Recognition of Human Actions using Visual Local Features

姫，艶麗
九州大学システム情報科学府情報知能工学専攻

https://doi.org/10.15017/25185

（別紙様式２）

氏　　　名　：姫　艶麗

論文題名　　：Recognition of Human Actions using Visual Local Features
　　　　　　　（局所特徴を用いた動画像中の人間動作認識）

区　　　分　：甲

論 文 内 容 の 要 旨

Recognition of human actions in videos is a process of naming actions which are captured by cameras, usually in a simple form of an action verb. Action recognition is an attractive research topic because it is widely applied to computer vision. Its application fields are not limited to video surveillance, human-computer interaction, sport video analysis, computer motion animation, and so on. However, human action recognition is a challenging problem because of two reasons. One reason is owed to quite a lot of appearance variations in human actions, such as various action classes, different physiques of humans and a variety of clothing styles and colors. Furthermore, camera based action recognition need to overcome some difficulties brought by motion sensing, for instance, occlusion, view point changes, scale variation of video screen, etc..

In this thesis, we aim to recognize human actions captured by cameras from basic to complex situations. For the target, we propose a method for local feature calculation, and design a recognition system using these local features. Furthermore, we propose a local feature based method to solve the problem of more complex action recognition: human interaction.

Firstly, a new local feature calculation method is proposed for human action representation. In the method, FAST detector is extended to spatio-temporal space to detect feature points from videos. Then a compact descriptor is proposed which represents actions with compact peak kept histograms of oriented spatio-temporal gradients (CHOG3D). It is calculated in a small spatio-temporal support region around the candidate feature point in order to obtain a compact descriptor. It employs the first order gradient in spatial and temporal orientations for descriptor calculation. In addition, it keeps the peak value of orientation quantized gradient to make the descriptor CHOG3D being able to represent actions more exactly and being distinguished more easily. The efficiency of peak kept is certified by comparing with threshold setting method for action recognition. By parameter training, the optimal parameters for CHOG3D are determined. The local features calculated with FAST and CHOG3D are applied for action recognition using SVM. Based on the computation cost comparison and performance evaluation, the compact descriptor CHOG3D performs well on human action recognition, and it has a lower

computation cost. Though CHOG3D has the limitation of containing less information, a proper quantity of feature points help to overcome the disadvantage.

Secondly, a self-organizing map (SOM) based recognition system is proposed for local feature used human action recognition. In the proposed system, the compact descriptor CHOG3D is adopted for local feature calculation to represent human actions. Then the SOM is employed to train local features and extract key features of actions because of its advantage in mapping data into a low dimension. After training, the key features are assigned action labels of the training data. For action recognition, we adopt k-Nearest Neighbor algorithm (k-NN) to classify features of a testing action sequence into different action classes. By calculating the statistics of feature classification, the action class of the testing sequence is determined. We search for the optimal map size of SOM for training and the proper value k for k-NN classification. With the optimal parameters, we test the proposed method for action recognition on three datasets, KTH, Weizmann and UCF sports datasets and the results certify the efficiency of the proposed recognition system. Compared with the method CHOG3D and SVM, the SOM based method performs better and faster.

Finally, we extend our research to recognize complex human actions, i.e. interactive actions, and propose a contribution estimation method for improving interactive action recognition. Unlike previous algorithms using both of two participants action information, the proposed algorithm estimates the action contribution of participants to select the major participant action for correct interaction recognition. To estimate contributions, we construct contribution interaction model for each interaction category in which both of two participants do major actions. Then we design a method making use of these contribution interaction models to estimate the contribution of participants and classify interaction samples to "co-contribution" or "single-contribution" interactions. Furthermore, we determine the major action in a "single-contribution" interaction. If a given interaction is determined to be "co-contribution," the actions of both the two participants are adopted for recognition. While for "single-contribution" interaction, the major action is selected for recognition. Experiments show that the method is effective for human interaction recognition, which outperforms other methods.