JOURNAL OF NORTHEASTERN UNIVERSITY

Volume 25 Issue 04, 2022 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

SEMANTIC SPARSE RECODING OF VISUAL CONTENT FOR IMAGE APPLICATIONS

Surender Reddy S [1], Dr.Neeraj Sharma[2], Dr.Y. Mohana Roopa[3]

[1]Research Scholar, Dept. of Computer Science and Engineering
Sri Satya Sai University of Technology and Medical Sciences,
Sehore Bhopal-Indore Road, Madhya Pradesh, India.

[2]Research Guide, Dept. of Computer Science and Engineering
Sri Satya Sai University of Technology and Medical Sciences,
Sehore Bhopal-Indore Road, Madhya Pradesh, India.

[3]Research Co-Guide, Professor. Dept. of Computer Science and Engineering
Institute of Aeronautical Engineering, Dundigal, Hyderabad

ABSTRACT

Sparse coding approximates the data sample as a sparse linear combination of several fundamental code words and then uses the sparse codes as new presentations. We study learning discriminative sparse codes using sparse coding in a semi-supervised manner with only a few labelled training samples. A novel semantic sparse recoding method is being developed to provide more descriptive and robust visual content representations for image annotation. Although it has been reported that the visual bag-of-words (BOW) representation achieves promising results in image annotation, its visual codebook is totally learned from low-level visual data using quantization approaches, leaving the so-called semantic gap unbridgeable. To address this difficult issue, we augment the original visual BOW representation by combining annotations from training photos with predicted annotations from test images. We learn the variable class labels for all the samples by exploiting the manifold structure spanned by the data set of labelled and unlabeled samples and the limitations imposed by the labels on the labelled samples. Additionally, to enhance the discriminatory ability of the learnt sparse codes, we assume that class labels may be predicted directly from the sparse codes using a linear classifier.

Keywords : Semantic chasm, sparse codes, annotation, and bag-of-words

INTRODUCTION

Sparse Coding (SC) has been a widely used and effective approach for representing data in a wide variety of applications, including pattern recognition, bioinformatics, and computer vision. Given a data sample and its feature vector, SC attempts to learn a codebook using some codeworks and approximates the data sample as the linear combination of the codewords. Because SC assumes that only a few codewords in the codebook are sufficient to represent the data sample, the combination coefficients should be sparse, with the majority of them being zeros and only a few being non-zero. The data sample's new representation might be its linear combination coefficients. The coefficient vector is frequently referred to as the sparse code due to its sparse nature. To solve a sparse code, one typically reduces the approximation error

Submitted: 21/10/2022 **Accepted**: 24/11/2022

between the codebook and the sparse code while also seeking the sparsity of the sparse code provided.

The purpose of this study is to discuss semantic sparse recoding of visual content in order to provide more descriptive and robust visual representations for image applications. The following section uses all of the photos' annotations (predicted by algorithms or given by users) as the high-level semantic information for such visual BOW representation refining. Additionally, to address the issue of noise introduced by improper quantization of visual features, we formulate the difficult task of visual BOW representation refinement as a sparse coding problem that can be efficiently tackled using a dimension reduction technique. Given that this type of sparse coding is primarily concerned with incorporating semantic information into the original visual codebook, it can be thought of as semantic sparse recoding (SSRC) of visual content for image applications.

The classic visual BOW representation, on the other hand, has two intrinsic disadvantages. To begin, its visual codebook is totally learned via quantization from low-level visual cues, leaving the so-called semantic gap unbridgeable. Second, because visual features are quantized incorrectly in the typical visual BOW representation, the noise issue becomes quite severe, resulting in significant performance reduction. Historically, little effort has been taken to alleviate these two disadvantages of the classic visual BOW representation simultaneously.

To demonstrate the efficacy of our semantic sparse recoding strategy, we apply the newly acquired visual BOW representation to automatic image annotation (AIA) and social image classification (SIC). For automatic image annotation, we first perform an AIA round to anticipate the annotations of test images in order to build the textual BOW representation for our semantic sparse recoding method. In comparison, for social image classification, all of the photos' annotations are contributed directly by users of photo-sharing websites (e.g., Flickr), and are frequently followed by a preprocessing step of keyword reduction. Subsequent experiments demonstrate that the revised new visual BOW representation created using our semantic sparse recoding method is much more descriptive and robust than the original visual BOW representation. Although our semantic sparse recoding method has been evaluated just in these two image applications, it can easily be adapted to additional difficult jobs when noisy image tags are provided initially. Finally, we describe the following distinguishing advantages of our semantic sparse recoding method:

• To the best of our knowledge, this is the first attempt to establish semantic sparse recoding of visual content in order to generate more descriptive and robust visual representations for image applications.

• Unlike existing visual codebook optimization methods, our semantic sparse recoding method can partially overcome the classic visual BOW representation's two shortcomings (i.e. semantic gap and inaccurate quantization).

• We demonstrate that our semantic sparse recoding strategy significantly improves automatic image annotation and social image classification, which is even more impressive given that we do not use additional semantic information. We find more promising results in these two

applications when we consider the spatial context of visual words and the global colour histogram.

• Our semantic sparse recoding method is easily applicable to other difficult image and video content analysis jobs. More specifically, it is widely used in the machine learning literature for graph-based learning.

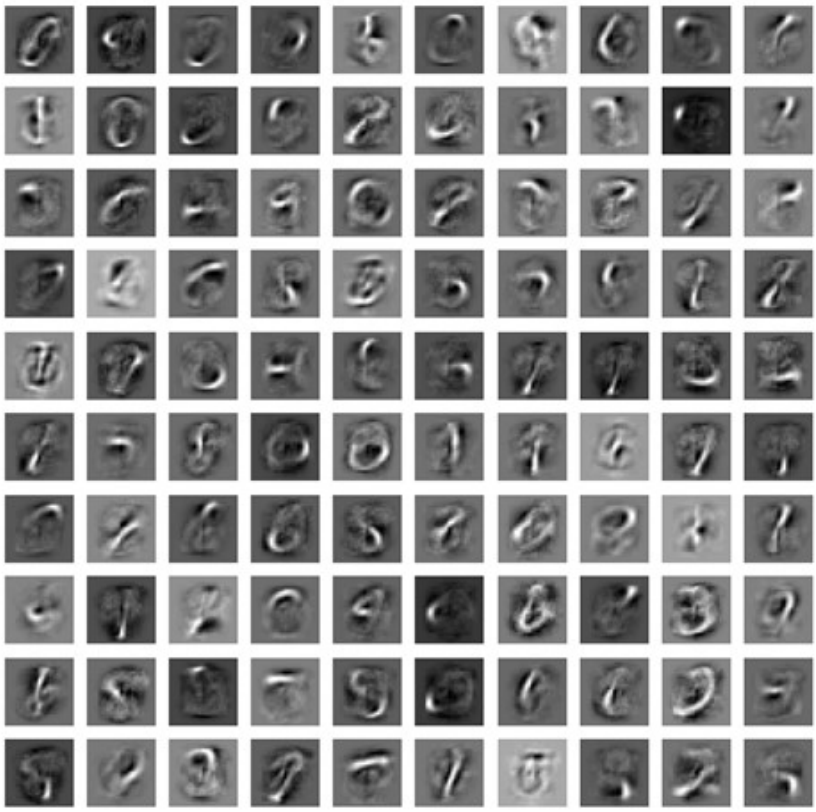

Figure 1 Sparse Coding

The classic visual BOW representation, on the other hand, has two intrinsic disadvantages. To begin, its visual codebook is totally learned via quantization from low-level visual cues, leaving the so-called semantic gap unbridgeable. Although supervisory information about local interest points has been used to optimise visual codebooks, this type of high-level semantics is frequently prohibitively expensive to collect in reality. Second, because visual features are quantized incorrectly in the typical visual BOW representation, the noise issue becomes quite severe, resulting in significant performance reduction. Numerous unsupervised approaches have been explored in the literature to address the noise problem related with the classic visual BOW representation. The primary disadvantage of these unsupervised approaches is that they ignore semantic information. In summary, to our knowledge, relatively few attempts have been made to overcome the two disadvantages of the standard visual BOW representation simultaneously.

When regions comparable to ROIs are located in separate parts of the database photos, this strategy fails to retrieve similar images. For instance, if the user searches for a horse in the image's left corner, the algorithm will not provide similar images having horses in the image's

right corner or other locations. This problem can be solved using the region-matching algorithm. The user-defined ROI sweeps across the entire image in block-by-block fashion. A similarity distance is calculated for each block. The image's output similarity distance is determined using the minimum similarity distance.

LITERATURE REVIEW

On the basis of the fusion of low-level features, Ashraf et al. (2020) created a subjective technique for the CBIR system (texture and color). Color characteristics were extracted using colour moments in the HSV colour space, while texture features were extracted using DWT and Gabor wavelets. To further enrich the feature vector, the colour and edge directivity descriptors were generated and included in the 1 250-dimensional feature vector. The bigger dimension of the feature vector provides more precise retrieval results, but requires more time for searching and comparing. The suggested system was evaluated against a variety of datasets (Corel 1000, Corel 15,000, Corel 5000, and GHIM-10 K) and demonstrated a good level of precision and recall on average. However, the proposed method, like many others in the literature, lacks texture and spatial information.

Huang et al. (2019) developed a new method for extracting regions of interest from colour photos using visual saliency in the HSV colour space. Color saliency is determined using a two-dimensional sigmoid function that incorporates both the saturation and brightness components in order to find regions of vibrant colour. The Discrete Moment Transform (DMT) can be used to determine the saliency of vast areas of interest. By merging colour saliency with DMT-based saliency, a visual saliency map is generated, referred to as the S image. The image calculates a criterion for local homogeneity termed the E image. The high visual saliency object seed point set and the low visual saliency object seed point set are determined using the S and E images. The growing and merging seeded regions are utilised to extract regions of interest. In contrast to our semantic sparse coding approach for visual BOW representation refinement, the problem of feature extraction or label propagation was not formulated as a sparse coding problem. Instead, group sparse coding was used directly for feature selection from multiple groups of visual features. However, these two methods neglected high-level semantic information during feature selection, in contrast to our semantic sparse recoding approach, which makes use of image annotations to refine visual BOW representations.

Zhang et al. (2017) split an image into 32 32 4 4 segments and then computed the average grey value for each segment. Due to the limited number of ROIs examined, the average grey values are classified into three groups using the K-means clustering algorithm. After segmentation, the segmentation image has only three values. For each ROI, colour characteristics based on hue histograms and texture features based on grey level co-occurrence matrices in four directions are extracted and utilised for indexing and similarity comparison. Similarly, in which treated image classification and annotation as group sparse coding over textual keywords and visual BOW representation, semantic refinement of visual words was similarly overlooked. That is, our semantic sparse coding can clean up the initial noisy visual words (i.e. noise reduction) while simultaneously propagating them to semantically similar images along the manifolds defined by the textual BOW representation (i.e. label propagation), resulting in a

more descriptive and robust visual BOW representation. Color data were extracted using the Canny edge histogram and the DWT transform in the YCbCr colour space, whereas texture characteristics were extracted using GLCM. The canny edge approach was used to extract shape characteristics in the RGB colour space.

RESEARCH METHODOLOGY

As illustrated in Figure 2, the general SRI(Sparse Recoding Image) framework comprises of several obligatory phases and several optional stages. The first stage in SRI is when the user submits the query image. All processes applied on the query image will be applied in the same order to all images in the database. Typically, these operations are done on the query image upon user submission and are referred to as online operations; however, the same operations can be conducted on dataset photos prior to query submission and are referred to as offline operations. An optional preprocessing stage may be added in the framework's architecture, which may involve resizing, segmentation, de-noising, and rescaling, among other operations. This optional stage is followed by the most critical stage, feature extraction, in which a visual concept is translated to a numerical representation. Low-level characteristics (i.e., colour, shape, texture, and spatial information) or local descriptors may be extracted. After feature extraction, another alternative preprocessing stage is normalisation or categorization. The final stage is to compare the retrieved features from the query image to all other photos in the dataset to determine which images are the most relevant.

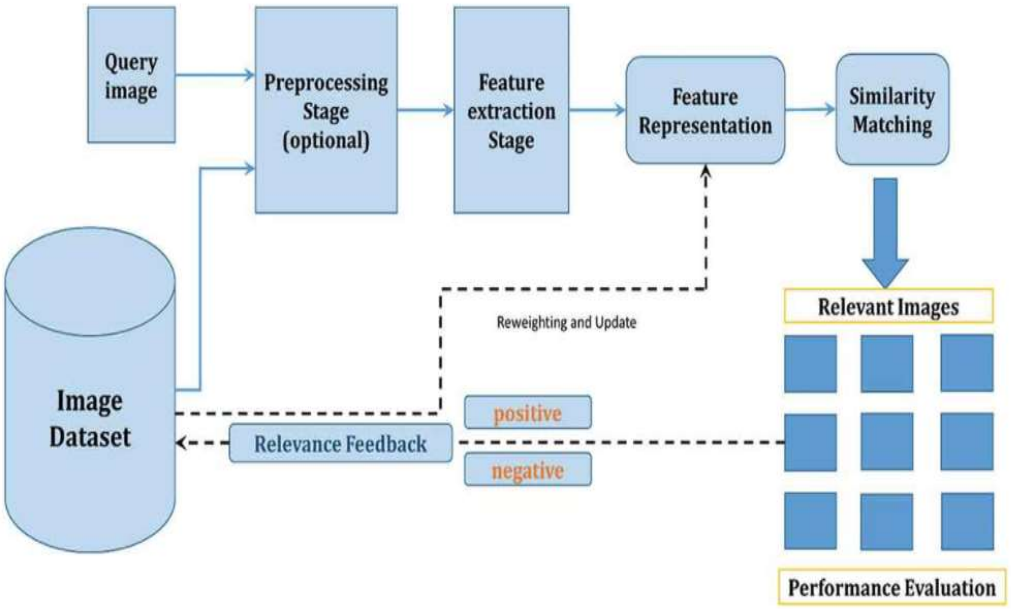

Figure 2 Framework Architecture

Another possible stage is relevance feedback, which enhances the results by user participation by determining which returned photos are relevant and which are not. Numerous strategies for applying relevant feedback to improve SPI performance have been proposed. As previously stated, SRI is primarily concerned with the extraction and selection of features that indicate the semantic content of images. These features can be classified as global features (colour, texture,

shape, and spatial information), which characterise the entire image, and local features, which are typically obtained by segmenting the image or by calculating some critical spots, such as corners, blobs, and edges. Local features are invariant with respect to changes in scale, translation, and rotation. Both categories will be discussed in detail in this section, along with a review of current research demonstrating the significance of the selected features and their effect on system performance.

The feature extraction procedure as well as the similarity measurement have an effect on the performance of image retrieval systems. The similarity measurement identifies which images in the dataset are regarded to be the most relevant to the query image and should be returned. Thus, the similarity measure indirectly impacts the CBIR's accuracy and has an effect on the system's computing complexity. The similarity measure chosen is influenced by the structure of the produced feature vector (type and dimensionality of input data). This selection is a significant and difficult assignment in the literature. The similarity metric is classified into two types: distance metric and similarity metric. Image annotation enables simple and rapid indexing and searching of big sets of photos.

Following the feature extraction and feature vector creation processes, clustering is undertaken, which is the process of grouping image descriptors into a single group that is semantically distinct from the other groups. Clustering is considered an unsupervised learning process because it does not know which group the images' data should belong to prior to clustering. K-means clustering is the most extensively used clustering algorithm in CBIR, particularly when systems rely on local feature extraction methods. Typically, these procedures are followed by a clustering operation to determine the semantic group to which the image belongs. Semantics is typically limited to its perceptual expression in most image annotation systems by learning a matching function that combines low-level information with higher-level visual ideas. Semantic. These methodologies function differently depending on the number of ideas and the type of the data being analysed. Any annotation system is effective if it improves a few of the defining characteristics. The researchers in this work assembled all the information from the image to its recovery using low- and high-level image attributes such as texture, shape, and colour.

RESULTS & DISCUSSION

Experiments are conducted on an image database. The library's image sources include image sets, image search engines, and user-contributed photo albums. It is the benchmark set most frequently used in the field of scene classification. According to the experimental design, 100 training photos were randomly chosen as training samples, while the remaining images were used as test samples. The Caltech-101 image collection (a image of which is shown in Figure 2) has 121 categories and 10,101 images. To compare with earlier methods, 15 photographs from each category were randomly picked for training, while the remaining images were assessed.

Figure 2 Part of the sample image of the Caltech-101 image dataset

On the scene 15 dataset, this method's average classification accuracy is 83.12 percent. It is a confusion matrix developed during the classification of the scene 15 image set. As demonstrated by the classification rate of 86.75 percent on the scene 13 image and 83.12 percent on the scene 15 image, this method is effective.

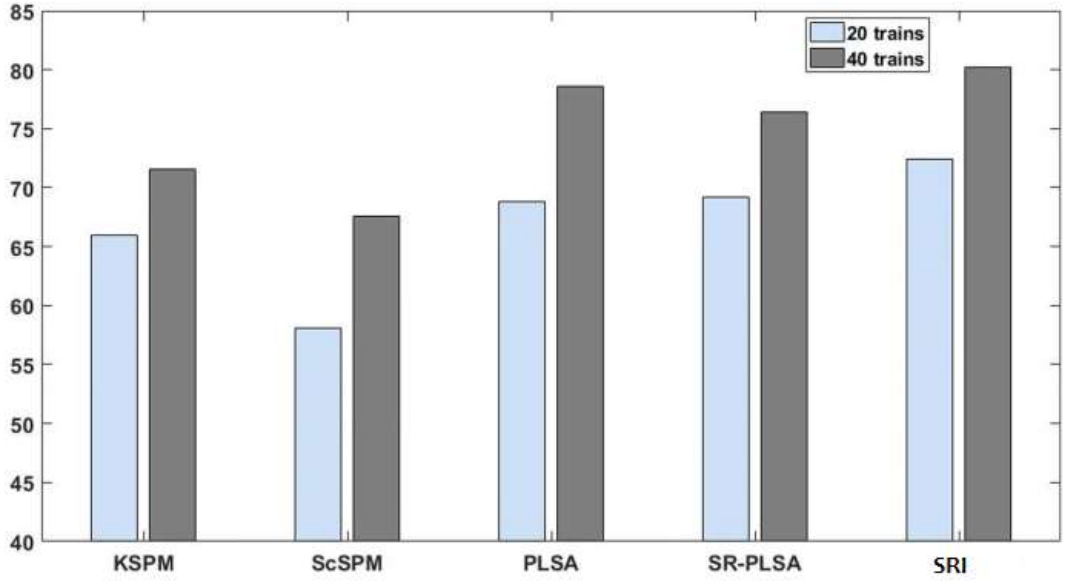

Figure 3 Comparison of classification accuracy of this method and other classification methods on Caltech-101 database (%)

As illustrated in Figure 3, when 20 training photos are used, this paper's SRI classification accuracy is 2.1 percent greater than that of ScSPM, and when 40 training images are used, its classification performance is 3.1 percent higher than that of ScSPM. In comparison to other methods, the suggested method greatly improves classification performance. The suggested image classification method based on sparse coding multi-scale spatial latent semantic analysis is successful and robust, as demonstrated by experimental findings.

CONCLUSION

The primary issue in image scene categorization is how to close the "semantic gap" between low-level characteristics and high-level semantics. It is a significant research idea to answer the fundamental challenge of scene categorization by extracting the local invariant properties of images and generating the images' local semantic concept representation. Sparse coding theory is applied to the study of image local semantic idea representation and has demonstrated great accuracy in classifying image scenes. The existing sparse coding models, on the other hand, are aimed at decreasing signal reconstruction error. It is more crucial to find a discriminant representation for image scene classification than it is to minimise the reconstruction error. As a result, this image presents a strategy for classifying images based on sparse coding and multi-scale spatial latent semantic analysis. The target's spatial position is extracted using image segmentation's spatial pyramid matching, and the target's co-occurrence matrix is formed using feature soft quantization based on sparse coding, which increases the accuracy of the original feature representation.

REFERENCES :

1. Ashraf, R., Ahmed, M., Ahmad, U., Habib, M. A., Jabbar, S., & Naseer, K. (2020, April). MDCBIR-MF: Multimedia data for content-based image retrieval by using multiple features. *Multimedia Tools and Applications*, 79(13–14), 8553–8579. https://doi.org/10.1007/s11042-018-5961-1 [Crossref], [Web of Science ®], [Google Scholar]
2. Huang Chaobing, Liu Quan, Yu Shengsheng (2019), “Regions of interest extraction from color image based on visual saliency”,Journal of Supercomp.
3. Zhang J., Marszalek M., Lazebnik S., and Schmid C. (2017), “Local features and kernels for classification of texture and object categories: A comprehensive study”, Int. J. Comput. Vision, 73, 2, 213–238.
4. P. Guo, T. Wan, and J. Ma. (2015) Experimental Studies of Visual Models in Automatic Image Annotation. Lecture Notes in Computer Science, Springer Berlin Heidelberg,Volume 6761/2011, 562-570, 2011.
5. J. He, M. Li, H.-J. Zhang, H. Tong, and C. Zhang. (2014) Mean version space: A new active learning method for content-based image retrieval. Proc. Multimedia Information Retrieval Workshop, NY, pp. 15–22, 2004.

6. P. S. Hiremath and J. Pujari. (2018) Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement. International Journal of Computer Science and Security, Volume(1) , Issue(4).
7. L. Hollink, G. Nguyen, G. Schreiber, J Wielemaker, B. Wielinga, and B Wielinga.(2017) Adding spatial semantics to image annotations. In in 4th International Workshop on Knowledge Markup and Semantic Annotation at ISWC04, pages 31–40, 2004.
8. J. Yang and Y. Zhang, "Alternating direction algorithms for l1-problems in compressive sensing," SIAM Journal on Scientific Computing, vol. 33, no. 1–2, pp. 250–278, 2011.
9. M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation," in Proc. ICCV, 2009, pp. 309–316.
10. M. Guillaumin, J. Verbeek, and C. Schmid, "Multimodal semisupervised learning for image classification," in Proc. CVPR, 2010, pp. 902–909.
11. A.Ulges, M. Worring, and T. Breuel, "Learning visual contexts for image annotation from Flickr groups," IEEE Trans. Multimedia, vol. 13, no. 2, pp. 330–341, 2011.
12. Wang, F. Jing, L. Zhang, and H.-J. Zhang, "Image annotation refinement using random walk with restarts," in Proc. ACM Multimedia, 2006, pp. 647–650.
13. Makadia, V. Pavlovic, and S. Kumar, "A new baseline for image annotation," in Proc. ECCV, 2008, pp. 316–329.
14. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Sch¨olkopf, "Learning with local and global consistency," in Advances in Neural Information Processing Systems 16, 2004, pp. 321–328.
15. X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in Proc. ICML, 2003, pp. 912– 919