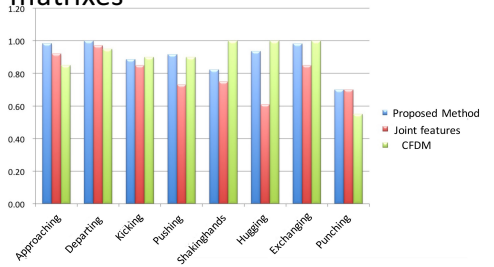


Fig. 2 Comparison between categories