主論文の要約

論文題目　A Study of Real-world People Re-identification
（現実世界におけるカメラでの人物照合）
氏 名　　ZHANG Guanwen　（張　冠文）

This thesis addresses the people re-identification problem in a real-world scenario. When a person appears in a camera view, people re-identification is used to decide whether this person has already been observed. People re-identification can be used in many wide-area surveillance applications, e.g., locating a person of interest spatio-temporally in a camera network or tracking a specific person globally across multiple cameras. Due to varying camera viewpoints, illumination, non-rigid deformations, and partial occlusions, the appearances of people change significantly. Thesame person may look different in different images, while different people may appear very similar, i.e., the intra-class variance of appearances is even larger than the inter-class variance. The resulting inter-/intra-class variance issue makes people re-identification especially challenging.

In this thesis, we first proposed the local distance comparison for the multiple-shot people re-identification. We demonstrated the multi-modal property of appearance distribution, in which appearance instances distribute with specific semantic meaning of features. An energy-based loss function was then defined to adapt to the multi-modal distribution. Unlike the pair-wise distance comparison, similarity between the image sets is only measured in the local neighbors. The performance of the local distance comparison is found to be superior to conventional approaches of multiple-shot re-identification.

Secondly, we investigated adaptive metric learning to improve the distance metric for local distance comparison. The generic knowledge for the re-identification was leveraged　from the existing data to assist metric learning in the target task. Different from previous approaches, proposed adaptive metric learning is not only related to the generic metric but also closely related to the target metric. The obtained adaptive distance metric is able to guard against over-fitting caused by small training data. Benefiting from adaptive metric learning, the performance of local distance comparison has been significantly boosted up.

Thirdly, we introduced a patch-wise learning for re-identification purpose. Due to non-rigid deformation of people's posture, the holistic appearance representations suffer inherent intra-class variance issues. To suppress such variances, we combine the advantages of hand-crafted representation and metric learning method. The appearance images were partitioned into patches to reduce the ambiguity of non-rigid posture deformation. The metric learning was then performed based on the corresponding patches to obtain optimal distance functions. With the patch-wise set-to-set comparison, the patch-wise learning could achieve competitive performance with the state-of-the-art method.

Finally, we developed a deep convolutional neural network (CNN) approach for single-shot people re-identification. Previous approaches employed either hand-crafted or existing features for appearance representation. Rather than neglecting the inter-/intra-class variance in the feature representation, we extracted feature information directly from input raw images. By feeding learned features into a linear support vector machine, we integrated the feature learning and similarity comparison into one framework for re-identification purpose. Although, the proposed deep CNN is a simple model, it still achieves promising performance.

The four approaches constitute a mutually complementary re-identification system. When there is enough training data with the target task, local distance comparison approach and patch-wise learning approach are able to achieve the re-identification purpose with high accuracy. Specifically, the latter is more robust to posture deformation than the former but costs little more. While, when the target task only has a small amount of training data, as often is the case, the adaptive learning approach is powerful to relate the generic knowledge in multiple tasks with the specific information of the target task and boost up the re-identification performance. In case of a vast amount of data available (this tendency is accelerating), since manually designing effective features becomes impossible, the deep CNN will be the promising approach for people re-identification. It can learn feature representation directly from raw input data to avoid missing important feature information.