Volume 6, Issue 12, December 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Image Classification by Linear Matching Technique with Sparse Coding using SVM Classifier

Aditya Kumar Kaushik
M. Tech. Scholar, CSE, VNS College,
Bhopal, India

Amit Nandanwar
AP CSE, VNS College,
Bhopal, India

Abstract— Image Classification is a vital errand inside the field of PC vision. Image classification alludes to the marking of images into one of various predefined classifications. Classification incorporates image sensors, image pre-handling, protest discovery, question division, include extraction and question order. Numerous arrangement procedures have been produced for image classification. As of late direct spatial coordinating (LSM) part has been very fruitful in image classification. In spite of its prevalence, these nonlinear have a multifaceted nature $O(n^2 \sim n^3)$ in preparing and O(n) in testing, where n is the preparation estimate, suggesting that it is nontrivial to scale-up the calculations to handle more than a large number of preparing images. In this paper we build up an expansion of the LSM strategy, by summing up vector quantization to inadequate coding took after by multi-scale spatial max pooling, and ace represent a direct LSM bit in light of SIFT meager codes. This new approach astoundingly lessens the many-sided quality of SVMs to O(n) in preparing and a consistent in testing with more exactness than past techniques. In various image arrangement tests, we discover order precision, the proposed straight SM in light of meager coding of SIFT descriptors dependably altogether outflanks the LSM portion on histograms, and is stunningly better than the nonlinear LSM pieces.

Keywords— Image Classification, Linear Match, SVM Classifier, Image Processing.

I. INTRODUCTION

Classification between the items is simple assignment for people yet it has ended up being a mind boggling issue for machines. The raise of high-limit PCs, the accessibility of high caliber and low-estimated camcorders, and the expanding requirement for programmed video examination has created an enthusiasm for question arrangement calculations. A basic classification framework comprises of a camera settled high over the intrigued zone, where images are caught and thus handled. Order incorporates image sensors, image pre-handling, protest identification, question division, highlight extraction and question classification. Arrangement framework comprises of database that contains predefined designs that contrasts and recognized protest order into legitimate classification. Image arrangement is a critical and testing errand in different application areas, including biomedical imaging, biometry, video-observation, vehicle route, mechanical visual examination, robot route, and remote detecting.

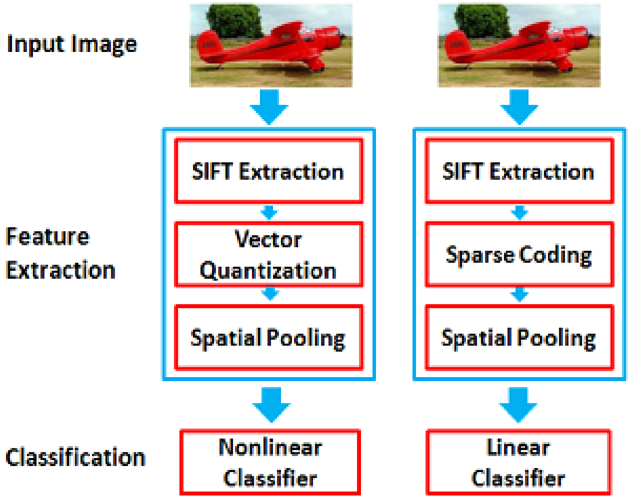
Classification process comprises of taking after strides:

- Pre-preparing air adjustment, clamor evacuation, image change, primary segment examination and so on.
- Discovery and extraction of a protest Detection incorporates identification of position and different qualities of moving article image got from camera.
- What's more, in extraction, from the distinguished protest assessing the direction of the question in the image plane.
- Preparing: Selection of the specific quality which best portrays the example.
- Classification of the protest Object order step classifies recognized items into predefined classes by utilizing appropriate technique that contrasts the image designs and the objective examples.

Lately the pack of-elements (BoF) display has been to a great degree famous in image arrangement. The strategy regards a image as an accumulation of unordered appearance descriptors extricated from nearby fixes, quantizes them into discrete "visual words", and after that processes a minimized histogram representation for semantic image classification, e.g. protest acknowledgment or scene classification.

The BoF approach disposes of the spatial request of neighbourhood descriptors, which extremely constrains the expressive force of the image representation. By defeating this issue, one specific augmentation of the BoF show, called linier spatial coordinating (LSM) [1], has made an exceptional accomplishment on a scope of image classification benchmarks like Caltech-101 [2], and was the significant part of the cutting edge frameworks, e.g., [3]. The strategy segments a image into 21 fragments in various scales l = 0; 1; 2, registers the BoF histogram inside each of the 21 portions, lastly connects every one of the histograms to shape a vector representation of the image. On the off chance that where just the scale l = 0 is utilized, LSM diminishes to BoF.

Individuals have exactly found that, to acquire great exhibitions, both BoF and LSM must be connected together with a specific kind of nonlinear Mercer portions, e.g. the crossing point bit or the Chi-square piece. In like manner, the nonlinear SVM needs to pay a computational many-sided quality O(n3) and a memory multifaceted nature O(n2) in the preparation stage, where n is the preparation estimate. Besides, since the quantity of bolster vectors develops directly with n, the computational unpredictability in testing is O(n). This adaptability infers a serious confinement — it is nontrivial to apply them to genuine applications, whose preparation size is commonly a long ways past thousands.



(a)Non-Linear LSM classification (b) Linear Proposed method classification

Fig. 1 Schematic correlation of the first nonlinear LSM with our proposed strategy. The hidden spatial pooling capacity for nonlinear LSM is averaging, while the spatial pooling capacity in proposed technique is max pooling.

In this paper, we propose an expansion of the LSM approach, which registers a spatial-pyramid image representation in light of inadequate codes (SC) of SIFT components, rather than the K-implies vector quantization (VQ) in the conventional LSM. The approach is actually determined by unwinding the prohibitive cardinality imperative of VQ. Besides, not at all like the first LSM that performs spatial pooling by figuring histograms, our proposed approach, utilizes max spatial pooling that is more hearty to nearby spatial interpretations and more organic conceivable [12]. The new image representation catches more striking properties of visual examples, and ends up working shockingly well with direct classifiers. Our approach utilizing basic direct SVMs significantly lessens the preparation unpredictability to O(n), and gets a steady multifaceted nature in testing, while as yet accomplishing a shockingly better arrangement precision in correlation with the conventional nonlinear LSM approach. Schematic examination between the first LSM with Proposed strategy is appeared in Fig. 1

II. RELATED WORK

Given a image, elements, for example, SIFT [13], HOG [14] and SURF [15], can be thickly removed and encoded with a code-book developed utilizing K-implies grouping. As of late, a wide range of highlight coding strategies have been proposed including hard-task coding (HC) [16], delicate task coding (SC) [17], limited delicate task coding (LSC) [18], meager coding (SCSPM) [19], region compelled straight coding (LLC) [20], Laplacian inadequate coding (LScSPM) [21], striking coding (SC) [22], and area obliged and spatially regularized coding (LCSRC) [23]. Subsequent to registering codes for neighborhood highlights, they should be pooled together to shape parallel measured element vectors each speaking to one image in a dataset.

Mainstream pooling techniques incorporate normal pooling (e.g. histogram) and max-pooling [19]. To incorporate the spatial design of neighborhood elements in a image, Spatial Pyramid Matching (SPM) [16] is typically performed to get a image level representation that can be utilized to separate diverse classifications of items, scenes, or activities. Utilizing this BoW representation, images can be ordered utilizing a plenty of discriminative models, for example, SVM or Boosting.

Late work demonstrates that given a visual codebook, the strategy for encoding nearby components has critical effect on classification execution. The most punctual strategy is hard-task coding (vector quantization) [16], a voting plan that is basic yet exceedingly touchy to the choice of codebook. A more powerful voting methodology is delicate task coding [17], which allocates a code coefficient for a specific neighborhood highlight to each visual word as per their pairwise remove. To enhance hard and delicate task coding, sparsity is authorized on neighborhood highlight codes by means of meager learning strategies [19].

Nonetheless, inadequate coding is tedious and for the most part prompts to non-reliable codes [20, 21], i.e. nearby components with comparable descriptors have a tendency to have distinctive inadequate codes. To reduce irregularity, creators in [24] present another coding property, called territory, which energizes that visual words used to speak to a neighborhood highlight be like the component's descriptor itself. This is typically guaranteed by building a component's codebook from its closest neighbors in the all inclusive codebook. Truth be told, a few usage of region have been proposed in [20, 18, 22], where every descriptor is coded on privately chosen bases. The work in [22] rebrands the territory property as codebook 'saliency'. Be that as it may, all the previously mentioned coding plans encode lo-cal includes autonomously.

III. LINEAR LSM USING SIFT SPARSE CODES

A. *Encoding SIFT: From VQ to SC*

Let X be a set of SIFT appearance descriptors in a D-dimensional feature space, i.e. X = [x1; . . . ; xM]>2RMD. The vector quantization (VQ) method applies the K-means clustering algorithm to solve the following problem

$$\min_{\mathbf{V}} \sum_{m=1}^{M} \min_{k=1\ldots K} \|\mathbf{x}_m - \mathbf{v}_k\|^2 \quad (1)$$

where V = [v1; . . . ; vK]^T are the K cluster centers to be found, called codebook, and k k denotes the L2-norm of vectors. The optimization problem can be re-formulated into a matrix factorization problem with cluster member-ship indicators U = [u1; . . . ; uM]^T

$$\min_{\mathbf{U},\mathbf{V}} \sum_{m=1}^{M} \|\mathbf{x}_m - \mathbf{u}_m\mathbf{V}\|^2 \quad (2)$$

subject to $Card(\mathbf{u}_m) = 1, |\mathbf{u}_m| = 1, \mathbf{u}_m \succeq 0, \forall m$

where Card(u_m) = 1 is a cardinality constraint, meaning that only one element of um is nonzero, um 0 means that all the elements of um are nonnegative, and jumj is the L1-norm of um, the summation of the absolute value of each element in um. After the optimization, the index of the only nonzero element in um indicates which cluster the vector x_m belongs to. In the training phase of VQ, the optimization Eq. (2) is solved with respect to both U and V. In the coding phase, the learned V will be applied for a new set of X and Eq. (2) will be solved with respect to U only

The constraint Card(u_m) = 1 may be too restrictive, giving rise to often a coarse reconstruction of X. We can relax the constraint by instead putting a L1-norm regularization on um, which enforces um to have a small number of nonzero elements. Then the VQ formulation is turned into another problem known as sparse coding (SC):

$$\min_{\mathbf{U},\mathbf{V}} \sum_{m=1}^{M} \|\mathbf{x}_m - \mathbf{u}_m\mathbf{V}\|^2 + \lambda|\mathbf{u}_m| \quad (3)$$

subject to $\|\mathbf{v}_k\| \leq 1, \quad \forall k = 1, 2, \ldots, K$

Similar to VQ, SC has a training phase and a coding phase. First, a descriptor set X from a random collection of image patches is used to solve Eq. (3) with respect to U and V, where V is retained as the codebook; In the coding phase, for each image represented as a descriptor set X, the SC codes are obtained by optimizing Eq. (3) with respect to U only.

We choose SC to derive image representations because it has a number of attractive properties. First, compared with the VQ coding, SC coding can achieve a much lower reconstruction error due to the less restrictive constraint; Second, sparsity allows the representation to be specialize, and to capture salient properties of images; Third, research in image statistics clearly reveals that image patches are sparse signals.

B. *Linear spatial matching*

For any image represented by a set of descriptors, we can compute a single feature vector based on some statistics of the descriptors' codes. For example, if U is obtained via Eq. (2), a popular choice is to compute the histogram

$$\mathbf{z} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{u}_m \quad (4)$$

The sack of-words way to deal with image arrangement processes such a histogram z for every image I spoke to by an unordered arrangement of nearby descriptors. In the more advanced LSM approach, the image's spatial pyramid histogram representation z is a link of neighborhood histograms in different parcels of various scales. After standardization z can be viewed as again a histogram. Let z_i signify the histogram representation for image I_i. For a twofold image order issue, a SVM intends to take in a choice capacity

$$f(\mathbf{z}) = \sum_{i=1}^{n} \alpha_i \kappa(\mathbf{z}, \mathbf{z}_i) + b \quad (5)$$

where $\{(\mathbf{z}_i, y_i)\}_{i=1}^{n}$ is the training set, and $y_i \in \{-1, +1\}$ indicates labels. For a test image spoke to by z, if f (z) > 0 then the image is delegated positive, generally as negative. In principle K(.,.) can be any sensible Mercer bit work, however practically speaking the crossing point part and Chi-square bit have been found the most appropriate on histogram representations. Our examination demonstrates that straight portion on histograms prompts to dependably significantly more terrible outcomes, incompletely because of the high quantization mistake of VQ. Be that as it may, utilizing these two nonlinear parts, the SVM needs to pay a high preparing cost, i.e. O(n^3) in calculation, and O(n^2) away (for the n bit framework). This implies it is hard proportional up the calculation to the situation where n is more than several thousands. Moreover, as the quantity of bolster vectors scales directly to the preparation estimate, the testing expense is O(n). In this paper we advocate an approach of utilizing direct SVMs based SC of SIFT. Give U a chance to be the after effect of applying the scanty coding to a descriptor set X, expecting the codebook V to be pre-learned and settled, we figure the accompanying image include by a pre-picked pooling capacity.

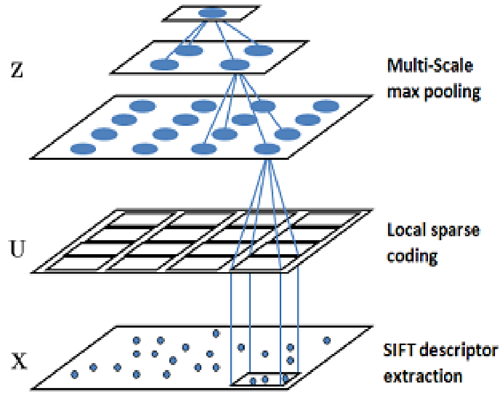


Fig. 2 The illustration architecture of our algorithm based on sparse coding. Sparse coding measures the responses of each local descriptor to the dictionary's "visual elements". These responses are pooled across different spatial locations over different spatial scales.

$$\mathbf{z} = \mathcal{F}(\mathbf{U}) \quad (6)$$

where the pooling capacity F is characterized on every segment of U. Review that every segment of U relates to the reactions of all the nearby descriptors to one particular thing in word reference V. Accordingly, unique pooling capacities develop diverse image measurements. For instance, in 4, the fundamental pooling capacity is characterized as the averaging capacity, yielding the histogram include. In this work, we de-fined the pooling capacity F as a maximum pooling capacity on unquestionably the scanty codes

$$z_j = \max\{|u_{1j}|, |u_{2j}|, ..., |u_{Mj}|\} \quad (7)$$

where z_j is the j^{th} element of z, u_{ij} is the matrix element at i^{th} row and j^{th} column of U, and M is the number of local descriptors in the region. This maximum pooling methodology is settled by biophysical prove in visual cortex (V1) [12] and is experimentally legitimized by numerous calculations connected to image classification. For our situation, we additionally find that maximum pooling outflanks other option pooling techniques. Like the development of histograms in LSM, we do max pooling Eq. (7) on a spatial pyramid developed for a image. By max pooling crosswise over various areas and over various spatial sizes of the image, the pooled highlight is more powerful to neighbourhood changes than mean measurements in histogram. Fig. 2 represents the entire structure of our calculation in view of inadequate coding. The pooled highlights from different areas and scales are then connected to shape a spatial pyramid representation of the image.

IV. PROPOSED WORK

A. Sparse Coding

The optimization problem Eq. (3) is convex in V (with U fixed) and convex in U (with V fixed), but not in both simultaneously. The conventional way for such a problem is to solve it iteratively by alternatingly optimizing over V or U while fixing the other. Fixing V, the optimization can be solved by optimizing over each coefficient um individually:

$$\min_{\mathbf{u}_m} \|\mathbf{x}_m - \mathbf{u}_m\mathbf{V}\|_2^2 + \lambda|\mathbf{u}_m| \quad (8)$$

This is essentially a linear regression problem with L1 norm regularization on the coefficients, well known as Lasso in the Statistical literature. The optimization can be solved very efficiently by algorithms such as the recently proposed feature-sign search algorithm. [4]. Fixing U, the problem reduces to a least square problem with quadratic constraints:

$$\min_{\mathbf{V}} \quad \|\mathbf{X} - \mathbf{U}\mathbf{V}\|_F^2 \quad (9)$$
$$s.t. \quad \|\mathbf{v}_k\| \leq 1, \quad \forall k = 1, 2, ..., K.$$

The optimization can be done efficiently by the Lagrange dual as used in [4].

In our experiments, we use 50,000 SIFT descriptors extracted from random patches to train the codebook, by iterating the steps Eq. (8) and Eq. (9). Once we get the code-book V in this off-line training, we can do on-line sparse coding efficiently as in Eq. (8) on each descriptor of an image.

B. 4.2. Multi-class Linear SVM

We introduce a simple implementation of linear SVMs that was used in our experiments. Given the training data $\{(\mathbf{z}_i, y_i)\}_{i=1}^n, y_i \in \mathcal{Y} = \{1, \ldots, L\}$, a linear SVM aims to learn L linear functions L $\{\mathbf{w}_c^\top \mathbf{z} | c \in \mathcal{Y}\}$, such that, for a test datum z, its class label is predicted by

$$y = \max_{c \in \mathcal{Y}} \mathbf{w}_c^\top \mathbf{z} \quad (10)$$

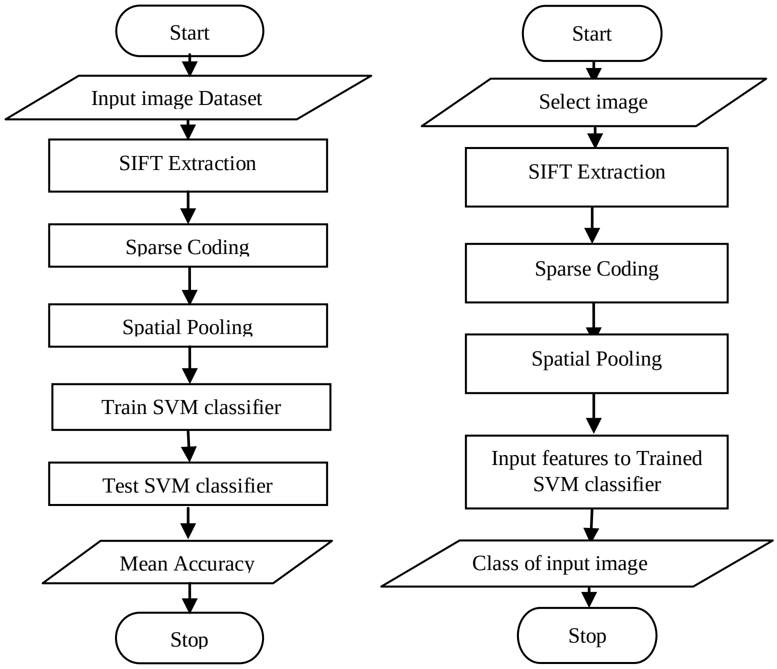
We take a one-against-all strategy to train L binary linear SVMs, each solving the following unconstraint convex optimization problem

$$\min_{\mathbf{w}_c} \left\{ J(\mathbf{w}_c) = \|\mathbf{w}_c\|^2 + C \sum_{i=1}^{n} \ell(\mathbf{w}_c; y_i^c, \mathbf{z}_i) \right\} \quad (11)$$

where $y_i^c = 1$ if yi = c, otherwise $y_i^c = -1$, and $\ell(\mathbf{w}_c; y_i^c, \mathbf{z}_i)$ is a hinge loss function. The standard hinge loss function is not differentiable everywhere, which hampers the use of gradient-based optimization methods. Here we adopt a differentiable quadratic hinge loss

$$\ell(\mathbf{w}_c; y_i^c, \mathbf{z}_i) = \left[\max\left(0, \mathbf{w}_c^\top \mathbf{z} \cdot y_i^c - 1\right)\right]^2$$

such that the training can be easily done with simple gradient-based optimization methods. In our work we used LBFGS. Other choices like conjugate gradient are also applicable. The only implementation on our side is providing the cost J (w) and the gradient ∂J(w)= ∂w. The computation linearly scans over the training examples and thus has the linear complexity O(n). In our experiment, the SVM training on about 200,000 examples with 5376-dimensional features was usually finished in 5 minutes.



(a) Proposed work SVM train and testing (b) Checking class for selected image using network.

Fig. 1 Flowchart of proposed method

V. EXPERIMENTS AND RESULTS

In the investigations, we executed and assessed three classes of LSM techniques on four various datasets: Caltech 101 [2]. The three techniques are

1. KLSM: the well known nonlinear piece LSM that utilizations spatial-pyramid histograms and Chi-square portions;
2. LSPM: the straightforward direct SPM that utilizations straight part on spatial-pyramid histograms;
3. Proposed strategy: the straight LSM that utilizations direct part on spatial-pyramid pooling of SIFT inadequate codes,

Other than our own particular executions, we additionally cite a few outcomes specifically from the writing, particularly those of KLSM from [1] and [5]. We take note of that occasionally we couldn't re-deliver their outcomes, to a great extent because of unpretentious building de-tails, e.g. the method for managing high-differentiation and low-differentiate patches. It in this manner bodes well to look at our own executions, since they depended on the very same arrangement of descriptors.

Our executions utilized a solitary descriptor sort, the prevalent SIFT descriptor, 3 as in [1, 6, 7]. The SIFT descriptors extricated from 16×16 pixel patches were thickly inspected from every image on a matrix with step measure 8 pixels. The images were all pre-prepared into dark scale. To prepare the codebooks, we utilized standard K-implies classification for KLSM and LSM, and the meager coding plan for our proposed technique calculation. For every one of the investigations we settled the codebook measure as 512 for LSM and 1024 for Proposed strategy, to accomplish ideal exhibitions for both. For preparing the direct classifiers, we utilized our executed SVM.

Taking after the regular benchmarking strategies, we rehash the trial procedure by 10 times with various arbitrary chose preparing and testing images to get dependable outcomes. The normal of per-class acknowledgment rates were recorded for every run. Furthermore, we report our last outcomes by the mean and standard deviation of the acknowledgment rates.

A. Patch Size

In our tests, we just utilized one fix size to separate SIFT descriptors, in particular, 16×16 pixels as in LSM [1]. In NBNN[6], they utilized four fix scales to extricate the descriptors keeping in mind the end goal to help their execution. In our investigations, we didn't watch any significant upgrades by pooling over different fix scales, most likely in light of the fact that maximum pooling over inadequate codes can catch the striking properties of neighborhood districts that are superfluous to the size of nearby fixes.

B. Codebook Size

We likewise explored the impacts of codebook sizes on these LSM calculations. Instinctively, if the codebook size is too little, the histogram highlight looses discriminant control; if the codebook size is too substantial, the histograms from a similar class of images will never coordinate. In Lazebnik et al's. work, they utilized two codebook sizes 200 and 400 and reported that there was little distinction. In our tests on proposed technique we attempted three sizes: 256, 512 and 1024. As appeared in Table 1, the execution for LSM builds at first and afterward diminishes as the codebook estimate becomes advance. The execution for proposed strategy keeps on expanding when the codebook scrutinizes goes to 1024.

Table I impacts of codebook size on Proposed technique and LSM separately on Caltech 101 dataset.

	Codebook size	**256**	**512**	**1024**
30 train	Proposed method	68.26	71.20	73.20
	LSM	57.42	58.81	58.56
15 train	Proposed method	61.97	63.23	69.70
	LSM	51.84	53.23	51.74

C. Sparse Coding Parameter

There is without one parameter λ as in Eq. (8) we have to decide when we do inadequate coding on every element vector. λ Enforces the sparsity of the arrangement; the greater λ is, more inadequate the arrangement will be. Observationally, we found that keeping the sparsity to be around 10% yields great outcomes. For every one of our tests, we essentially settled λ to be 0.3~0.4 and the mean number of backings (non-zero coefficients) is around 10.

D. Caltech-101 Dataset

The Caltech-101 dataset contains 101 classes (counting creatures, vehicles, blooms, and so on.) with high shape changeability [11]. The quantity of images per class shifts from 31 to 800. Most images are medium determination, i.e. around 300×300 pixels. We took after the regular examination setup for Caltech-101, preparing on 15 and 30 images for every classification and testing on the rest. Itemized correlation results are appeared in Table 2. As appeared, our scanty coding plan beats direct SM by more than 14 percent, and even outflanks the nonlinear SPM [1] by a vast edge (around 11 percent for 15 preparing and 9 percent for 30 preparing per class). One work needs to say is the Kernel Code-books [8], where the writer appointed every descriptor into various containers rather than hard task. This plan for the most part enhances their standard LSM by 5~6 percent. Nonetheless, their strategy is still in view of nonlinear parts.

Table III Classification rate (%) comparison on Caltech-101

Algorithms	**15 training**	**30 training**
Zhang et al. [28]	59.10±0.60	66.20±0.50
KSPM [1]	56.40	64.40±0.80
NBNN [6]	65.00±1.14	70.40
ML+CORR [7]	61.00	69.60
KC [8]	–	64.14±1.18
Proposed method	**67.0±0.45**	**73.2±0.54**

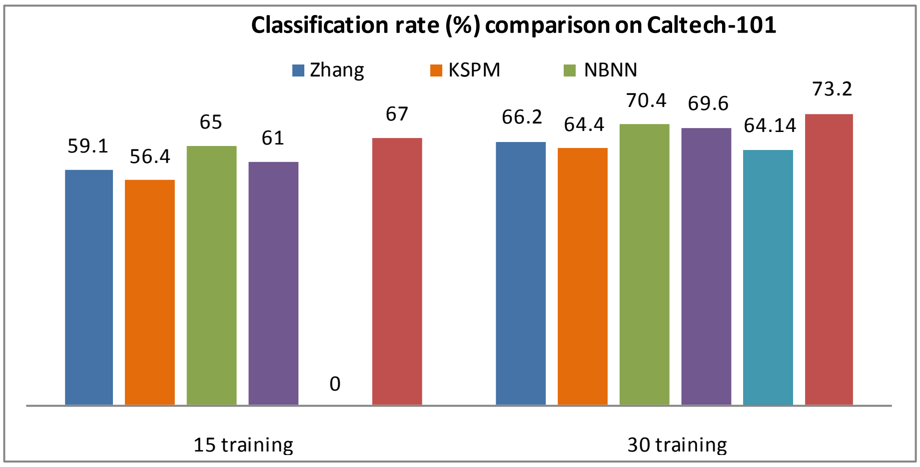


Fig. 4 Proposed method comparison with previous methods.

VI. CONCLUSION AND FUTURE WORK

In this paper we proposed a spatial methodology in view of SIFT meager codes for image classification. The technique utilizes specific meager coding rather than customary vector quantization to concentrate remarkable properties of appearance descriptors of nearby image patches. Promote more, rather than averaging pooling in the histogram, scanty coding empowers us to work nearby max pooling on various spatial scales to fuse interpretation and scale invariance. The most promising consequence of this paper is the acquired image representation works shockingly well with basic direct SVMs, which significantly enhance the versatility of preparing and the speed of testing, and even enhances the classification precision. Our investigations on an assortment of image classification undertakings exhibited the viability of this approach. Since the nonlinear LSM in view of vector quantization is extremely prominent in top-performing image order frameworks, we trust the recommended direct LSM will significantly enhance best in class by permitting utilizing much bigger arrangements of preparing information.

As a sign from our work, the scanty codes of SIFT components may serve as a superior nearby appearance descriptor for general image handling assignments. Additionally research of this in observational review and hypothetical comprehension is an intriguing heading. Another issue is the proficiency of en-coding. As of now encoding the SIFT descriptors of each Caltech image takes around one second in normal. A late work demonstrates that inadequate coding can be significantly quickened by utilizing a nourish forward system [9]. It will enthusiasm to attempt such strategies to make our approach speedier. Besides, the precision could be further enhanced by taking in the codebook in a directed manner, as proposed by another late work [10].

REFERENCES

[1] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features. Spatial pyramid matching for recognizing natural scene categories. In CVPR, 2006.
[2] F.-F. Li, R. Fergus, and P. Perona. Learning generative visual models from few training examples. an incremental Bayesian approach tested on 101 object categories. In CVPR Workshop on Generative-Model Based Vision, 2004.
[3] A. Bosch, A. Zisserman, and X. Munoz. Image classification using random forests and ferns. In ICCV, 2007.
[4] H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In NIPS, 2006.
[5] Griffin, G. Holub, and P. AD. Perona. Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology, 2007.
[6] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest neighbor based image classification. In CVPR, 2014.
[7] P. Jain, B. Kullis, and K. Grauman. Fast image search for learned metrics. In CVPR, 2013.
[8] J. C. van Gemert, J.-M. Geusebroek, C. J. Veenman, and A. W. M. Smeulders. Kernel codebooks for scene categorization. In ECCV, 2014.
[9] K. Kavukcuoglu, M. Ranzato, and Y. LeCun. Fast infer-ence in sparse coding algorithms with applications to object recognition. Technical report, Computational and Biological Learning Lab, NYU, 2008.
[10] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Supervised dictionary learning. In NIPS, 2009.
[11] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples. An incremental bayesian approach tested on 101 object categories. In CVIU, 2007.
[12] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In CVPR, 2005
[13] D. G. Lowe. Distinctive image features from scaleinvariant keypoints. In IJCV,2004.
[14] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.
[15] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. In ECCV, 2006.
[16] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In CVPR, 2006.
[17] J. C. van Gemert, J.-M. Geusebroek, C. J. Veenman, and A. W. M. Smeulders. Kernel codebooks for scene categorization. In ECCV, 2008.
[18] L. Liu, L. Wang, and X. Liu. In defense of softassignment coding. In ICCV, 2011.
[19] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In CVPR, 2009.
[20] J.Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In CVPR, 2010.
[21] S. Gao, I. Tsang, L. Chia, and P. Zhao. Local features are not lonely – laplacian sparse coding for image classification. InCVPR, 2010.
[22] Y. Huang, K. Huang, Y. Yu, and T. Tan. Salient coding for image classification. In CVPR, 2011.
[23] A. Shabou and H. Le-Borgne. Locality-constrained and spatially regularized coding for scene categorization. In CVPR, 2012.
[24] K. Yu, T. Zhang, and Y. Gong. Nonlinear learning using local coordinate coding. In NIPS, 2009.