International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 3, Issue 12, December 2014

Survey on Supervised Image Classification Techniques

Anju Davis, Suhara S

***Abstract*—Image classification is one of the most challenging problems in computer vision, especially in the presence of intra-class variation,clutter,occlusion and pose changes. Image classification refers to the labeling of images into one of the predefined categories. In image classification it is very difficult to deal with background information. The main two methods for image classification are supervised and unsupervised classification. The objective of this survey is to provide a summarization of major advanced supervised image classification techniques. This literature survey suggests that the selection of the classification procedure depends on the dataset to be dealt with. Efficient classification can be achieved by multiscale information fusion method which has a classification accuracy of 78.24%.**

***Index Terms*—Image Classification, Supervised Classification, Support Vector Machine, Bag Of Words Model.**

I. INTRODUCTION

In Supervised classification, the nature of the data and their feature distribution is learnt from the training images and this learnt information is used for the classification. In unsupervised classification no prior knowledge about the data is required. Unsupervised classification is based on clustering. Clustering starts with a set of unclassified data and a measure to find the similarity between the data. The data is repeatedly clustered and a label is given to each of the cluster. In supervised classification discriminative classification using Support Vector Machines and variants of the boosted decision trees are two of the leading techniques used in supervised image classification. Advantage of SVM is that non-linear decision boundaries can be learnt using the kernel trick. The SVMs have greater training speed but the runtime complexity of non-linear SVM is high. Boosted decision trees have faster classification speed but they are slow to train. Linear SVMs are popular because of the faster training and classification speed. In this survey the discussed methods are mainly based on SVM classifier. Bag of features model is efficient for image classification. In Bag of features model a visual vocabulary is created from the training images and this visual word are predictive of a certain class of object in the image.BOF model can be incorporated into any of the classification framework.

Manuscript received Dec 02, 2014.

***Anju Davis**,Computer Science, KMCT College of Engineering,Calicut, India,8943093511.*

***Suhara .S**, Computer Science, KMCT College ofEngineering,Calicut, India.*

II. SUPERVISED IMAGE CLASSIFICATION METHODS

In supervised classification, the prior knowledge about the data is required. All the methods described in this section are based on bag of words model. In bag of words model features are extracted from each of the training images. The extracted features are vector quantized to form the visual vocabulary. Each of the images is then represented as histograms based on the codebook from the visual vocabulary. During the testing time the test image will also be represented as a histogram. Histogram matching is then done by any of the similarity measure. This similarity measure will then be incorporated into a kernel and this kernel will be used with a classifier for image classification.

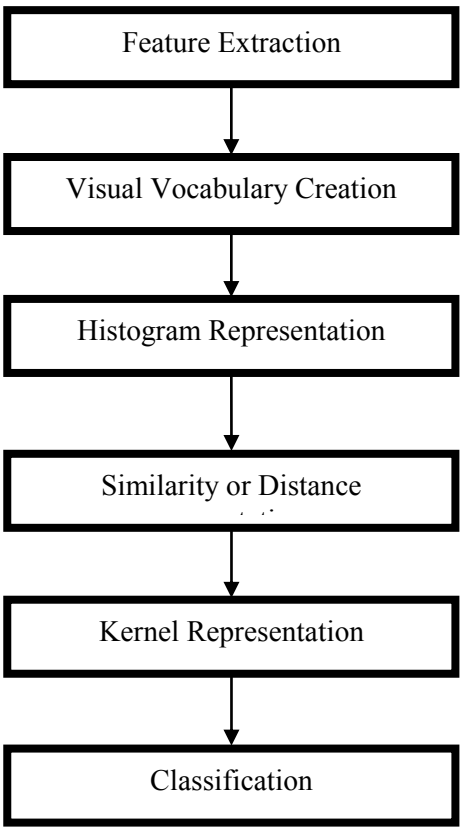

Fig 1. Supervised Classification using Bag of Features Model

HU et al [1], proposed a saliency driven nonlinear diffusion

ISSN: 2278 – 909X

filtering based method for image classification. The image is classified using the multi-scale information fusion based on the original image, the image at the final scale at which the diffusion process converges, and the image at a midscale. Background regions and foreground regions can often be identified using the image saliency. Saliency detection is combined with nonlinear diffusion filtering where the image gradients in the salient regions are increased while those in non-salient regions are decreased. Once the image is represented at three scales, features are extracted and represented using four color SIFT descriptors (Opponent SIFT, rgSIFT, C-SIFT and RGB-SIFT).Bag of words model is used to create a frequency histogram. During the training time the images were clustered to create a visual vocabulary using k-means clustering. The χ^2 distance is used to find the similarity between the histograms. The distances at three scales are combined by weighted sum of the distances and this distance is transformed to a Gaussian kernel and the classification is done using SVM classifier. Multi-scale fusion method has a classification accuracy of 81.51% in the OXFORD flower dataset, which outperforms the existing classification methods. This method has a classification accuracy of 78.24% in the car category of the PASCAL dataset. The results have demonstrated that saliency driven multi-scale information fusion improves the accuracy of image classification.

GAO *et al* [2], classified images by incorporating kernel sparse representation into spatial pyramid matching.KSR-based feature coding is an extension of Sc-based SPM.KSRSPM can be applied to image classification, face recognition and low-rank kernel approximation. Normally in bag of words model k-means clustering is used for feature quantization, which will lead to assigning the feature into more than one cluster. In KSR based feature coding instead of using the k-means a sparse constraint is imposed. SIFT descriptor is used for feature description. SIFT descriptors are extracted at three different patch scales: 16, 25 and 31.For codebook generation and feature coding instead of using Euclidean distance, HIK based similarity measure is used. Features are randomly selected to generate codebook for each dataset. Soft assignment feature coding based on HIK is used in this method. Finally SVM is used for image classification. Experiments were done with KSRSPM using Gaussian kernel and polynomial kernel.KSRSPM using polynomial kernel has better performance with 72%.

Ponce *et al*. [3], proposed a method for identifying natural scene categories based on geometric correspondence. This method is an extension of orderless bag of features model. It is based on a "subdivide and disorder" strategy. This technique works by dividing the image into finer sub-regions and computing the histograms at different resolutions. At a fixed resolution two points are said to match if they fall into the same cell of grid. The classification is done by taking weighted average of the match found at each resolution. More weight is assigned to the matches found at finer resolution. Two types of features are extracted weak features and strong features. Weak features are the oriented edge points whose gradient magnitude in a given direction exceeds a minimum threshold. Strong features are SIFT descriptors of 16×16 pixel patches computed over a grid with spacing of 8 pixels. K-means clustering of random subset of patches from the training set is selected to form the visual vocabulary. In pyramid matching, performance improves as it move from L=0 to a multilevel step. Increasing the size of the vocabulary does not have any effect on classification at higher pyramid levels. Experiments demonstrate that strong features perform well than weak features. Geometric cues have higher discriminative power than visual vocabulary. Here the Multiclass classification is done using one versus all rule. Results show that geometric cues have higher discriminative power than visual vocabulary. This method achieves high accuracy on Caltech-101 dataset. The classification rate is 72.2%.

Varma et al. [4], investigated the problem of learning optimal descriptors for a classification task. The trade-off between discriminative power and invariance distinguishes one descriptor from another. Knowledge of the trade-off leads to improved classification. The focus of this method is to learn the optimal trade-off for classification. The optimal domain specific kernel is learned as a combination of base kernels corresponding to base features.Weightage are given to the kernels according to the preference for the base feature. The descriptors are selected according to the nature of the dataset. Experiments were conducted by using different descriptor combinations according to the dataset available. The optimization is carried out in SVM framework. Experiments were conducted in UIUC textures, Oxford flowers and Caltech101 object categorization databases. In UIUC dataset Rotation, scale , similarity invariant descriptors were obtained and a better performance is obtained by rotationally invariant descriptors.In Oxford flowers dataset a combination of shape,colour,texture descriptors were used and a better classification is obtained from shape descriptor. In Caltech101 dataset four SIFT descriptors were used .Images are represented as bag of words and similarity is calculated using spatial pyramid match kernel.

Maji et al. [8], showed that one can build histogram intersection kernel SVMs (IKSVMs) with runtime complexity of the classifier logarithmic in the number of support vectors. This method is a technique to speed up the histogram comparison in kernalized SVMs.The proposed IKSVM method has a complexity of O (n logm), where n is the dimension of the feature vectors and m is the number support vectors. The proposed method provides speed up and space savings compared to standard implementation. Features are based on a multi-level version of the HOG descriptor and histogram intersection kernel SVM based on spatial pyramid match kernel is used. The oriented edge energy responses in 8 directions using the magnitude of the odd elongated oriented filters from at a fine scale, with non max suppression performed independently in each orientation. The response is then L_1 normalized over all the orientations in non overlapping cells of fixed size. The features at each level are weighted by a factor and concatenated to form our feature vector, which is used to train an IKSVM classifier.

ISSN: 2278 – 909X

Experiments show the multi-level histogram of oriented edge energy features lead to better classification accuracies on INRIA dataset while being as good as the state of the art on Daimler-Chrysler dataset, when used with IKSVM and are significantly better than a linear classifier trained on the same features.

Nilsback et al. [5], proposed a method to extent 'bag of visual words' models to distinguish categories which have significant visual similarity. They demonstrated that by developing a visual vocabulary that explicitly represents the various aspects of colour, shape, and texture that distinguish one flower from another, can overcome the ambiguities that exist between flower categories. This method is based on nearest neighbor classifier architecture. In flower classification the greatest challenge arises from intra-class vs inter-class variability, i.e. there is a smaller variation between images of different classes than within a class itself. Visual vocabulary is created for each of the shape, texture and colour, a combined vocabulary is created using these three vocabularies for classification. HSV colour space is used to represent the Colour.In order to obtain a good generalization, the HSV values for each pixel in the training images are clustered using k-mean clustering. Each image is then represented by normalized frequency histogram. Shape is represented by SIFT descriptors SIFT descriptors are computed on a regular grid to optimize over three parameters. Texture vocabulary is created using MR8 filters. The filter bank contains filter at multiple orientation. Rotation invariance is obtained by choosing the maximum response over orientations. A vocabulary is created by clustering descriptors and the frequency histogram is obtained. Three vocabularies are combined to form a joined flower vocabulary, to obtain a joint frequency histogram. Weights are assigned to each of the descriptors and the weights can be adjusted to give preference to the descriptors.

Results show that the final classifiers have a superior performance over individual classifiers. Experiments demonstrate that the colour feature has a performance of 73.7%, and the shape features achieve a performance of 71.8%.

Zhangh et al. [6], demonstrated that a combination of multiple detectors and descriptors usually achieves better results than even the most discriminative individual detector/descriptor channel. It is an approach to represent images as distributions (signatures or histograms) of features extracted from a sparse set of key point locations and learns a Support Vector Machine classifier with kernels based on two effective measures for comparing distributions, the Earth Mover's Distance and the $\chi 2$ distance. This method is based on bag of key points that uses both background and foreground features to make a classification as a whole. In this method two types of detectors are used, the Harris-Laplace detector and the Laplacian detector. Both the descriptors output circular regions and an affine invariant version is obtain through an affine adaptation procedure. Three different descriptors are used SIFT, SPIN and RIFT. The SIFT descriptor computes a gradient orientation histogram within the support region. For each of 8 orientation planes, the gradient image is sampled over a 4×4 grid of locations, thus resulting in a 4×4×8 = 128-dimensional feature vector for each region. For SPIN descriptors 10 bins for distance and 10 bins for intensity values are used and thus resulting in 100-dimensional feature vectors. To obtain the RIFT descriptor the image region is divided into concentric rings of equal width, and a gradient orientation histogram is computed within each ring. To obtain rotation invariance, gradient orientation is measured at each point relative to the direction pointing outward from the center. Four rings and eight histogram orientations are used to yield a 32-dimensional feature vector. After computing the features, the image is represented as distribution of features called signatures. A visual vocabulary is created from the signatures and a histogram is computed using the vocabulary. The distances between two histograms are computed using $\chi 2$ distance. This distance measure will be incorporated into a Gaussian kernel and finally given to a SVM classifier for classification. This method has a classification accuracy of 53.95% in Calteach101 dataset.

Galleguillos et al. [7], showed that spatial context information is useful for image classification. This method is a novel approach to object categorization by incorporating two types of context co-occurrence and relative location with local appearance features. This method uses a Conditional Random Field to maximize object label agreement according to both semantic and spatial relevance. By vector quantizing the feature space a small set of prototypical spatial relationships are learned from the data. The spatial configuration is learned from MSRC and PASCAL databases. Four types of spatial pair wise relationships are learned from the training images they are above, below, inside and around. The pair wise relationships are represented as a frequency matrix, co-occurrence counts are computed from the frequency matrix which denote the number of times each object co-occur with another object. For each of the image multiple stable segmentations are computed, each of the segments is treated as an individual image .Each of the segment is given as an input to the bag of features model. Segments are models as nodes of CRF, where location and object co-occurrence constraints are imposed. Finally based on local appearance, contextual agreement and spatial arrangements, each segment receives a category label. The average categorization for MSRC database is 68.38% and 36.7% for PASCAL database.

TABLE 1
CLASSIFICATION RATE OF VARIOUS IMAGE CLASSIFICATION METHODS

METHOD	CLASSIFICATION RATE
Multi Scale fusion	78.24%
KSRSPM	72%
SPM	72.2%
Spatial context	68.38%

ISSN: 2278 – 909X

III. Conclusion

This survey is an attempt to summarize few of the popular supervised image classification methods. Better classification accuracy can be obtained by combining different descriptors and learning power invariance trade-off between discriminative power and invariance. Combination of multiple edge detectors and descriptors also result in improved classification. Representing the image at different scales and using the feature from each of the scale for classification also results in better classification. Finally we conclude that image classification using multiscale information fusion have a better classification accuracy with 78.24%.

REFERENCES

[1] W. Hu, R. Hu, N. Xie, H. Ling, and S. Maybank, "Image classification using multiscale information fusion based on saliency driven nonlinear diffusion filtering *," IEEE Trans. Image Process.,* vol. 23, no. 4,pp. 1513 – 1526,April 2014.

[2] S. Gao, I. W. Tsang, and L. Chia, "Sparse representation with kernels," *IEEE Trans. Image Process.*, vol. 22, no. 2,pp. 423 – 434, February 2013.

[3] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2169–2178, vol. 2. Jun. 2006.

[4] M. Varma and D. Ray, "Learning the discriminative power invariancetrade-off," in Proc. IEEE Int. Conf. Comput. Vis., pp. 1–8,Oct. 2007.

[5] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flowerclassification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,vol. 2. pp. 1447–1454. Jun. 2006.

[6] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local featuresand kernels for classification of texture and object categories: A comprehensive study," Int. J. Comput. Vis., vol. 73, no. 2, pp. 213–238,Jun. 2007.

[7] C. Galleguillos, A. Rabinovich, and S. Belongie, "Object categorizationusing co-occurrence, location and appearance," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1–8,Jun. 2008.

[8] S. Maji, A. C. Berg, and J. Malik, "Classification using Intersection Kernel Support Vector Machines is Efficient," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 40 – 48 ,Jun. 2008.

Ms.Anju Davis completed her B.Tech in Computer Science & Engineering from University Of Calicut in the year 2007.She has 6 years of IT Industry experience with iGATE global Solutions Bangalore. She is currently pursuing M.Tech in Computer Science & Engineering from KMCT College Of Engineering, Calicut, Kerala.Her research interests include image processing.

Ms.Suhara .S completed her B.Tech in Information Technology from Muslim Association college of engineering, Kerala in the year 2006 and M.Tech in Computer Science and Engineering from M.S university in the year 2012.Currently she is an assistant professor with KMCT College Of Engineering, Calicut, Kerala since 2013.Her research interests include image processing.

ISSN: 2278 – 909X