

報告番号	※甲	第	号
------	----	---	---

主 論 文 の 要 旨

論文題目 Visual Analysis of Human Activity with Middle-high Level Knowledge (ミドル-ハイレベル知識を用いた人間行動の映像解析)
氏 名 WANG Yu

論 文 内 容 の 要 旨

In this study, we seek approaches for practical human activity analysis tasks. Our core idea is to encode the middle-high level knowledge, which human relies on when understanding visual information, into computational activity analysis approaches. Based on such an idea, we deal with the following three challenging practical human activity analysis tasks.

- Human detection in traffic videos.
- Action recognition under a varying environment.
- Summarization of surveillance videos based on activity recognition.

For the first task, we propose an approach which reflects high level human knowledge on the relationship between different types of information. The proposed approach utilizes the sparse depth information as a supplementary cue to the appearance-based detection mechanism, and updates the outputs from an image-feature-only human detector probabilistically using a simplified graphical model. The final detection outputs are contributed by both image features and depth information, and could provide stronger ability to discriminate human from background and other objects. Through the experiment on challenging real world traffic videos, we confirmed that the proposed approach could improve the image-feature-only human detector significantly, with minor additional processing time.

For the second task, we propose a novel descriptor, known as bundled pose descriptor (BP descriptor), to represent actions. BP descriptor is designed to

reflect the middle level human knowledge on feature encoding. Different to popular local spatio-temporal descriptors which only capture low level changes in videos, BP descriptor explicitly captures the information of an actor's pose and pose order, which are thought to be extremely important when human recognizes actions. Such a property makes it a stronger discriminative ability even under a varying environment. We evaluated the BP descriptor on the KTH dataset, and confirmed that it can not only provide promising recognition accuracy when work alone, but also significantly boost up the performance of many other popular local features.

For the last task, we propose a novel approach for activity recognition in nursery school surveillance videos. This approach makes full use of a learned distance metric and reflects the middle level human knowledge on recognition mechanisms. In the proposed approach, the learned distance metric is combined with supervised classification and unsupervised clustering, to categorize daily raw surveillance videos into individual event categories in a coarse-to-fine fashion. Its practical advantage is that it requires only limited labelling efforts, which means it is easy to be implemented. Additionally, we also confirmed that it has a reliable performance on activity recognition for uncontrolled surveillance videos.

Through this study, we encoded the middle-high level knowledge, which plays important roles in the recognition process of human, into computational approaches for practical human activity analysis tasks. By evaluating these approaches on challenging real world data, our belief in the power of middle-high level knowledge has been plentifully verified. We confirmed that properly utilizing such knowledge in practical human activity analysis tasks can result in solutions with better performance and easier implementation.