A Comparative Study of Encoding, Pooling and Normalization Methods for Action Recognition

Xingxing Wang[1] (xx.wang@siat.ac.cn) Limin Wang [1,2](lm.wang@siat.ac.cn) Yu Qiao [1,2] (yu.qiao@siat.ac.cn)

[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China, [2]Department of Information Engineering, The Chinese University of Hong Kong

Introduction

Motivation. Bag of visual words (BoVW) models have been widely and successfully used in video based action recognition. One key step in constructing BoVW representation is to encode feature with codebook. Recently, a number of new encoding methods have been developed to improve the performance of BoVW based object recognition and scene classification, but their effects for action recognition are still unknown.

Overview. The main objective of this paper is as follows,

I. evaluate and compare these new encoding methods in the context of video based action recognition
II. analyze and evaluate the combination of encoding methods with different pooling and normalization strategies.

Results. Our experiments show that new encoding methods can significantly improve the recognition accuracy compared with classical VQ.

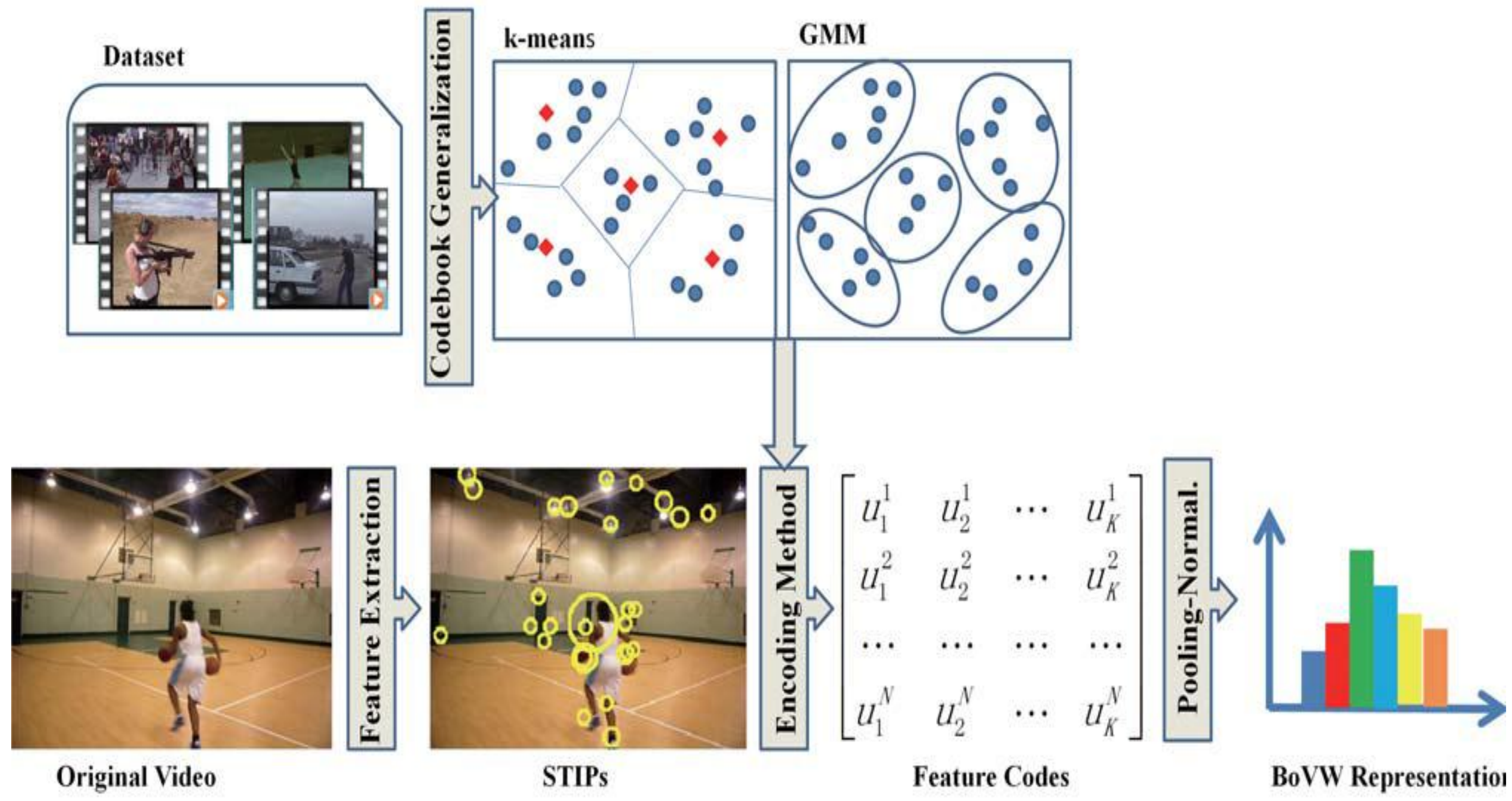

Figure 1. BoVW model for action recognition

Methods

Codebook Generation Methods :

I. K-means:

$$\min \mathcal{J}(\{r_{mk}, d_k\}) = \sum_{m=1}^{M}\sum_{k=1}^{K} r_{mk}\|\mathbf{x}_m - \mathbf{d}_k\|^2.$$

II. GMM:

$$p(\mathbf{x};\theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x};\mu_k,\Sigma_k).$$

Encoding Methods:

I. Vector Quantization (VQ).

$$u_{nk} = \begin{cases} 1. & \text{if } k = \arg\min_k \|\mathbf{x}_n - \mathbf{d}_k\|^2. \\ 0. & \text{otherwise.} \end{cases}$$

II. Soft-assignment Encoding (SA).

$$u_{nk} = \frac{\exp(-\beta\|\mathbf{x}_n - \mathbf{d}_k\|^2)}{\sum_{j=1}^{K}\exp(-\beta\|\mathbf{x}_n - \mathbf{d}_j\|^2)}, \quad u_{nk} = \frac{\exp(-\beta\hat{d}(\mathbf{x}_n,\mathbf{d}_k))}{\sum_{j=1}^{K}\exp(-\beta\hat{d}(\mathbf{x}_n,\mathbf{d}_j))},$$

$$\hat{d}(\mathbf{x}_n,\mathbf{d}_k) = \begin{cases} \|\mathbf{x}_n - \mathbf{d}_k\|^2 & \text{if } \mathbf{d}_k \in N_k(\mathbf{x}_n), \\ \infty & \text{otherwise,} \end{cases}$$

III. Sparse Encoding (SPC).

$$\mathbf{u}_n = \arg\min_{\mathbf{u}\in\mathbb{R}^K} \|\mathbf{x}_n - \mathbf{D}\mathbf{u}\|^2 + \lambda\|\mathbf{u}\|_1.$$

IV. Locality-constrained Linear Encoding (LLC).

$$\mathbf{u}_n = \arg\min_{\mathbf{u}\in\mathbb{R}^K} \|\mathbf{x}_n - \mathbf{D}\mathbf{u}\|^2 + \lambda\|\mathbf{s}_n \odot \mathbf{u}\|^2.$$
$$\text{s.t.} \quad \mathbf{1}^T\mathbf{u}_n = 1.$$

$$\mathbf{s}_n = \exp\left(\frac{\mathrm{dist}(\mathbf{x}_n,\mathbf{D})}{\sigma}\right),$$

V. Fisher Kernel Encoding (FK).

$$\mathcal{G}^{\mathbf{X}}_{\mu,k} = \frac{1}{T\sqrt{\pi_k}}\sum_{t=1}^{T}\gamma_t(k)\left(\frac{\mathbf{x}_t - \mu_k}{\sigma_k}\right),$$

$$\mathcal{G}^{\mathbf{X}}_{\sigma,k} = \frac{1}{T\sqrt{\pi_k}}\sum_{t=1}^{T}\gamma_t(k)\left[\frac{(\mathbf{x}_t - \mu_k)^2}{\sigma_k^2} - 1\right].$$

Pooling and Normalization methods:

I. Pooling

Sum pooling, With sum pooling scheme , the k^{th} component of p is $p_k = \sum_{n=1}^{N} u_{nk}$

Max pooing, With max pooling scheme , the k^{th} component of p is $p_k = \max\{u_{1k}, u_{2k}, \cdots, u_{nk}\}$

II. Normalization

L1, In $\ell 1$ normalization , feature $\boldsymbol{p}$ is normalized by its $\ell 1$-norm: $p = p/\sum_{k=1}^{K}|p_k|$

L2, In $\ell 2$ normalization [4], feature $\boldsymbol{p}$ is normalized by its $\ell 2$-norm: $p = p/\sqrt{\sum_{k=1}^{K} p_k^2}$

Power, In power normalization , we apply the following function for each dimension of feature $\boldsymbol{p}$: $f(p_k) = \mathrm{sign}(p_k)|p_k|^{\alpha}.$

Evaluation

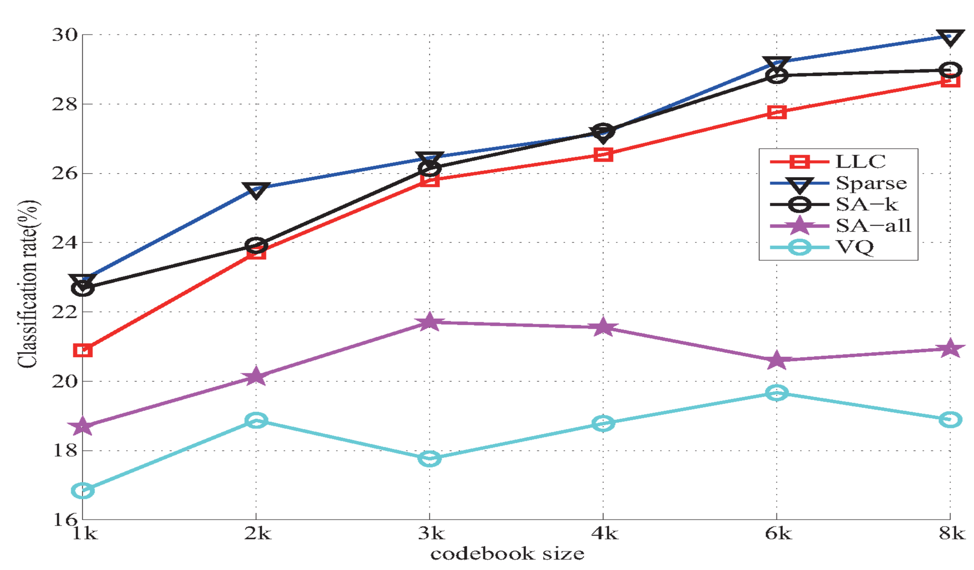

Figure 2. Exploration of performance of different encoding methods with changing codebook size

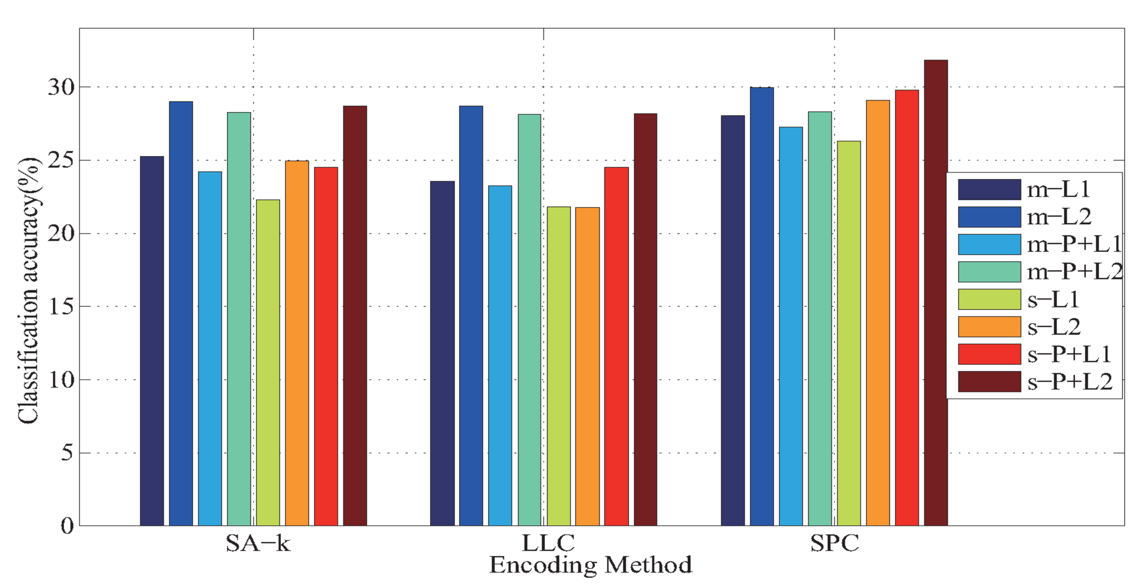

Figure 3. Comparison of different pooling-normalization strategies on HMDB51

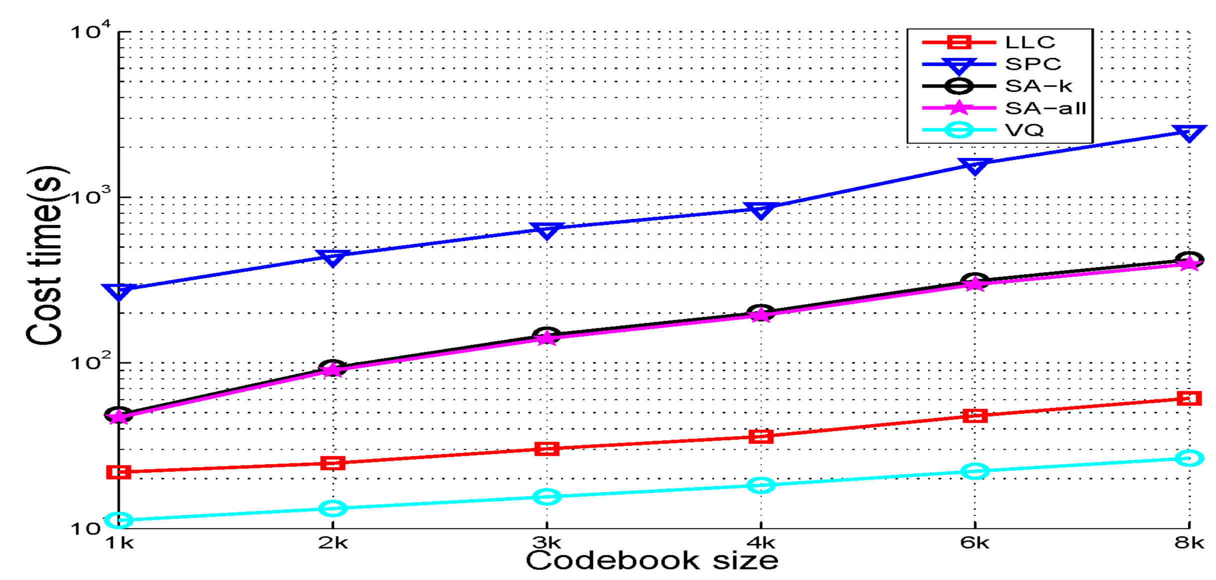

Figure 4. Exploration of the computational cost of different encoding methods

Method	Ours(FK)	Schuldt	Laptev	Ryoo	Liu	Sada
Accuracy(%)	92.1	71.7	91.8	91.1	91.6	98.2

Table 1. Comparison the proposed methods with state of the art on KTH.

Method	Our(SPC-s-P_l2)	HOG/HOF	C2	Action Bank
Accuracy(%)	**31.82**	20.44	22.83	**26.9**

Table 2. Comparison the proposed methods with state of the art on HMDB51.