A Survey on Image Classification using Data Mining Techniques

Vyoma Patel[1] G. J. Sahani[2]
[1]Student [2]Professor
[1,2]Department of Computer Engineering
[1,2]Sardar Vallabhbhai Patel Institute of Technology, Vasad, Gujarat, India

Abstract— Image classification is one of the most useful and essential research field in computer vision domain. The growing demands for image classification in computer vision having application such as video surveillance, image and video retrieval, Web content analysis, biometrics etc. has pushed application developers to search ways to manage and classify images more efficiently. Classification includes image pre-processing, object detection, object segmentation, feature extraction and object classification. The main goal of image classification is to classifying image into different classes according to their visual characteristics. Recently there has been new trend to merge two or three field to take the advantage of each field. In this paper presents a survey on various image classifications by data mining techniques and image representation methods.

General Terms: Image Classification, Image Representation

Key words: K-NN, SVM, DT, BoW, SPM, B2S, FV, Bag-of-FLHs

I. INTRODUCTION

Image classification remains to be a major challenge in field of computer vision. With the development of computer, image classification have been widely used in many military areas and civilian areas such as safety monitoring, human-computer interaction, medical diagnosis, vehicle navigation, industrial visual inspection, robot sensing and remote sensing. Despite some research progress, the ability of programs to automatically interpret the content of images is still limited in terms of accuracy and scalability. For example, searching keyword by Google image thousands of images are easily retrieved but typically they contain outlier images that do not match with what the user wanted. Image classification consists of database that contains predefined patterns and compares those patterns to detected object to classify in to proper category. For identifying target images from image database use appropriate algorithm which can express image correctly and accurately. Image classification steps involve preprocessing, feature extraction, feature selection and classification of the object.

Now a days data mining technique, Frequent itemset mining (FIM) is used to tackle computer vision problem like image classification, action recognition, scene understanding, object recognition and object-part recognition. It is an important data mining technique to extract informative patterns. It allows the construction of compound features which can capture more discriminative information from images. The main aim is to optimize the process of finding patterns which should be efficient, scalable and can detect the important patterns which can be used in various ways. Frequent pattern-based classification framework can achieve good scalability and high accuracy in classifying large datasets in image classification [2].

The main objective of image classification is to identify the feature occurring in an image. Local features are at the heart of the most successful approaches to object class detection and image classification [4]. The most common visual features include color, texture and shape, etc. Feature extraction is the process of generating features to be used in the selection and classification tasks. Feature selection reduces the number of features provided to the classification task. Those features which are likely to assist in discrimination are selected and used in the classification task. The SIFT Descriptor [5], used for local feature extraction. After extracting most suitable features image is classified by using image classification method.

This survey focuses on image classification and representation methods and also describes their advantages and disadvantages.

II. IMAGE CLASSIFICATION METHODS

A. K-Nearest-Neighbor (K-NN):

K-nearest neighbor algorithm is a method for classifying objects based on closest training label in the feature space. Training process for this algorithm only consists of storing feature vectors and labels of the training images. K value which gives the minimum error rate may be selected for K-Nearest Neighbor classification. Distance function for K-Nearest-Neighbor is Euclidean distance [1]. It uses distance based comparison to assign equal weight to each attribute. They can suffer from poor accuracy when noisy and irrelevant attributes are given. It is classifying the pattern by comparing a given test pattern with training pattern that are similar to it. It is widely used in pattern recognition. The object is classified based on the labels of its nearest neighbors and if k=1, the object is simply classified as the class of the object nearest to it [7]. K-NN has a wide application including text mining, agriculture, finance and medicine.

B. Support Vector Machine (SVM):

A support vector machine (SVM)[6] , a promising method for classification of both linear and nonlinear data. SVM classification uses different planes in space to divide data points. It gains flexibility in the choice of threshold and handles more input data very efficiently. Its performance and accuracy depend upon the selection of hyper plane and kernel parameter. The goal of SVM Classification is to produce a model, based on the training data, which will be able to predict class labels of the test data accurately. SVM have been applied to a number of areas including object recognition, handwritten digit recognition and speaker identification. A main advantage of SVM is that: reduction in computational complexity, simple to manage error frequency [7]. SVM achieve good accuracy among all other classification methods. Most image classification used this method for classifying images from large database. SVM with a small number of support vectors can have good generalization capability.

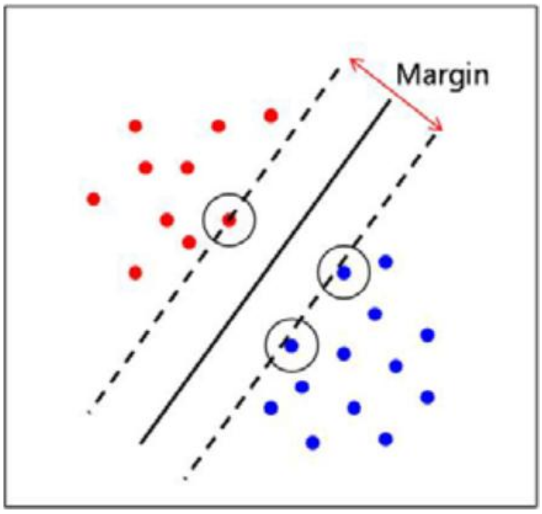

Fig. 1: SVM Classification (Dot circled represent support Vectors) [7]

C. Decision Tree (DT):

Decision tree consist of mainly three parts: Partitioning the nodes, find the terminal nodes and allocate class label to terminal nodes. It is based on hierarchical rule. It handles high dimensional data and representation of knowledge in tree form which is easy to humans for understanding purpose. When decision tree built, many of branches reflects noise in the training pattern so, tree pruning attempts to identify and remove such branches and improve the accuracy of classification. Decision tree induction algorithms have many applications in classification including medicines, manufacturing and production, financial analysis, astronomy and molecular biology [1]. It can handle nonparametric training data and does not required an extensive design and training [3].

Classification methods	**Advantages**	**Disadvantages**
K-Nearest Neighbor	– Simple and easy to learn – Training is very fast – Robust to noisy training data – Effective if training data is large	– Memory limitation – Computation complexity
Support Vector Machine	– Good accuracy – Flexibility in choice of threshold – Problem of over fitting is eliminated	– Training is time consuming – Structure of algorithm is difficult to understand
Decision Tree	– Not require extensive training – Efficiency is good – Easy to understand	– Becomes complex calculation when various values are undecided outcomes are correlated – Suffer from over fitting

Table 1: Comparison of different image classification methods

III. IMAGE REPRESENTATION METHODS

A. Bag of Words (BoW):

The bag-of-words (BoW) method was first proposed in the text retrieval domain problem for text document analysis, and it was further adapted for computer vision applications. BoW is histograms over vector-quantized local features. The main advantage of the method is its simplicity, its computational efficiency and its invariance to affine transformations, as well as occlusion, lighting and intra-class variations [8]. To extract the BoW feature from images involves the following steps: (i) it automatically detect points of interest, (ii) compute local descriptors over those points, (iii) quantize the descriptors into words to form the visual vocabulary, and (iv) find the occurrences in the image of each specific word in the vocabulary for constructing the BoW feature (or a histogram of word frequencies) [9].

B. Spatial Pyramid Matching (SPM):

Spatial Pyramid matching (SPM) [10], repeatedly subdividing the image and computing histograms of local features at increasingly fine resolutions. This method significantly improves performance over a basic bag-of-Word representation. It is capture perceptually salient features from the images. There is L level in SPM and matches found in level L also include all the matches found at the finer scale L-1[11]. Figure 2 shows the construction of pyramid and it consist of three feature types: circles, diamonds, and crosses. At the top subdivide an image into three different level of resolution. . Next, for each level of resolution and each channel, count the features that fall in each spatial bin. Pyramid matching works by placing a sequence of increasingly coarser grids over the feature space and taking a weighted sum of the number of matches that occur at each level of resolution. At any fixed resolution, two points are said to match if they fall into the same cell of the grid; matches found at finer resolutions are weighted more highly than matches found at coarser resolutions. Let X and Y be two sets of vectors in a d-dimensional feature space. The number of matches found at level l also includes all the matches found at the finer level l + 1. Therefore, the number of new matches found at level l is given by $I^l - I^{l+1}$ for l = 0. . . L – 1. H_X^l and H_Y^l denote the histograms of X and Y at this resolution, so that H_X^l(i) and H_Y^l(i) are the numbers of points from X and Y that fall into the i^{th} cell of the grid. Then the number of matches at level l is given by the histogram intersection function defined by,

$$I(H_X^l, H_Y^l) = \sum_{i=1}^{D} \min(H_X^l\,(i), H_Y^l\,(i))$$

and for pyramid match kernel

$$k^L(X,Y) = I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} (I^l - I^{l+1})$$

By using above equation weight each spatial histogram. Each channel m gives us two sets of two-dimensional vectors, X_m and Y_m, representing the coordinates of features of type m found in the respective images. The final kernel is then the sum of the separate channel kernels:

$$k^L(X,Y) = \sum_{m=1}^{M} K^l(X_m, Y_m)$$

It must be also observed that not any significant improvement beyond level L=2. Thus, this method has shown promising results on three large-scale and diverse datasets.

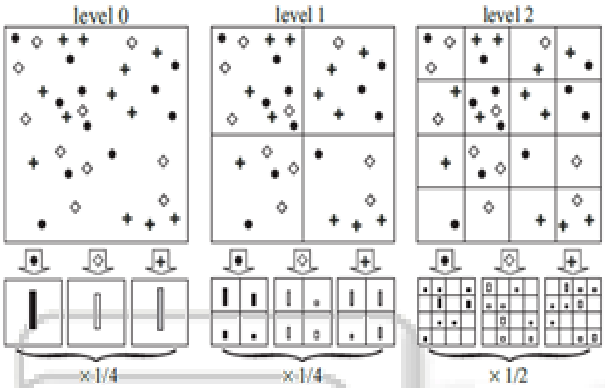

Fig. 2: Example of Toy constructing a three-level pyramid[10]

C. Bag To Set (B2S):

B2S (Bag to Set) image representation method [12], is more discriminative than histogram based Bag-of-Words representation and it keeps all frequency information of the image. This method convert a histogram based bag representation into a set representation. It is creates a new feature for each value of a feature in the bag representation. B2S is a one-to-one function so it always recovered to its original representation. B2S enables the datasets to be more separable which makes it more discriminative than previous histogram based bag representation, while containing all original information at the same time. In B2S representation the frequency of any feature is either 0 or 1, which makes it a set expression. B2S is useful in area including text mining and image classification and it enabled to translate the frequency information (used in computer vision and text mining fields) into a set representation (used in most data mining algorithms) without any loss of information. It scans dataset only ones and runs in a linear time of the number of total items in a dataset. This method leads to higher classification accuracy.

D. Fisher Vetor (FV):

The Fisher Vector[13,15] extends the BOW counting (0-order statistics) to encoding second order statistics. It provides a more general way to define a kernel from a generative process of the data. Let X = {X_t, t = 1. . .T} be the set of T local descriptors extracted from an image. The generation process of X can be modeled by a probability density function u_λ with parameters λ.

$$G_\lambda^X = \frac{1}{T} \nabla_\lambda log u_\lambda(X).$$

The dimensionality of this vector depends only on the number of parameters in λ, not on the number of patches T. A kernel on these gradients is K(X,Y)= $G_\lambda^{X\prime} F_\lambda^{-1} G_\lambda^Y$ where, F_λ is the Fisher information matrix of u_λ and defined by,

$F_\lambda = E_{x\sim u_\lambda}[\nabla_\lambda log u_\lambda(x) \nabla_\lambda log u_\lambda(x)']$.

$F_\lambda = L'_\lambda L_\lambda$ and K(X, Y) can be rewritten as a dot-product between normalized vectors : $g_\lambda^X = L_\lambda G_\lambda^X$ where, g_λ^X refer as the Fisher vector of X. It transforms a variable length feature set into a fixed sized representation and it is suitable for supervised (classification) and unsupervised (clustering) tasks. It is much richer representation of the low level feature distribution in the image. Fisher Vector (FV) has many advantages: it bring large improvements in terms of accuracy, it is efficient to compute, it leads to excellent results even with efficient linear classifiers and it can be compressed with a minimal loss of accuracy. The dimensionality of vector depends upon the selection of parameter.

E. Bag-Of-FIHs:

Bag-of-FLHs [16], is a new image representation technique which is used data mining technique. It is more discriminative than traditional Bag-of-Words (BOW). Mining methods allow the construction of high-level sets of compound features which can, capture more discriminative information from an image. Each images is described by keypoint which is identified by SIFT descriptor. For each keypoint generate Local histogram also called Local Bag-of-Words (LBOW) using K nearest neighbour. For each LBOW generate transaction by applying Bag to set (B2S) method. Figure 3 shows Bag-of-FLHs creation process. Local histogram mining discover unique set of local pattern called Frequent Local Histograms or FLHs. Set of FLHs pattern find by applying frequent itemset mining algorithm. These generated FLH pattern are huge so, relevant pattern mining method used to select discriminative, representative non-redundant FLH pattern. These selected FLH patterns semantically meaningful for capturing the relevant shape information in an image. Most suitable pattern for classification can be finding by gain of pattern t. After computing the most relevant and non-redundant FLHs, represent each image using a bag-of-FLHs by counting the occurrences of FLHs in the image. Let L be such a bag-of-FLHs for the image I_L and M be the bag-of-FLHs for the image I_M. Now by using the kernel function,

K(L,M) = $\sum_i min(\sqrt{L(i)}, \sqrt{M(i)})$

Find the similarities between the bag-of-FLHs of L and M. Good classification accuracies provides by this kernel for frequent pattern-based image representation This method gives high accuracy among all other methods.

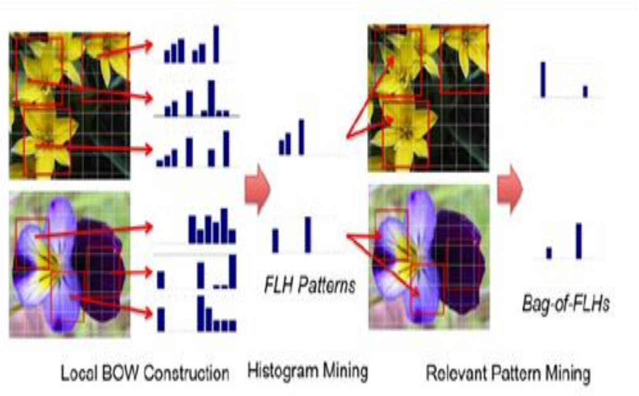

Fig. 3: Bag-of-FLH creation process [16]

Image Representation Methods	Advantages	Disadvantages
Bag-of-Words	– It is simple method – Computation Efficiency is good – Used Feature histogram to represent an image	– Limited Descriptive ability – Incapable of capturing shape or of a segmenting of objects from its background. – Ignore spatial relationship among the patches
Spatial Pyramid Matching	– It can capture salient feature from an image – Achieve high accuracy on large database	– Sensitive to image clutter and geometric deformation
Bag to Set	– High classification accuracy – B2S translate the frequency information into a set representation without any loss of information	– B2S generate artificial visual pattern which is not actually present in dataset
Fisher Vector	– Features are described by identifying various stochastic relationships – Improvement in terms of accuracy	– It uses stochastic approach – Dimensionality of the vector depends on selection of parameter
Bag-of-FLH	– Not generate artificial visual pattern which is not actually present in dataset. – Robust to occlusions and image clutter	– Precise solution depend upon the selection of FLH pattern

Table 2: Comparison of different image representation methods

IV. CONCLUSION

This survey gives theoretical knowledge about various image classification and representation methods. We also discussed their advantages and disadvantages. Effective image classification achieves good accuracy by selecting best image representation method. By merging data mining techniques for classification achieves good result and accuracy. The main objective of this survey is to help the researchers to select best technique for image classification.

REFERENCES

[1] Han, J., Kamber, M. 2006. Data Mining Concept and Techniques, Second ed. San Francisco, CA.

[2] Cheng, H., Yan, X., Han, J., and Hsu, C.-W. 2007. Discriminative frequent pattern analysis for effective classification. IEEE 23rd International Conference on Data Engineering, 716-725

[3] Pooja, K., Sonam, S., Sonu, A. 2013. A survey on image classification approaches and techniques. Department of Computer Science & Applications-IJARCCE, Volume.2, Issue 1, 1005-1009.

[4] Quack, T., Ferrari, V., Leibe, B., Van Gool, L. 2007. Efficient mining of frequent and distinctive feature configurations. IEEE 11th International Conference on Computer Vision, 1 – 8.

[5] David, G. L., 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision-Springer, Volume.60, Issue 2, 91–110.

[6] Christoper, J.C.B. 1998. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery,121–167.

[7] Jinho, K., Byung-soo, K., Silvio, S. 2012. Comparing Image Classification Methods: K-Nearest-Neighbor and Support-Vector-Machines. International conference on Computer Engineering and Applications-ACM, 133-138.

[8] Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C. 2004. Visual categorization with bags of

keypoints. International Workshop on Statistical Learning in Computer Vision, ECCV, 1–22.
[9] Chih-Fong, T. 2012. Bag-of-Words Representation in Image Annotation: A Review. ISRN Artificial Intelligence, Volume.2012, 1-19.
[10] Lazebnik, S., Lazebnik, S., Ponce, J. 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Computer Vision and Pattern Recognition-IEEE, Volume.2, 2169–2178.
[11] Zhang, E., Mayo, M. 2010. Improving Bag-of-Words Model with Spatial Information. Image and Vision Computing New Zealand-IEEE, 1 - 8.
[12] Kim, S., Jin, X., Han, J. 2010. Disiclass: discriminative frequent pattern-based image classification. Tenth International Workshop on Multimedia Data Mining- ACM, Volume.12, 178-188.
[13] Perronnin, F., S´anchez, J., Mensink, T. 2010. Improving the Fisher Kernel for Large- Scale Image Classification. Copmuter Vision- Springer, Volume.6314,143–156.
[14] Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J. Image Classification with the Fisher Vector: Theory and Practice. 2013. International Journal of Computer Vision-Springer, Volume.105, Issue 3, 222-245.
[15] Perronnin, F., Dance, C. 2007. Fisher kernels on visual vocabularies for image categorization. Computer Vision and Pattern Recognition-IEEE, 1–8.
[16] Fernando, B., Fromont, É., and Tuytelaars, T. 2012. Effective use of frequent itemset mining for image classification**.** Computer Science-Springer, Volume.7572, 214–227.
[17] Fernando, B., Fromont, É., and Tuytelaars, T. 2014. Mining Mid-level Features for Image Classification. International Journal of Computer Vision, Volume. 108, 186-203.