Discriminative Key Pose Extraction using Extended LC-KSVD for Action Recognition

A. Method

0. Ideas

Inspired by the fact that human can easily recognize an action by looking at a few poses rather than the entire sequence. This paper aims to extract key poses, which is the most helpful to recognize the action in the pose sequence.

[Problems to Discuss]

1. How to select key poses?
2.

1. Skeletal Representation

Firstly, normalize all the 3d coordinates. Let P(t) = [P1, P2, …, PN] represent N skeleton joints, where Pi = [xi, yi, zi] stands for the 3d location. A feature vector is combine all the joints' 3d location, velocity and acceleration. Velocity at time t, say V(t), can be represented by P(t+1) - P(t-1). And acceleration at time t, say A(t), can be represented by V(t+1) - V(t-1) = P(t+2) + P(t-2) - 2 * P(t). Thus, [P(t) V(t) A(t)] is the feature vector at time t.

2. The proposed Algorithm

This paper extends the label consistent K-SVD algorithm, so one should be ratehr familiar with K-SVD to understand this paper. But I don't want to spend too much time in this because I study it a little and find that it need many linear mathmetical knowledge that I don't hold now. In the following, Now suppose that we have apply the proposed algorithm correctly, we get a common dictionary and multiple action-specific dictionaries. According to the idea, the action-specific dictionaries is discriminative, so they are considered as the key poses.

[Action-based features]

Once get the action-specific dictionaries, max pooling method(???) is used to construct an action-based feature. Though different action sequence are vary in length, after max pooling, all the features' dimension become the same.

1) Training

Finally, a non-linear lib-SVM classifier is applied to train.

2) Action Recognition

SVM predict, input is the action-based feature of the key poses extracted from the new actions.

B. New things

1) Taken velocity and acceleration into consideration is a rather normal mesure, although I don’t think this will help too much in practice. The result is not very high.

2) Action-based feature is a good idea to master the macroscopic view, because frame-level often involve with noise and variance trouble.

C. Shortcomings

1) SVD will lose some data, and the velocity and acceleration information didn't be explicitly made use of. It gives me an impression that the frame-base feature is combined casually instead of a considerate design.