

Panoramic Human Activity Recognition

Supplementary Material

Ruize Han¹, Haomin Yan^{1*}, Jiacheng Li^{1*},
Songmiao Wang¹, Wei Feng^{1†}, and Song Wang^{2†}

¹ Intelligence and Computing College, Tianjin University, China

² University of South Carolina, Columbia, USA

{han_ruize, yan_hm, ljcthreeswords, smwang, wfeng}@tju.edu.cn,
songwang@cec.sc.edu

1 Dataset Annotation

We first discuss the problem setting and dataset annotation of this work and previous works, e.g., JRDB-Act [2]. In our PAR problem setting, the group activity is *related to* the individual actions but ***their label candidate sets are different***. For example, as shown in the below figure (a), the group activity in [2] is labeled as ‘listening to someone, talking to someone’, which is *the combination of the contained individual action labels*. In our setting, we label such group activity as ‘chatting’, which *is regarding to the whole group but not each individual*. Another issue is that some redundant or unbecfitting labels appear in the group activity in [2], e.g., the label ‘sitting’ as shown in (b,c). Therefore, we relabel the group activity by including some new group activity labels different from the individual labels.



For the global activity in this work, it actually depicts the ***general and abstract activity*** according to ***majority of the people***. This *follows the setting in the classical group activity recognition (GAR) problem*, e.g., CAD dataset [1], that provides an overall activity of all people in the scene. For example, as shown in the above figure, a global activity of ‘commuting’ is composed of individual/group actions of walking, chatting, a global activity of ‘office working’ contains the human sitting, looking at the computer, etc. To obtain such annotations, we first scan all the videos and discuss repeatedly to confirm the activity label sets, and then arrange three graduate students for independent annotation and cross validation to avoid subjectivity.

Overall, the proposed PAR problem studies the human activity understanding in *multiple spatial granularity* (individual, group and global) with *different levels of activity categories*. This work and JRDB-Act are both the extensions for activity analysis based on the JRDB dataset but in different directions, which, we believe, can complement each other.



References

1. Choi W, Shahid K, S.S.: What are they doing?: Collective activity classification using spatio-temporal relationship among people. In: ICCV (2009)
2. Ehsanpour, M., Saleh, F.S., Savarese, S., Reid, I.D., Rezatofghi, H.: JrdB-act: A large-scale dataset for spatio-temporal action, social group and activity detection. In: arXiv preprint (2021)