Efficient Sparsity Estimation via Marginal-Lasso Coding

Tzu-Yi Hung[1], Jiwen Lu[2], Yap-Peng Tan[1], and Shenghua Gao[3]

[1] School of Electrical and Electronic Engineering,
Nanyang Technological University, Singapore
[2] Advanced Digital Sciences Center, Singapore
[3] ShanghaiTech University, Shanghai, China

Abstract. This paper presents a generic optimization framework for efficient feature quantization using sparse coding which can be applied to many computer vision tasks. While there are many works working on sparse coding and dictionary learning, none of them has exploited the advantages of the marginal regression and the lasso simultaneously to provide more efficient and effective solutions. In our work, we provide such an approach with a theoretical support. Therefore, the computational complexity of the proposed method can be two orders faster than that of the lasso with sacrificing the inevitable quantization error. On the other hand, the proposed method is more robust than the conventional marginal regression based methods. We also provide an adaptive regularization parameter selection scheme and a dictionary learning method incorporated with the proposed sparsity estimation algorithm. Experimental results and detailed model analysis are presented to demonstrate the efficacy of our proposed methods.

Keywords: Sparsity estimation, marginal regression, sparse coding, lasso, dictionary learning, adaptive regularization parameter.

1 Introduction

Sparse coding has been successfully applied to many machine learning and computer vision tasks, including image classification [30,31,8], face recognition [28,5], activity-based person identification [22,15], etc. Sparse coding refers to a feature quantization mechanism which obtains a codebook to encode each input signal as a sparse histogram feature based on the linear combination of a few visual words. Due to its soft assignment principal, it can make the quantization error much smaller [9,22]. Moreover, sparse coding is more robust to the noise and easily incorporating into the bag-of-feature (BoF) framework.

The least absolution shrinkage and selection operator which is also known as lasso [27] in short is a popular sparse coding model in statistics signal processing which has wildly used in sparsity estimation. It simultaneously minimizes the quantization error, and impose an L_1 penalty with a regularization parameter λ which controls the sparsity of the coefficients. While there exists many efficient algorithms [20,7,29], however, finding the lasso solutions remain a computational task with the complexity $O(p^3 + dp^2)$ where d refers to the input feature dimension, and p refers to the amount of visual

D. Fleet et al. (Eds.): ECCV 2014, Part IV, LNCS 8692, pp. 578–592, 2014.

words in the dictionary. Thus, it hardly supports large scale data analysis efficiently and effectively [6,23,11,1].

Recently, marginal regression, has been revisited and shown its efficient performance for visual word selection and sparsity estimation [6,11,1]. For each feature, it calculates the correlation with each visual word and imposes a tuning parameter to achieve the sparsity. While marginal regression is simple and fast with the complexity $O(dp)$, it usually considers and utilizes a fixed tuning parameter which is not always the good cut-off point for each feature and could result in a large quantization error since some coefficients may be shrunk too much, but some may be included too much noise. Hence, how to determine a suitable cut-off point becomes an typical problem in featuring coding. To address this, the sure independent screening (SIS) [6] method imposes an L_1-norm penalty and an energy constraint E to each marginal regression coefficient vector. It sorts coefficients in terms of their absolute values and selects the top k coefficients whose L_1 norm is bounded by E so that each sparse code can be represented by similar sparsity level. However, this method may also cause a large quantization error, because marginal regression works well when visual words have low correlation in the dictionary [1,23]; however, in general, they are highly correlated, and thus may fail to clarify the relationship between visual words and the input feature.

While there are many works working on sparse coding and dictionary learning, none of them has exploited the advantages of the marginal regression and the lasso simultaneously to provide more efficient and effective solutions. To this end, we provide in this paper such an approach with a theoretical support. We propose an efficient sparsity estimation approach using Marginal-Lasso Coding (MLC) which quantizes each feature into a small set of visual words using marginal regression combined with the lasso framework. Our model represents each sparse code with similar sparsity level bounded by a global sparsity energy and simultaneously shrinks individual coefficients to alleviate the bias of sparse codes caused by the highly correlated visual words. Moreover, our approach automatically determines the shrinking regularization parameter for each feature and further designs a self-ratio energy constraint of sparsity level which is different from the traditional constraint with a case-dependent chosen value.

Contribution: 1) We propose an efficient Marginal-Lasso Coding (MLC) framework for feature quantization with sparsity estimation, regularization parameter selection and dictionary learning. 2) We exploit the advantages of the marginal regression and the lasso simultaneously to provide more efficient and effective solutions. 3) We determine the regularization parameter automatically and adaptively for each individual feature and provide a self-ratio energy bound. 4) We provide a theoretical support for the proposed model. 5) We successfully apply the proposed method to various recognition tasks.

2 Related Work

Our approach is related to the general sparse coding model which has been widely used for sparse representation [28,27,30,22], and there exist many algorithms to estimate sparse codes, such as least angle regression [4], gradient descent [7,29], and feature-sign search [20]. The feature-sign search method [20,8,31] is a popular and efficient solution for sparsity estimation. It continuously selects and updates potential candidates in an active set to maintain these nonzero coefficients and their corresponding signs until reaching the constraint. However, with a large dictionary size, the active set would become large and the feature-sign searching process may be hard to terminate [1,23]. While these methods aim to obtain a small quantization error, they have a large computational cost [6,11]. Therefore, different from traditional sparsity estimation techniques which estimate sparse codes by continuously selecting variables to the active set based on the searching criteria, we incorporate marginal regression coefficients into the sparse coding model which is more efficient for sparsity estimation.

Recently, Fan and Lv [6] proposed a sure independent screening (SIS) method using marginal regression to obtain sparse codes efficiently and can easily deal with large-scale problems. Genovese *et al.* [11] provided a statistical comparison of the lasso and marginal regression. Due to the promising performance in terms of accuracy and speed, Krishnakumar *et al.* [1] applied marginal regression to learn sparse representation for visual tasks. While these solutions use hard-thresholding-type methods, we propose in this work a soft-thresholding-type method with a theoretical support to use marginal regression combined with the lasso model for sparsity estimation.

3 Efficient Sparsity Estimation via Marginal-Lasso Coding

3.1 Sparsity Estimation via Marginal Regression

Consider a regression model $x = Us + z$ with an input feature $x \in \mathbb{R}^d$, a coefficient vector $s \in \mathbb{R}^p$, a column-wisely normalized dictionary $U = [u_1, \ldots, u_p] \in \mathbb{R}^{d \times p}$ ($p \gg d$) and a noise vector $z \in \mathbb{R}^d$. The feature quantization task is to quantize an input feature x into a feature histogram s with a few non-zero elements so that the linear combination of U and s can gain a minimized quantization error. The mathematical expression can be shown as follows:

$$\min_{s} \frac{1}{2}\|x - Us\|_2^2 \tag{1}$$

To estimate the sparse code s, we first compute the component-wise marginal regression coefficients $\hat{a}$:

$$\hat{a} \equiv U^T x \tag{2}$$

where $\hat{a}^{(k)} = u_k^T x$ is the kth element of $\hat{a}$ and u_k is the kth column of the dictionary U. Then, we threshold the coefficients in terms of their absolute values using a tuning parameter $t > 0$ [3,11] so that the coefficients whose absolute values are larger than t are kept and the rest are set to zero. The mathematical expression can be written as follows:

$$\hat{s}^{(k)} = \begin{cases} \hat{a}^{(k)} & \text{if } |\hat{a}^{(k)}| > t \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

According to Eq. 3, however, utilizing a fixed tuning parameter may cause a large quantization error since it is not always the good cut-off point for each feature. Instead of cut-off point selection, we can select the top k large coefficients in terms of their absolute values whose L_1-norm is bounded by a constraint E [6,1]. The equation can be reformulated as follows:

$$\hat{s}^{(k)} = \begin{cases} \hat{a}^{(k)} & \text{if } k \in B \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where

$$B = \{J_1, \ldots, J_r : r \leq p : \sum_{k=1}^{r} |\hat{a}^{(J_k)}| \leq E\} \tag{5}$$

We sort the coefficients in terms of their absolute values in descending order which denotes by indexes $J_1, \ldots, J_p$ where $|\hat{s}^{(J_1)}| > |\hat{s}^{(J_2)}| > \cdots > |\hat{s}^{(J_p)}|$. The two thresholding approaches above are hard-thresholding-type methods.

3.2 Marginal-Lasso Coding

While marginal regression has shown its effectiveness, there still exists three limitations: (i) Marginal regression works well when the columns of dictionary have low correlation [1,23]. However, in general, the columns of over-complete dictionaries are usually highly correlated, and thus, in a linear regression model, their coefficients can become poorly determined and exhibit high variance. (ii) In addition, the fixed regularization parameter may cause a large quantization error. Some coefficients may be shuck too much and some may be included too much noise. (iii) Moreover, the existing marginal regression solutions are hard-thresholding-type methods, and there are no theoretical supports.

To address this, we consider the following conditions: (1) The regularization parameter should be chosen adaptively and automatically for each feature to minimize individual estimate of expected quantization error. (2) The L_1-norm in the lasso framework should be considered locally for visual word selection and coefficient shrinkage to alleviate the bias of sparse codes caused by the highly correlated visual words. (3) The L_1-norm in the constraint should be bounded locally and globally for sparse representation. To this end, we introduce a Marginal-Lasso Coding (MLC) method to better characterize input features and provide a theoretically soft-thresholding-type solution. Similar to the existing works, we first compute the marginal regression coefficients of each feature via Eq. 2. Then, we consider the following optimization problem:

$$\begin{aligned} &\min_{s,\lambda} \frac{1}{2}\|U^T x - s\|_2^2 + \lambda\|s\|_1 \\ &\text{subject to } \|s\|_1 \leq E \end{aligned} \tag{6}$$

The first term is the quantization error, and the second term is a local L_1 penalty with an adaptive regularization parameter λ which controls the cut-off point selection and the shrinkage of coefficients. The global sparsity constraint E bounds each sparse code to the similar sparsity level. In this model, we aim to estimate a sparse code and a parameter λ so that the quantization error can be minimized and the coefficients can be shrunk simultaneously for sparse representation. When considering multiple features, the optimization problems can be refined as follows:

$$\begin{aligned} \min_{\substack{s_1,\ldots,s_N \\ \lambda_1,\ldots,\lambda_N}} \ & \frac{1}{2}\|U^T X - S\|_F^2 + \sum_{i=1}^{N} \lambda_i \|s_i\|_1 \\ & \text{subject to } \|s_i\|_1 \le E \quad \forall i \end{aligned} \tag{7}$$

where $X = [x_1, \ldots, x_N] \in \mathbb{R}^{d \times N}$ be a set of features, and $S = [s_1, \ldots, s_N] \in \mathbb{R}^{p \times N}$ be a set of corresponding sparse codes. Since the sparsity energy E is case-dependent variable, we provide here another self-ratio energy e such that each feature can obtain its sparse code by proportionally shrinking their marginal coefficients to the desired level. The constraint can be reformulated as follows:

$$\begin{aligned} \min_{\substack{s_1,\ldots,s_N \\ \lambda_1,\ldots,\lambda_N}} \ & \frac{1}{2}\|U^T X - S\|_F^2 + \sum_{i=1}^{N} \lambda_i \|s_i\|_1 \\ & \text{subject to } \frac{\|f_\lambda(U^T x)\|_1}{\|U^T x\|_1} \le e \quad \forall i \end{aligned} \tag{8}$$

where $f_\lambda(A)$ refers to the solution of Eq. 3 with $\hat{a} = A$ and $t = \lambda$. While Eq. 7 considers the similar L_1-norm between sparse codes, Eq. 8 considers the similar self-ratio between each other.

3.3 Sparsity Estimation

When λ is fixed, we can solve the optimization problem of Eq. 6 by rewriting the objective as follows:

$$\min_s \ \frac{1}{2}\|\hat{a} - s\|_2^2 + \lambda \|s\|_1 \tag{9}$$

Minimizing the Eq. 9 is equivalent to minimize individual element errors with their corresponding absolute values. Thus, we can rewrite the objective as follows:

$$\min_s \ \frac{1}{2}\sum_{k=1}^{p} (\hat{a}^{(k)} - s^{(k)})^2 + \lambda |s^{(k)}| \tag{10}$$

Since the marginal regression coefficients are computed independently for each element, hence, we can decompose the problem into p separate optimization tasks. For each task k, the optimization problem can be defined as follows:

$$\min_{s^{(k)}} \ (\hat{a}^{(k)} - s^{(k)})^2 + \lambda |s^{(k)}| \tag{11}$$

Then, by solving the above optimization problem, we can obtain a soft-thresholding-type optimal solution of each element $s^{*(k)}$:

$$\begin{aligned} s^{*(k)} &= \begin{cases} \hat{a}^{(k)} - \lambda & \text{if } \hat{a}^{(k)} > \lambda \\ \hat{a}^{(k)} + \lambda & \text{if } \hat{a}^{(k)} < -\lambda \\ 0 & \text{otherwise} \end{cases} \\ &= \operatorname{sign}(\hat{a}^{(k)})(|\hat{a}^{(k)}| - \lambda)_+ \end{aligned} \tag{12}$$

where $sign(j)$ refers to the sign of j, and $(l)_+$ means to keep the value when l is positive and set it to zero when l is negative. The L_1-norm can be defined as the sum of absolute elements $\|s^*\|_1 = \sum_{k=1}^{p} |\text{sign}(\hat{a}^{(k)})(|\hat{a}^{(k)}| - \lambda)_+|$. In Eq. 12, the regularization parameter λ controls the element selection and the shrinkage of the coefficients. If λ equals to zero, then $\|s^*\|_1 = \|\hat{a}\|_1$. Instead, if $\lambda > 0$, then some coefficients will be directly set to zero and others will be shrunk towards zero automatically so that $\|s^*\|_1 < \|\hat{a}\|_1$. By maximizing $\|s^*\|_1$ bounded by the global sparsity constraint E, the individual λ^* can be calculated as follows:

$$\begin{aligned} \lambda^* &= \arg\max_{\lambda} \|s^*\|_1 = \arg\max_{\lambda} \sum_{k=1}^{p} |\text{sign}(\hat{a}^{(k)})(|\hat{a}^{(k)}| - \lambda)_+| \\ &\text{subject to } \|s^*\|_1 \leq E \end{aligned} \tag{13}$$

By doing so, we can estimate each element of the sparse code $s^{*(k)} = \text{sign}(\hat{a}^{(k)})(|\hat{a}^{(k)}| - \lambda^*)_+$ which is a soft-thresholding-type solution. When considering a set of features, the optimization problem of Eq. 7 or 8 can be solved via Algorithm 1.

4 Marginal-Lasso Coding for Dictionary Learning

The dictionary U is a set of normalized visual words denoting by each column u_k. To learn the dictionary, we iteratively optimize U and S by obtaining S with a fixed U and updating U based on a given S. The sparsity estimation S via the marginal-lasso model has been introduced in the previous subsection. In this subsection, we introduce the dictionary learning method. Since marginal regression works well when the columns in the dictionary have low correlation [6,11,1,23], we formulate the optimization problem of the dictionary learning as follows:

$$\begin{aligned} &\min_{U} \sum_{i=1}^{N} \|x_i - Us_i\|_2^2 + \gamma \|U^T U - I\|_F^2 \\ &\text{subject to } u_k^T u_k \leq 1 \quad \forall k \end{aligned} \tag{18}$$

We simultaneously minimize the quantization errors denoting by the first term and the correlations between columns in the second term with a regularization parameter γ. This optimization problem can be solved via the first-order gradient descent method [23,1]:

$$U_{q+1} = \Pi_U \{U_q - \beta \nabla F(U_q)\} \tag{19}$$

where

$$F(U) = \sum_{i=1}^{N} \|x_i - Us_i\|_2^2 + \gamma \|U^T U - I\|_F^2 \tag{20}$$

and

$$\nabla F(U_q) = 2(U_q S S^T - X S^T) + 4\gamma (U_q U_q^T U_q - U_q) \tag{21}$$

$\nabla F(U_q)$ is the gradient of $F(U)$ with respect to U at each iteration q, parameter β is the step size, and $\mathit{\Pi}_U$ is the projection function which maps each column u_k to the

Algorithm 1. Efficient Sparsity Estimation via Marginal-Lasso Coding

Require: Data $X = [x_1, \dots, x_N] \in \mathbb{R}^{d \times N}$, a learned dictionary $U = [u_1, \dots, u_p] \in \mathbb{R}^{d \times p}$, and a specific value of the sparsity constraint E or a self-ratio constraint e

Ensure: Sparse coefficients $S = [s_1, \dots, s_N] \in \mathbb{R}^{p \times N}$

1: For x_i $\forall i$, compute the marginal regression coefficients

$$\hat{a}_i{}^{(k)} = \frac{u_k^T x_i}{\|u_k\|_2^2} \quad \forall k \tag{14}$$

2: Calculate the regularization parameter λ_i for each x_i $\forall i$ based on the sparsity constraint E of Eq. 6 and 13

$$\begin{aligned} \lambda_i^* &= \arg\max_{\lambda_i} \|s_i^*\|_1 = \arg\max_{\lambda_i} \sum_{k=1}^{p} |\text{sign}(\hat{a}_i{}^{(k)})(|\hat{a}_i^{(k)}| - \lambda_i)_+| \\ &\text{subject to } \|s_i\|_1 \leq E \end{aligned} \tag{15}$$

Or calculate the regularization parameter λ_i for each x_i $\forall i$ based on the self-ratio constraint e in Eq. 8

$$\begin{aligned} \lambda_i^* &= \arg\max_{\lambda_i} \|s_i^*\|_1 = \arg\max_{\lambda_i} \sum_{k=1}^{p} |\text{sign}(\hat{a}_i^{(k)})(|\hat{a}_i^{(k)}| - \lambda_i)_+| \\ &\text{subject to } \frac{\|f_{\lambda_i}(U^T x_i)\|_1}{\|U^T x_i\|_1} \leq e \end{aligned} \tag{16}$$

3: Obtain sparse codes

$$s_i^{*(k)} = \begin{cases} \hat{a}_i^{(k)} - \lambda_i^* & \text{if } \hat{a}_i^{(k)} > \lambda_i^* \\ \hat{a}_i^{(k)} + \lambda_i^* & \text{if } \hat{a}_i^{(k)} < -\lambda_i^* \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

L_2-norm unit ball. More specifically, the dictionary is learned iteratively by finding the local minimum along the gradient direction based on the small step size and normalizing each column vector with length smaller or equal to one until convergence. The sparsity estimation and the dictionary learning process is shown in Algorithm 2.

5 Experimental Results

In this section, we describe the experimental settings and analyze the proposed method on the image classification, action recognition and activity-based human identification tasks. We denote the proposed marginal-lasso coding approach by MLC with Matlab implementation. We compare the proposed method with 5 algorithms: (1) The lasso model with feature-sign search (LASSO-FS): the code has been implemented by its authors in Matlab [20]. (2) The lasso model with least angle regression (LASSO-LAR): we used the code from SPAMS software implemented by C++ [23]. (3) The marginal regression model with the hard-thresholding method (MR-Hard): we implemented it in

Algorithm 2. Sparsity Estimation and Dictionary Learning

Require: Data $X = [x_1, \ldots, x_N] \in \mathbb{R}^{d \times N}$, an initial normalized dictionary $U = [u_1, \ldots, u_p] \in \mathbb{R}^{d \times p}$, and an energy bound with value E or self-ratio e

Ensure: Learned dictionary $U \in \mathbb{R}^{d \times p}$, sparse coefficients $S = [s_1, \ldots, s_N] \in \mathbb{R}^{p \times N}$

1: **repeat**

2: Obtain sparse code via Algorithm 1 using the latest dictionary.

$$\begin{aligned} \min_{\substack{s_1,\ldots,s_N \\ \lambda_1,\ldots,\lambda_N}} \; & \frac{1}{2}\|U^T X - S\|_F^2 + \sum_{i=1}^{N} \lambda_i \|s_i\|_1 \\ \text{subject to } & \|s_i\|_1 \leq E \quad \forall i \end{aligned} \tag{22}$$

3: Update the dictionary based on the sparse codes obtained from the previous step.

$$\begin{aligned} \min_{U} \; & \sum_{i=1}^{N} \|x_i - U s_i\|_2^2 + \gamma \|U^T U - I\|_F^2 \\ \text{subject to } & u_k^T u_k \leq 1 \quad \forall k \end{aligned} \tag{23}$$

4: **until** convergence

Matlab according to Eq. 3. (4) The marginal regression model with the soft-thresholding method (MR-Soft): we implemented it in Matlab according to Eq. 3 with soft thresholding. (5) The marginal regression model with the sure independence screening method (MR-SIS): we implemented it in Matlab according to Eq. 4.

5.1 Image Classification

We evaluate the performance of the proposed method for the image classification task on USPS and Scene 15 datasets. USPS is a handwritten digits dataset [14,31] which consists of 7291 training images and 2007 test images of size 16×16 with digits 0 to 9. We represent each sample by a 256-dimensional vector. The dictionary sizes are selected from $32, 64, 128, 256, 512, 1024$ under different training sizes. For the penalty constraint, we adopt self-ratio in Eq. (8), and the parameters are set empirically via cross-validation. To evaluate the performance, we train a linear SVM classifier and compare to original regression (OR) and sparse coding (SC) methods. Table 1 shows the results under different sizes of training samples from 100 to 7291. As shown, the proposed MLC method outperforms the other two methods in most of the cases and obtains the best accuracy result 95.3 under 5000 training sample size.

Scene 15 [24,21,19] contains 4485 images with 15 categories. Each category consists of ranging from 200 to 400 images with average size of 300×250 pixels. This dataset is very diverse from indoor such as living room, kitchen, office, store, etc., to outdoor including street, inside city, building, mountain, forest, etc. Following the same setting as the previous works [19,30], we randomly select 100 images per category as training samples and the rest as test samples and incorporate the proposed method with spatial pyramid matching which denotes by MLcSPM. Table 2 shows the results. As shown, our MLcSPM method outperforms other methods, especially for the

Table 1. Recognition accuracy (%) of different methods on the USPS dataset

Method \ Size of training set	100	500	1000	2000	5000	7291
OR	81.7	88.4	90.1	91.4	92.3	93.8
SC	79.6	88.0	90.4	91.8	92.9	94.5
MLC	74.5	91.1	92.9	94.2	95.3	94.0

Table 2. Recognition accuracy (%) of different methods on the Scene 15 dataset

Method \ Size of training set	100
KSPM [19]	81.40
KC [10]	76.67
ScSPM [30]	80.28
HardSPM	75.72
SoftSPM	79.48
SISSPM	81.37
MLcSPM	82.54

popular lasso-based ScSPM scheme. Further, the proposed MLcSPM outperforms other marginal regression based methods.

5.2 Action Recognition

We analyze the characteristic of the proposed method using the KTH action recognition dataset. The KTH dataset [26] contains 25 persons. Each person performed 6 different activities, including boxing, handclapping, handwaving, jogging, running, and walking, respectively. There are 4 scenarios under each activity, including outdoors, outdoors with scale variation, outdoors with different clothes and indoors, respectively, resulting 599 video sequences in total. We follow the original experimental setup which divides clips into 2391 subclips with 863 clips for the test set (9 subjects) and 1528 clips for the training set (16 subjects). We use three space-time local feature descriptors in the experiments, HOG, HOF and HOGHOF [18] obtained by a 3D-Harris detector [17], and quantize each local feature into a sparse code via Algorithm 1. After that, we represent each video clip using the maximum pooling method and use a non-linear support vector machine [2] with a X^2 kernel [18]. We select the parameters using cross validation. To speedup the dictionary learning process, we firstly learn an initial dictionary using online dictionary learning method [23] with 100 runs as the warm start.

Comparison between Hard-Thresholding and Soft-Thresholding: We encode each HOF local feature using a hard-thresholding technique, MR-Hard, and a soft-thresholding technique, MR-Soft, respectively, with dictionary size 4000 in the training set as the tunning parameter t varies. Fig. 1 (a) shows the average L_1-norm energy of sparse codes under different t. As shown, using the soft-thresholding scheme is a more effective way for variable selection and coefficient shrinkage. Further, the soft-thresholding

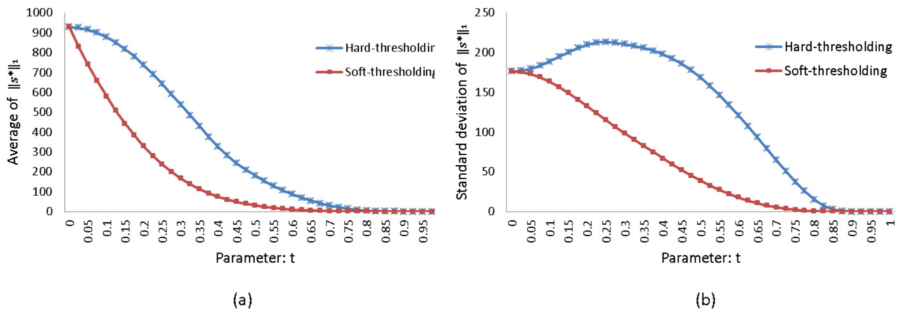

Fig. 1. Comparison of the soft and hard-thresholding schemes. (a) The average L_1-norm energy of sparse codes under different t. As shown, using the soft-thresholding scheme is a more effective way for variable selection and coefficient shrinkage. (b) The average standard deviation of the L_1-norm energy between sparse codes as t changes. Sparse codes obtained via the soft-thresholding scheme are much stable than those obtained via the hard-thresholding scheme under different t.

scheme achieves smaller standard deviation between sparse codes as t varies as shown in Fig. 1 (b). In the recognition point of view, Fig. 2 shows the recognition accuracy of MR-Soft and MR-Hard with the HOG, HOF, and HOGHOF features respectively under different t. The accuracy of the MR-soft method is much better than that of the MR-Hard method. Therefore, we can conclude that soft-thresholding technique such as MR-Soft is much stable and robust than hard-thresholding technique such as MR-Hard.

Energy Discussion: The sparsity constraint E in Eq. 6 needs to be specified a value, and the value may vary with different datasets. Instead, the self-ratio constraint e in Eq. 8 aims to keep the certain ratio of the marginal regression coefficients in terms of L_1-norm energy. For example, the self-ratio $e = 5\%$ means that, for each feature, only 5% of its marginal regression coefficients in terms of the L_1 norm can be kept and the rest will be set to zero. Thus, we can obtain sparse codes more easily by assigning a desired ratio. Fig. 3 shows the recognition accuracy between different values of the self-ratio constraint e. The results show that we obtain around 90% recognition accuracy by setting the self-ratio e from 5% to 35%, and obtain poor recognition accuracy (below 10%) by setting e from 75% to 100%. Therefore, we can conclude that sparse representation helps signal interpretation.

Speed Comparison: In this subsection, we examine the feature quantization speed using different sparsity estimation methods under three different feature descriptors, HOG, HOF and HOGHOF, respectively. The dictionary sizes are set from 1000 to 4000. It is worth mentioning that all the methods except for LASSO-LAR were implemented in Matlab, and thus it would be unfair to compare them with the LASSO-LAR method together since C++ program of LASSO-LAR has a built-in speed advantage. Nevertheless, as shown in Table 3, the marginal regression related methods still take significantly less time than those of the lasso solutions. In addition, MR-SIS and MLC have

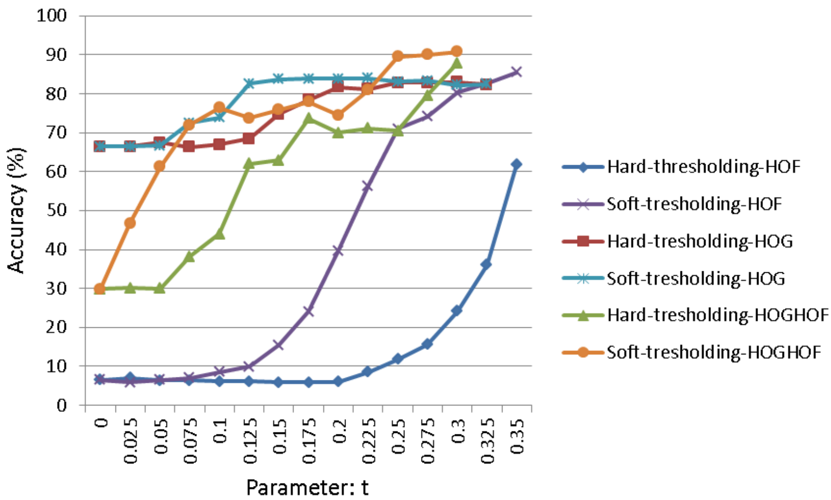

Fig. 2. Accuracy of soft and hard-thresholding with the HOG, HOF and HOGHOF features as t varies. The accuracy of MR-soft is much better than that of MR-Hard in terms of recognition accuracy under three different features.

comparable quantization speed but perform slower than MR-Hard and MR-soft since the former two consider the L_1-norm energy constraint, and the latter two do not.

Feature Quantization Error: Except for the speed comparison, Table 3 also shows the quantization errors. As shown, LASSO-FS and LASSO-LAR achieve the lowest quantization error. For the marginal regression related methods, the energy-based methods (MR-SIS and MLC) achieve lower quantization error than those without L_1-norm energy (MR-Hard and MR-Soft) constraint. In addition, the soft-thresholding-like methods achieve lower quantization error than the hard-thresholding-like methods pairwisely, i.e. MR-Soft against MR-Hard, and MLC against MR-SIS.

Recognition Accuracy: Table 4 shows the recognition accuracy under three different features with the dictionary size 4000. In this experiment, we select the sparsity constraint E via cross validation, and compare the proposed MLC model to other marginal related methods: MR-Hard, MR-Soft and MR-SIS. As shown, the proposed MLC approach obtains the best results, and the MR-Hard method obtains the worst and performs poorly under the HOF feature. In addition, the energy-based methods (MLC and MR-SIS) performs better than those without the energy constraint (MR-Hard and MR-Soft), and the soft-thresholding-type methods have better performance against the hard-thresholding-type methods pairwisely (MLC against MR-SIS and MR-Soft against MR-Hard).

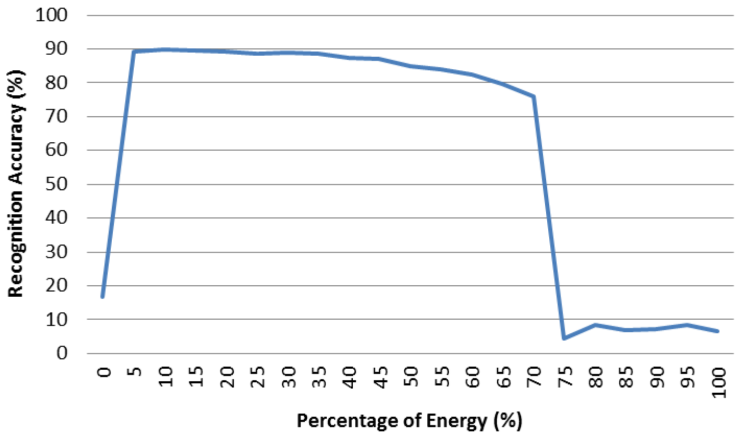

Fig. 3. Recognition accuracy under different self-ratios

Table 3. Comparison with the existing methods for the KTH dataset. The results are shown in terms of speed and quantization error. Notice that here we random sampled $1,258$ features from the training set. The actual numbers of the local features in the KTH dataset are $261,946$ for the training set and $136,219$ for the test set.

	time (sec)				Reconstruction error			
Method \ Dictionary size	1000	2000	3000	4000	1000	2000	3000	4000
HOG-LASSO-FS [20]	8.08	9.71	18.83	19.77	0.04	0.03	0.03	0.03
HOG-LASSO-LAR [4]	1.69	3.61	5.81	7.67	0.04	0.03	0.03	0.03
HOG-MR-Hard	0.03	0.05	0.08	0.11	781.94	3438.42	8079.61	14863.70
HOG-MR-Soft	0.03	0.06	0.09	0.12	68.40	319.07	764.57	1423.95
HOG-MR-SIS [6]	0.54	1.01	1.48	2.03	2.92	2.90	12.13	4.68
HOG-MLC	0.47	0.92	1.44	1.93	2.20	2.43	2.47	2.57
HOF-LASSO-FS [20]	5.47	6.95	13.89	14.96	0.03	0.03	0.03	0.02
HOF-LASSO-LAR [4]	1.15	2.64	4.61	6.07	0.03	0.03	0.03	0.02
HOF-MR-Hard	0.03	0.06	0.09	0.12	7317.76	32218.46	77688.51	142454.52
HOF-MR-Soft	0.03	0.07	0.10	0.14	1258.00	5818.47	14413.56	26806.27
HOF-MR-SIS [6]	0.65	1.19	1.80	2.34	17.02	29.80	39.24	65.60
HOF-MLC	0.64	1.21	1.86	2.48	3.35	3.27	3.12	3.09
HOGHOF-LASSO-FS [20]	7.73	9.34	17.83	18.46	0.07	0.06	0.05	0.05
HOGHOF-LASSO-LAR [4]	1.60	3.52	5.79	7.78	0.07	0.06	0.05	0.05
HOGHOF-MR-Hard	0.03	0.06	0.09	0.12	595.99	2366.80	5247.93	9215.52
HOGHOF-MR-Soft	0.03	0.06	0.10	0.14	47.14	192.52	431.67	765.61
HOGHOF-MR-SIS [6]	0.52	0.97	1.45	1.93	2.90	2.95	3.03	2.91
HOGHOF-MLC	0.44	0.88	1.34	1.79	2.18	2.33	2.43	2.47

Table 4. Comparison with the existing methods for the KTH dataset

Feature-Method	KTH
HOG-MR-Hard	82.27
HOG-MR-Soft	82.39
HOG-MR-SIS [6]	84.01
HOG-MLC	85.98
HOF-MR-Hard	61.76
HOF-MR-Soft	85.52
HOF-MR-SIS [6]	89.11
HOF-MLC	92.47
HOGHOF-MR-Hard	87.83
HOGHOF-MR-Soft	90.73
HOGHOF-MR-SIS [6]	90.03
HOGHOF-MLC	92.35

Summary: While the lasso-like models (LASSO-FS and LASSO-LAR) achieve the lowest quantization error, they take more time to obtain sparse codes. Instead, the marginal regression models without energy constraint (MR-Hard and MR-Soft) can reach the fastest speed to generate sparse codes; however, they also gain the highest quantization errors. While MR-SIS can compete with our model in terms of the quantization error, we reach better recognition accuracy than that of MR-SIS as shown in Table 4 because the proposed approach is a soft-thresholding method which is more stable and robust. There are two main differences between the MR-SIS model and the proposed MLC model: 1) While the idea of MR-SIS is frequently used of the hard-thresholding technique in the applications, it has no theoretical support. We integrates marginal regression and the lasso model with a theoretical support which is a soft-thresholding scheme. 2) We estimate the regularization parameter for each individual input feature and gain the lower quantization error than that of the MR-SIS model.

5.3 Activity-Based Human Identification

The activity-based human identification task aims to recognize the identity of a person based on his/her activities. We performed the experiment on two publicly available databases, weizmann [13] and MOBISERV-AIIA [16], respectively. In the dataset, each clip contains a person performing one activity. For each image frame, we extract human body silhouette which is similar to the way in [25,12,16,22,15]. An initial dictionary is learned via the online learning method proposed in [23] to speed up the learning process. The weizmann dataset contains 215 clips with 9 persons performing 9 different activities. The bend activity is excluded due to lack of samples. We randomly select one clip per action per class as training data, and use the remaining as test data. The MOBISERV-AIIA dataset contains 12 persons performing eating and drinking activities with 2 clothing scenarios in four different days. We adopt 2 kinds of the activities, drinking with a cup and eating with a fork, with 776 clips in total. We randomly choose one-day sequences as test samples and use the rest as the training samples.

The parameters are determined via cross validation. Table 5 shows the recognition accuracy based on the 10% self-ratio sparsity constraint. In this application, the soft-thresholding-type methods (MLC and MR-Soft) work better than the hard-thresholding-type methods (MR-Hard and MR-SIS). In addition, MLC outperforms MR-Soft.

Table 5. Comparison with the existing methods for the weizmann and the MOBISERV-AIIA datasets

backslashboxMethodDataset	weizmann	MOBISERV-AIIA
MR-Hard	69.40	62.69
MR-Soft	79.85	70.47
MR-SIS [6]	77.61	67.88
MLC	85.82	73.06

6 Conclusion

In this paper, we have proposed a novel feature quantization approach for sparsity estimation by exploiting the advantages of the marginal regression and the lasso simultaneously to provide more efficient and effective solutions. The proposed approach has been evaluated on three visual applications: image classification, action recognition and activity-based human identification. Experimental results have shown that our method can achieve excellent performance in terms of speed, quantization error and recognition accuracy. All these sufficiently demonstrate the efficacy of our proposed methods.

Acknowledgment. This work is supported by the research grant for the Human Cyber Security Systems (HCSS) Program at the Advanced Digital Sciences Center from the Agency for Science, Technology and Research of Singapore.

References

1. Balasubramanian, K., Yu, K., Lebanon, G.: smooth sparse coding via marginal regression for learning sparse representations. In: ICML (2013)
2. Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2(3), 27:1–27:27 (2011)
3. Donoho, D.L.: For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. Comm. Pure Appl. Math. 59, 797–829 (2004)
4. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Annals of Statistics (2004)
5. Elhamifar, E., Vidal, R.: Robust classification using structured sparse representation. In: CVPR (2011)
6. Fan, J., Lv, J.: Sure independence screening for ultrahigh dimensional feature space. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70(5), 849–911 (2008)
7. Friedman, J., Hastie, T., Hofling, H., Tibshirani, R.: Pathwise coordinate optimization (2007)
8. Gao, S., Tsang, I., Chia, L.: Laplacian sparse coding, hypergraph laplacian sparse coding, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence(2012)

9. Gao, S., Tsang, I.W.H., Chia, L.T., Zhao, P.: Local features are not lonely - laplacian sparse coding for image classification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (2010)
10. van Gemert, J.C., Geusebroek, J.-M., Veenman, C.J., Smeulders, A.W.M.: Kernel codebooks for scene categorization. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 696–709. Springer, Heidelberg (2008)
11. Genovese, C.R., Jin, J., Wasserman, L., Yao, Z.: A comparison of the lasso and marginal regression. J. Mach. Learn. Res. 13(1), 2107–2143 (2012)
12. Gkalelis, N., Tefas, A., Pitas, I.: Human identification from human movements. In: ICIP, pp. 2585–2588 (2009)
13. Gorelick, L., Blank, M., Shechtman, E., Irani, M., Basri, R.: Action as space-time shapes. PAMI 29(12), 224–2253 (2007)
14. Hull, J.: A database for handwritten text recognition research. PAMI 16(5), 550–554 (1994)
15. Hung, T.-Y., Lu, J., Tan, Y.-P.: Graph-based sparse coding and embedding for activity-based human identification. In: ICME (2013)
16. Iosifidis, A., Tefas, A., Pitas, I.: Activity based person identification using fuzzy representation and discriminant learning. TIFS 7(2), 530–542 (2012)
17. Laptev, I.: On space-time interest points. International Journal of Computer Vision 64(2-3), 107–123 (2005)
18. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: CVPR (2008)
19. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: CVPR, pp. 2169–2178 (2006)
20. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: Schölkopf, B., Platt, J., Hoffman, T. (eds.) Advances in Neural Information Processing Systems, vol. 19, pp. 801–808. MIT Press, Cambridge (2007)
21. Li, F.F., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: CVPR, pp. 524–531 (2005)
22. Lu, J., Hu, J., Zhou, X., Shang, Y.: Activity-based human identification using sparse coding and discriminative metric learning. In: ACM Multimedia (2012)
23. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res. 11, 19–60 (2010)
24. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. IJCV 42(3), 145–175 (2001)
25. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The humanID gait challenge problem: data sets, performance, and analysis. PAMI 27(2), 162–177 (2005)
26. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: A local svm approach. In: ICPR, pp. 32–36 (2004)
27. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58(1), 267–288 (1996)
28. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2), 210–227 (2009)
29. Wu, T., Lange, K.: Coordinate descent algorithms for lasso penalized regression. Annals of Applied Statistics (2008)
30. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 1794–1801 (June 2009)
31. Zheng, M., Bu, J., Chen, C., Wang, C., Zhang, L., Qiu, G., Cai, D.: Graph regularized sparse coding for image representation. IEEE Transactions on Image Processing 20(5), 1327–1336 (2011)