Action Recognition based on Subdivision-Fusion Model

Zongbo Hao[1]
zbhao@uestc.edu.cn
Linlin Lu[1]
fountainous@gmail.com
Qianni Zhang[2]
qianni.zhang@qmul.ac.uk
Jie Wu[1]
jiewu.uestc@gmail.com
Ebroul Izquierdo[2]
ebroul.izquierdo@qmul.ac.uk
Juanyu Yang[1]
yangjuanyu2008@163.com
Jun Zhao[1]
zhaojun.hjh@gmail.com

[1] School of Information and Software Engineering
University of Electronic Science and Technology of China
[2] Multimedia and Vision Group
School of Electronic Engineering and Computer Science
Queen Mary, University of London

This paper proposes a novel Subdivision-Fusion Model (SFM) to recognize human actions. In most action recognition tasks, overlapping feature distribution is a common problem leading to overfitting. In the subdivision stage of the proposed SFM, samples in each category are clustered. Boundaries for the more concentrated subcategories are easier to find and as consequence overfitting is avoided. In the subsequent fusion stage, we convert the multi-subcategories classification back to the original classification problem. Two methods to determine the number of clusters are provided. We evaluated our model SFM on four popular datasets to demonstrate the enhancement in performance.

Motivation

In most recognition tasks, the overlapping feature distribution is popular, as presented in Figure 1(a). Category 3 and 4 are both distributed in two regions and overlapped, instead of distributing in separating regions as supposed, like category 1 and 2 do. In this case it is not wise to find the boundary that enclose each category into one region and not overlapped with others, as shown in Figure 1(b). Trying to find that kind of boundaries, overfitting will be obviously resulted in. Hence, category 3 and 4 can be divided into two subcategories respectively, and the smooth boundaries can be found as shown in Figure 1(c).

Subdivision-Fusion Model (SFM)

Suppose there are n samples in the dataset, which are grouped into L categories. The features are extracted by some kind of models:

$$F(X) = [f_1, f_2, ..., f_n] = h([x_1, x_2, ..., x_n]) \quad (1)$$

where f_i is the feature vector for each sample x_i, h is the transformation by the network from the sample space $\mathbb{R}^s$ to feature space $\mathbb{R}^F$.

As presented in the motivation part, we can subdivide each category using clustering algorithm. The data of the i^{th} category in the feature space are grouped into K_i clusters, and the labels are updated by the results of clustering. Let M be the total number of updated subcategories. Sparse Subspace Clustering (SSC) [1] is employed in clustering in this paper. By clustering and updating the labels, we made a transformation from the feature space to the sample space:

$$Y' = [y_1', y_2', ..., y_n'] = c([f_1, f_2, ..., f_n]) \quad (2)$$

where c is the clustering operation, which transforms features to labels.

In Figure 2, Layer V with M nodes acts as an M-subcategory classifier. We connect an output Layer O with L nodes to Layer V. Let $W(i,k)$ be the weight between node k in Layer V and node i in Layer O:

$$W(i,k) = \begin{cases} 1, V_k \in O_i \\ 0, V_k \notin O_i \end{cases} \quad (3)$$

where V_k is the k^{th} subcategory output in Layer V, O_i is the i^{th} category output in Layer O. If subcategory k belongs to category i, $W(i,k) = 1$, otherwise $W(i,k) = 0$. The final classification result R is calculated by:

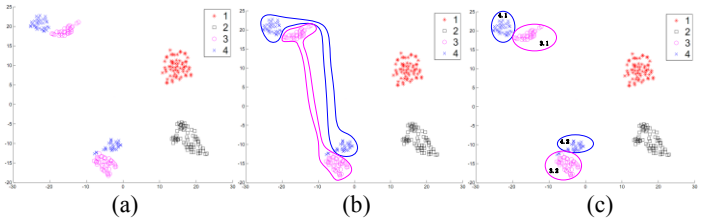
$$R = \max(O_i), O_i = \sum_{k=1}^{M} W(i,k) * V_k, 1 \le i \le L, 1 \le k \le M \quad (4)$$

According to (4), the M-subcategory classification is converted back to the L-category classification which is identical to the original problem.

The rule of determining cluster number K_i

1) Considering the overlapping feature distribution:

Our first method to determine the proper number K_i of subcategories is based on observing the t-Distributed Stochastic Neighbor Embedding (t-SNE) [2] 2-dimensional visualization directly.



(a) (b) (c)

Figure 1: Subdivision in each class. (a) Sample distribution. (b) Finding boundaries to distinguish category 3 and 4 in a usual way, it is prone to be overfitting. (c) Clustering in each class makes it easier to distinguish the classes and avoid overfitting.

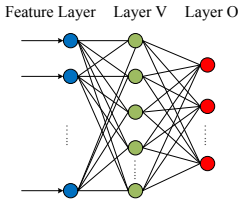


Figure 2: Classifier in the Fusion Stage

2) Considering the class imbalance problem:

Our second method to determine the proper number K_i of subcategories is based on the ratio of the number of each category samples to the number of the minority category samples.

Experimental results

SFM has been thoroughly tested with four popular datasets:

1. Hollywood2 dataset: the accuracy has been improved from 53.3% with Convolutional ISA [4] to 79.4%, significantly outperforming the state-of-the-art accuracy of 64.3% of Improved Trajectories [3].
2. YouTube Action dataset: the performance has been improved from 75.8% with Convolutional ISA [4] to 82.5%.
3. KTH dataset: the accuracy has been improved from 90.2% with 3D CNN [5] to 94.0%.
4. UCF50 dataset: the performance has been slightly improved from 76.4% with Action Bank features [6] to 76.9%.

[1] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(11), 2765-2781, 2013.

[2] L. Maaten and G.E. Hinton. Visualizing data using t-SNE. Journal of Machine Learning, 85(9), 2579-2605, 2008.

[3] H. Wang and C. Schmid. Action recognition with improved trajectories. In Computer Vision, IEEE International Conference on, 3551-3558, 2013.

[4] Q. V. Le, W. Zou, S. Yeung, and Andrew Y. Ng. Learning hierarchical invariant spatio-temporal feature for action recognition with independent subspace analysis. In Computer Vision and Pattern Recognition, IEEE Conference on, 3361-3368, 2011.

[5] S. Ji, W. Xu, M. Yang, and K.Yu. 3D convolutional neural networks for human action recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 221-231, 2013.

[6] S. Sadanand and J. Corso. Action bank: A high-level representation of activity in video. In Computer Vision and Pattern Recognition, IEEE Conference on, 1234-1241, 2012.