

Recognition of Transitional Action for Short-Term Action Prediction using Discriminative Temporal CNN Feature

Hirokatsu Kataoka¹

hirokatsu.kataoka@aist.go.jp

Yudai Miyashita²

undermusic@gmail.com

Masaki Hayashi^{3,4}

m.hayashi@liquidinc.asia

Kenji Iwata¹

kenji.iwata@aist.go.jp

Yutaka Satoh¹

yu.satou@aist.go.jp

¹ National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Ibaraki, Japan

² Tokyo Denki University
Adachi-ku, Tokyo, Japan

³ Liquid Inc.
Tokyo, Japan

⁴ Keio University
Yokohama, Kanagawa, Japan

Transitional actions belong to a class between actions for short-term action prediction (see Figure 1). Early action recognition is necessary for producing action predictions in the early frames of an objective action. Earlier prediction in the initial frames of an objective action is desirable for early action recognition problems, but the solutions depend on the action itself. On one hand, within the setting of a short-term action prediction, understanding a pending human action change is more natural if we have a firm grasp on transitional actions. In a traffic scene, short-term action predictions are particularly crucial for avoiding accidents between humans and vehicles. Figure 1 shows sequential actions that include *Walk straight*, *Walk straight - cross*, and *cross*. Where *Walk straight* and *cross* are conventional action definitions, our proposal adds a transitional action between actions (here *Walk straight - cross*) in order to provide a better action approach to predictions. Our proposed short-term predictions achieve earlier prediction than so-called early activity recognition, since they can recognize a dangerous *cross* action while it is transitional. Intuitively, the recognition difficulty arising from action and transitional action is that they tend to partially overlap each other. We believe that the use of a subtle motion descriptor (SMD) will allow us to identify sensitive differences between actions and transitional actions.

In this paper, we address the recognition of transitional action for short-term action prediction. We also propose a discriminative temporal convolutional neural network (CNN) feature that can be used for recognizing transitional actions in order to overcome the difficulty of indistinguishable feature classification in transitional

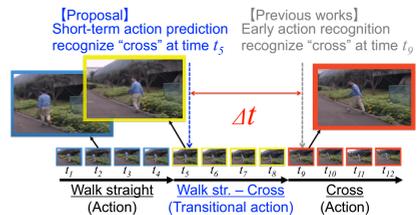


Figure 1: Recognition of transitional actions for short-term action prediction: Identification of transitional actions allow us to understand the next activity at time t_5 before an early action recognition approach at time t_9 .

actions. To accomplish this, we employ an SMD that captures subtle differences between consecutive frames. Our paper contains two main contributions: (i) the definition of transitional action for short-term action prediction that achieves earlier prediction than early action recognition, and (ii) identifying CNN-based SMD to create a clear distinctions between action and transitional action. The feature is simply updated from a spatio-temporal CNN feature Pooled Time Series (PoT) proposed in [1].

Our CNN-based SMD demonstrated the best rate of success on three different trial datasets. Even when using the shortest (3-frame) feature accumulation for recognition tuning, we confirmed outstanding results with 85.78% (NTSEL), 69.77% (UTKinect), and 49.93% (Watch-n-Patch) on the three different datasets.

[1] M. S. Ryoo, B. Rothrock, and L. Matthies. Pooled motion features for first-person videos. CVPR, 2015.