IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-2, Issue-5 E-ISSN: 2347-2693

Shape And Texture Based Scene Classification

Bibin Prasad[1*] and Usha Kinsly Devi [2]

[1*,2] *Regional Center of Anna University, Tirunelveli,India*

www.ijcaonline.org

Received: 14/04/2014 Revised: 10/05/2014 Accepted: 22/05/2014 Published: 31/05/2014

Abstract— Humans are extremely proficient at perceiving natural scenes and understanding their contents. Scene recognition in Human is the natural activity by which human can easily recognize the scene even if the scene is complex, partially occluded or blurred. In machine vision the recognition rate is less compared with human vision. To improve the recognition rate of the machine vision an efficient structural and textural based features are extracted from the image. H-Descriptor with Local Binary Pattern (LBP) [24] and H-Descriptor with Local Gradient Pattern (LGP) can effectively extract structural arrangement and textural arrangement of pixels in an image. LGP is invariant to local intensity variation so it is efficient for scene classification. LBP and LGP [23] is applied for each slices when the input image is separated into three different slices. Then Haar wavelet is applied for the input image and three different slices. The HOG is applied for each Haar wavelet transformed images to produce H-Descriptor with Local Binary Pattern and H-Descriptor with Local Gradient Pattern. Then by taking the H-Descriptor with Local Binary Pattern and H-Descriptor with Local Gradient Pattern as two independent feature channels, and combined them to arrive at a final decision using SphereSVM [22] for achieving an effective scene categorization.

Keywords— H-Descriptor; Local Binary Pattern; Local Gradient Pattern; Haar wavelet; SphereSVM

1.INTRODUCTION

One of the long-term goals of computational vision is to be able to understand the world through visual images. Humans are extremely proficient at perceiving natural scenes and understanding their contents. The eye is one of the most important organs of the human body and our skills greatly depend on our ability to see, recognize, and distinguish objects and to estimate distances. Most jobs depend on our ability of visual perception. Computers do not 'see' in the same way those human beings are able to. Cameras are not equivalent to human optics and while people can rely on inference systems and assumptions, computing devices can 'see' by examining individual pixels of images, processing them and attempting to develop conclusions with the assistance of knowledge bases and features such as Pattern recognition engines. Although some machine vision algorithms have been developed to mimic human visual perception, a number of unique processing methods have been developed to process images and identify relevant image features in an effective and consistent manner. As amazing as the human sense of vision may be used in machines to identify and automatically recognize the surroundings. Machine vision refers the vision technology according to the human vision system. It describes the understanding and interpretation of technically obtained images for controlling natural processes. It has evolved into one of the key technologies in automation's, which is used in virtually all industries and surveillance applications. The ability of humans to recognize thousands of object categories in cluttered scenes, despite variability in pose, changes in illumination and occlusions, is one of the most surprising capabilities of visual perception, still unmatched by computer vision algorithms.

Corresponding Author: *Bibin Prasad*

Navneet Dalal [8] et al (2005), proposed [9] the technique Histogram of Gradient (HOG) for human detection. In which the original image is converted into small grid of same size and the Histogram Of Gradient for each grid is find out to extract the shape invariant feature. The normalized block representation of each grid provide efficient output if there any variation or scale changes. This technique is more efficient for finding the right person in the scene. The accuracy of the system is minimum compared with other person finding methods. SugataBanerji, et al (2013), proposed [24] new image descriptors called H-Descriptor. H-Descriptor is based on incorporating additional useful and important features for object and scene image classification, such as shape and local features. H-Descriptor integrates the 3D-LBP and the HOG of its wavelet transform, to encode color, shape, and local information. The H-Descriptor achieves better image classification performance than other popular descriptors, such as the Scale Invariant Feature Transform (SIFT), the Pyramid Histograms of visual Words (PHOW), and the Pyramid Histograms of Oriented Gradients (PHOG). Bongjin Jun, et al (2013), proposed [23] a local transform based texture feature called Local Gradient Pattern (LGP), which makes the local intensity variations along the edge components robust. The local intensity variation will affect the performance of the scene classification and object recognition. To avoid such inconvenience, the LGP technique is employed. It is much similar to Local Binary Pattern (LBP), but it varies only in the computation of the gradient value to improve the recognition rate of the system. Robert Strack, et al (2013), introduces [22] Sphere Support Vector Machines (SVMs) as the new fast classification algorithm based on combining a minimal enclosing ball approach. The nearest neighbor point classification is difficult and complex in the case of other SVM techniques. This method mainly improve the accuracy of the system as well as it can carries large

dataset at the same time, that is the advantage compared with the other SVM techniques.

Human can easily recognize any object, if the scene is blurred, partially occluded and even it may at long distance. Thus human vision is more efficient and recognition rate is high. But in the case of machines it is much difficult to recognize the object if the image or scene is partially occluded, blurred or even at long distance. This general challenging task involves a number of vision tasks such as object recognition, scene classification is contextual relationship among image's components such as objects, regions, blocks, etc. is difficult to extract. This difficulty can be solved by extracting and combining the structural and textural features from the image then classified with a Sphere SVM classifier to achieve better recognition rate.

II.BLOCK DIAGRAM FOR SCENE RECOGNITION

The proposed block diagram representing the effective scene classification as follows,

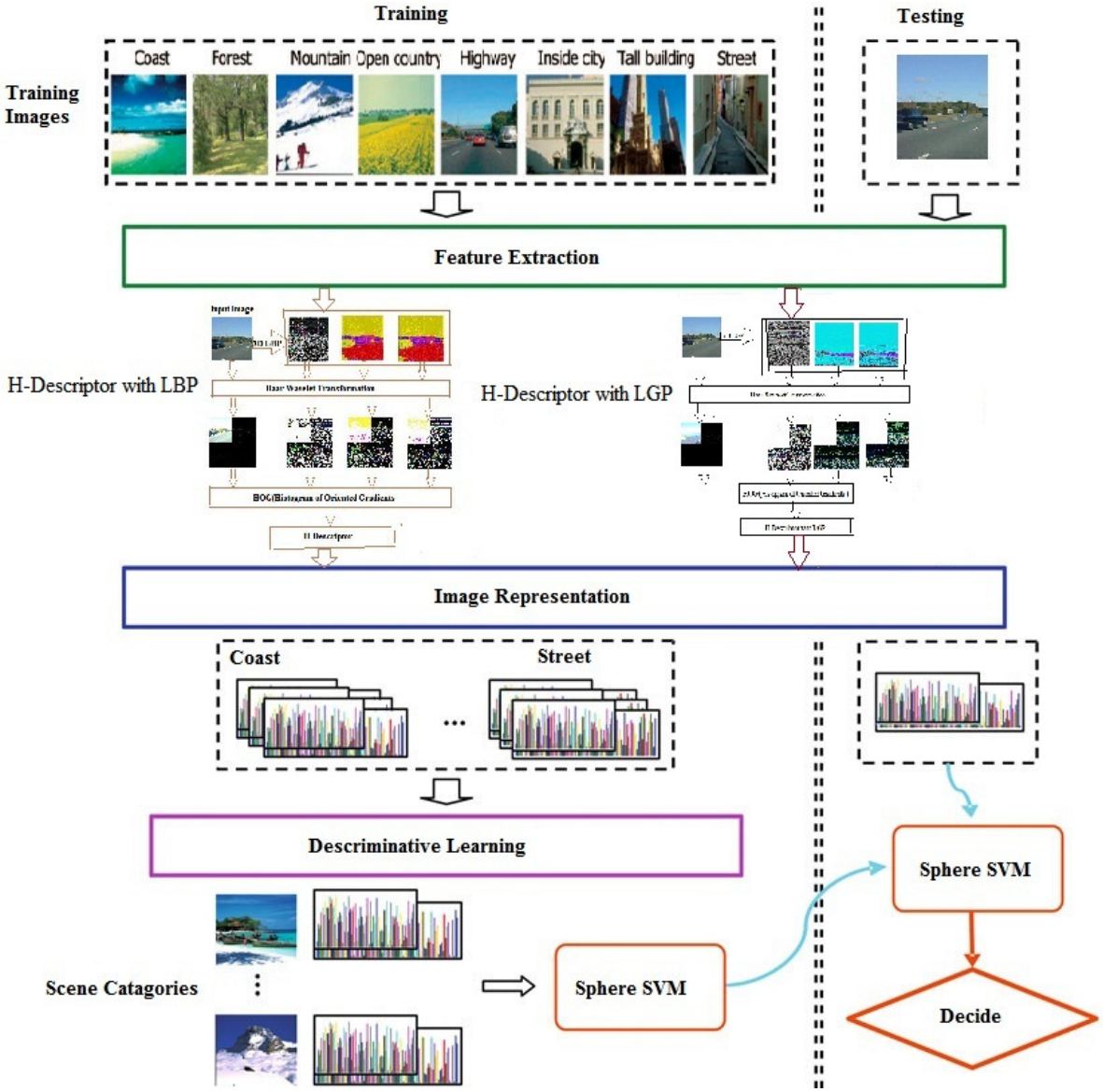

Figure 1. Block diagram for scene recognition

Implementation of shape and texture based scene classification involves extracting texture and shape based features from the input image. Fig.1 shows the proposed methodology for the shape and texture based scene classification. The shape based feature will gives the structural arrangement of pixel in an image so that it is called as the structural feature. The structural feature is the type of shape invariant feature. So that it is efficient for effective scene

classification. The regular repetition of pixels in an image can be effectively represented by using the texture filter. The structural feature can be represented by using the H-Descriptor with LGP and H-Descriptor with LBP [24], and then the textural feature can be represented by using Local Gradient Pattern (LGP) and Local Binary Pattern (LBP). Then finally these features are combined by using the Sphere SVM classifier to achieve an efficient scene classification.

III. FEATURE EXTRACTION

In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant, then the input data will be transformed into a reduced representation set of features. Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which overfits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. In computer vision and image processing the concept of feature is used to denote a piece of information which is relevant for solving the computational task related to a certain application. To improve the recognition rate of the machine vision two different features are extracted from the scene or the image. The feature extraction mainly enhances the recognition rate of the system. The features mainly act as the function which the human visual cortex can do.

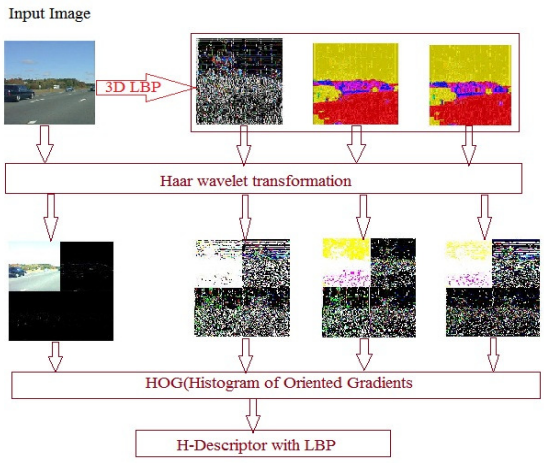

Figure 2. H- Descriptor with LBP

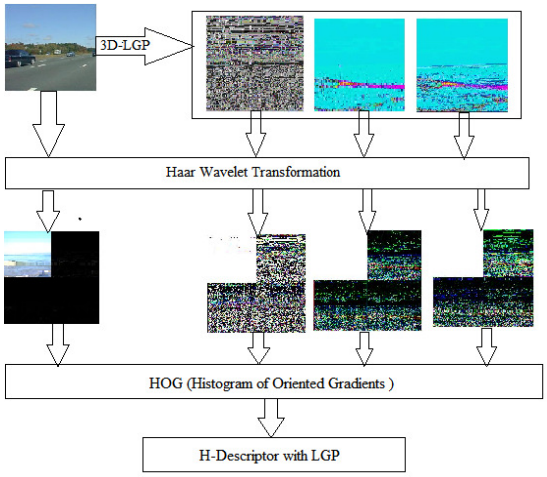

Figure 3. H- Descriptor with LGP

In computer vision structural features are efficient for scene recognition because of it is a shape invariant feature. H-Descriptor is the one of the important structural feature which can effectively increase the recognition rate of the scene classification. This method of feature extraction involves three different steps which are combined to form H-Descriptor with LBP and LGP. 3D-LBP, 3D-LGP, Haar wavelet transformation and HOG [8] calculation are the three important steps behind the H-Descriptor extraction. Fig.2 and Fig.3 shows the basic steps involved in the H-Descriptor with LBP and LGP extraction. Computation of LGP and LBP only varies.

A. Textural Feature Extraction

In many machine vision and image processing algorithms, simplifying assumptions are made about the uniformity of intensities in local image regions. However, images of real objects often do not exhibit regions of uniform intensities. For example, the image of a wooden surface is not uniform but contains variations of intensities which form certain repeated patterns called visual texture. The patterns can be the result of physical surface properties such as roughness or oriented strands which often have a tactile quality, or they could be the result of reflectance differences such as the color on a surface. Texture is the most important visual cue in identifying these types of homogeneous regions. This is called texture classification. The goal of texture classification then is to produce a classification map of the input image where each uniform textured region is identified with the texture class it belongs. The texture has an important role in scene recognition. An image texture is a set of metrics calculated in image processing designed to quantify the perceived texture of an image. Image Texture gives us information about the spatial arrangement of color or intensities in an image or selected region of an image. Image textures can be artificially created or found in natural scenes captured in an image. Image textures are one way that can be used to help in classification of images. In this method 3D-LBP and 3D-LGP are represented as textural feature.

B. 3D-LBP

The motivation for the three Dimensional Local Binary Patterns (3D-LBP) descriptor rests on the extension of the conventional LBP method to incorporate the color cue when encoding a color image. Specifically, for a color image, the 3D-LBP descriptor generates three new color images by applying three perpendicular LBP encoding schemes. The three Dimensional Local Binary Patterns (3D-LBP) descriptor that produces three new color images for encoding both color and texture information of an image. The Local Binary Patterns (LBP) method derives the texture description of a grayscale image by comparing a center pixel with its neighbors. In particular, for a 3×3 neighborhood of a pixel p=[x,y], p is the center pixel used as a threshold. The neighbors of the pixel p are defined as N(p,i)=[x_i,y_i], i = 0,1, . . . ,7, where i is the number used to label the neighbor. The value of the LBP code of the center pixel p is calculated as follows,

$$LBP(P)=\sum_{I=0}^{7} 2^{I}\, S\{G[N(P,I)] - G(P)\} \qquad (1)$$

Where G(p) and G[N(p,i)] are the gray level of the pixel p and its neighbor N(p,i), respectively. S is a threshold function that is defined as follows,

$$S(x_i,x_c)=\begin{cases}1 & ,xi>xc\\ 0 & ,otherwise\end{cases} \qquad (2)$$

LBP tends to achieve grayscale invariance because only the signs of the differences between the center pixel and its neighbors are used to define the value of the LBP code as shown in Equation (1).

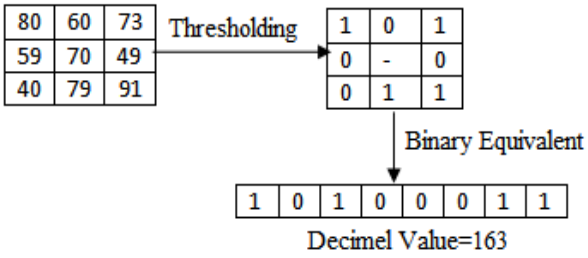

Figure 4. LBP Calculations

Fig. 4 illustrates how the LBP code is computed for the center pixel whose gray level is 70. In particular, the center pixel functions as a threshold and after thresholding the right3×3 matrix reveals the signs of the differences between the center pixel and its neighbors. The signs are derived from Equation (1) and (2), and the threshold value is 70, as the center pixel is used as the threshold in the LBP. The binary LBP code is 10100011, which corresponds to 163 in decimal. LBP, however, does not encode color information, which is an effective cue for pattern recognition such as object and scene image classification. Normally, after performing an LBP operation, the outer pixels could be discarded. However, since the number of color planes is just three, it is not possible to simply discard the top and bottom planes after performing the LBP operations. To solve this problem, it is better to replicate the existing planes in a manner that puts an extra plane on either side of the three existing planes without copying a plane next to itself. For example, if the image is RGB, our new five-plane matrix will be BRGBR. After the 3D-LBP operation is done, these two new planes, i.e. the first and fifth planes of the five-plane image, are discarded to give us a three plane image again.

The 3D-LBP descriptor thus encodes the color and texture information to generate three new color images which will be further processed in order to extract shape and local information. This is the first step of generating the H-Descriptor in which the three 3D-LBP color images are obtained from the input color image.

C. 3D-LGP

LGP [23] have been applied to tasks such as face detection, face recognition, facial expression recognition, gender recognition, face authentication, gate recognition, image retrieval, scene classification, shape localization, and object detection. LGP is similar to Local Binary Pattern (LBP), but it only varies in the computation of the gradient values that is effective for the scene classification.Similar to 3D-LBP descriptor, 3D-LGP also generates three new color images by applying three perpendicular LGP encoding schemes. The three Dimensional Local Binary Patterns (3D-LGP) descriptor that produces three new color images for encoding both color and texture information of an image. LGP generates constant patterns irrespective of local intensity variations along edges. The LGP operator uses the gradient values of the eight neighbors of a given pixel, which are computed as the absolute value of the intensity difference between the given pixel and its neighboring pixels. The average of the gradient values of the eight neighboring pixels is then assigned to the given pixel and used as the threshold value for LGP encoding. A pixel is assigned a value of 1 if the gradient value of a neighboring pixel is greater than the threshold value, and a value of 0 otherwise. The LGP code for the given pixel is then produced by concatenating the binary 1s and 0s into a binary code.

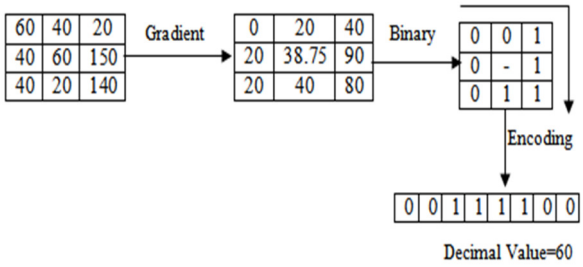

Figure 5.Calculation of LGP

The calculation involved in the computation of LGP can be easily understood from the above fig.5. Initially a 3×3 matrix or the pixel with 8 neighborhoods is assigned. The gradient values are computed by calculating the absolute difference between the center pixel and the 8 different neighborhoods. After calculating the gradient value the center pixel is updated by taking average value of the 8 neighborhoods as 38.75. Then the neighborhoods are encoded as if the neighborhoods

are greater than the center pixel it is assigned as 1, elsewhere it is assigned as 0.The decimal equivalent for the encoded pixels are calculated as 60.

The LGP can be expressed as follows,

$$LGP(x_c,y_c)=\sum_{i=0}^{7} S(g_n-\bar{g})2^i \quad (3)$$

Where $g_n$is the Average gradient value of the neighborhood pixels, $\bar{g}$ is the Adapted threshold value of neighborhood pixels and (x_c,y_c) be the Center pixel position. The function S(x) can be expressed as follows,

$$S(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{otherwise} \end{cases} \quad (4)$$

From the above definition, the LGP can generates the codes by thresholding the absolute intensity difference with an adapted threshold ($\bar{g}$), while LBP generates the codes by thresholding the signed intensity difference between two pixels with a fixed value. When the intensity levels of both the background and the foreground are changed together, LGP and LBP both generate invariant patterns. However, when the intensity level of the background or the foreground is changed locally, LGP generates invariant patterns, but LBP generates variant patterns. Thus 3D-LGP is efficient for scene classification.

D. Haar Wavelet Transformation

The Haar wavelet [5] was introduced by Haar in 1910. It is a bipolar step function. Haar is chosen over other wavelets due to its simplicity and computational efficiency. The 2D Haar wavelet transform is defined as the projection of an image onto the 2D Haar basis functions, which are formed by the tensor product of the one dimensional Haar scaling and wavelet functions. The Haar wavelet function ψ(x) is defined below in equation 5,

$$\psi(x) = \begin{cases} 1, & 0 \le x < 1/2 \\ -1, & 1/2 \le x < 1 \\ 0, & \text{Otherwise} \end{cases} \quad (5)$$

The Haar wavelet is discontinuous in time. The Haar wavelets are generated from the mother wavelet by scaling and translation function as,

$$\psi_{i,j}(x)=2^{i/2}\psi(2^i x-j) \quad (6)$$

The Haar wavelet transformed images reveal both local and shape information. The Haar wavelet transform [5], which extracts local information by means of enhancing local contrast, is applied to every component image of the color image and its three 3D-LBP and 3D-LGP color images. The image in the upper left quadrant of the Haar wavelet transformed image is a lower resolution version of the original image while the other three quadrants contain the high-frequency information from the images along separate orientations. This is the second step of generating the H-Descriptor in which the Haar wavelet transform of each of the three 3D-LBP and 3D-LGP color images are carried out.

E. Structural Feature Extraction

Structural feature is the one of the important type of feature which gives the shape based arrangement of each pixel in an image. The structural feature can be represented by Histogram Of Gradient. H-Descriptor with LBP and H-Descriptor with LGP involves same steps to calculate the HOG.

F. HOG

Histogram of Oriented Gradients [8] mainly used to encode local and shape information of the Haar wavelet transformed images. The idea of HOG rests on the observation that local object appearance and shape can often be characterized well by the distribution of local intensity gradients or edge directions. Since 3D-LBP and Haar wavelet transform both work towards enhancing edges and other high-frequency local features, the choice of HOG as the next step seems logical as an image with enhanced edges is likely to yield more shape information than an unprocessed image. HOG features are derived based on a series of well-normalized local histograms of image gradient orientations in a dense grid. In particular, the image window is first divided into small cells. For each cell, a local histogram of the gradient directions or the edge orientations is accumulated over the pixels of the cell. All the histograms within a block of cells are then normalized to reduce the effect of illumination variations. The blocks can be overlapped with each other for performance improvement. The final HOG features are formed by concatenating all the normalized histograms into a single vector. In this method the image is divided into 3×3 parts or grids and each histogram divides the gradients into nine bins. That makes the HOG vector 81 elements long. In the case of a color image, this process could be repeated separately for the three components images and then concatenate the histograms. The length of a color HOG feature vector is 81×3, i.e. 243.

The HOG descriptors could be derived from the four quadrants of a Haar wavelet transformed image and then concatenate them to get the HOG descriptor of a Haar wavelet transformed image. Fig. 5 shows a color Haar wavelet transformed image, its four quadrant color images, their HOG descriptors, and the concatenated HOG descriptor. Finally integrate the HOG descriptors from the Haar wavelet transform of the component images of the color image and its 3D-LBP and 3D-LGP color images to form the H-descriptor, which encodes color, texture, shape, and local information for object and scene image classification. In particular, for a color image, the 3D-LBP and 3D-LGPdescriptors first generate three new color images. The Haar wavelet transform then produces twelve wavelet transformed images from the twelve color component images of the color image and its three 3D-LBP, 3D-LGP color images. The HOG process further generates four HOG descriptors corresponding to each of the Haar wavelet transformed images. The HOG descriptors from all the Haar wavelet transformed images are finally concatenated to form a new descriptor, the H-descriptor. This is the third step of generating the H-Descriptor in which the Histogram of Oriented Gradients is derived from the Haar wavelet transformed images.

G. H-Descriptor

The 3D-LBP [24] and 3D-LGP descriptor can improves upon the conventional LBP method by means of encoding both

color and texture information of a color image. The H-descriptor is useful and important features for object and scene image classification because of it can extract shape and local features. After calculating the 3D-LBP and 3D-LGP, the Haar wavelet transform of the color image and its new 3D-LBP and 3D-LGP images are extracted. Histogram of Oriented Gradients (HOG) of the Haar wavelet transformed images can encode both shape and local features. And then finally integrate these HOG features corresponding to the Haar wavelet transform of both the original color images and the 3D-LBP, 3D-LGP color images to form the H-descriptor, which encodes color, texture, shape, and local information for object and scene image classification. The dimensionality of this descriptor is 3888 which is the product of the size of the grayscale HOG vector and the total number of quadrants from all the twelve component images of the four Haar transformed color images (81×4×12). The time taken to compute the H-descriptor [24] from an image is empirically seen to be directly proportional to the number of pixels in the image.

IV. SVM CLASSIFIER

A. Introduction

Support Vector Machines are considered to be among the best classification tools available today. Many experimental results achieved on a variety of classification tasks complement the highly appreciated theoretical properties of SVMs. However, there is one property of SVM learning algorithm that has required, and still requires, special attention. Classification in SVM is an example of Supervised Learning. Known labels help indicate whether the system is performing in a right way or not. This information points to a desired response, validating the accuracy of the system, or be used to help the system learn to act correctly.

A classification task usually involves with training and testing data which consist of some data instances. Each instance in the training set contains one target values and several attributes. The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes.

B. SphereSVM

Sphere SVM [22] is the new fast classification algorithm which is similar to minimal enclosing ball approach but it varies only in updating the radius of the center of the sphere. The sphereSVM algorithm is a reformulation of the Ball Vector Machine (BVM). Therefore, some parts of both algorithms are similar. The initialization procedure and the way of finding the violating vectors are same. The important difference is the way updates the center are performed. In the case of a BVM algorithm, all weights α_i corresponding to vectors $\tilde{x}i$ belonging to the data points are modified in each updating step. In the SphereSVM algorithm only two weights α_v and α_u are updated. The first weight α_v corresponds to the vector that is furthest from the ball center while the other weight α_u belongs to the support vector closest to the center.

During the initialization part of the algorithm the random support vector is chosen and its weight is initialized to 1. The approximate center can be represented as follows,

$$C=\sum_{i=1}^{m} \alpha_i x_i \quad (7)$$

Where, C is the approximate center, $\alpha_i$be the weight factor and x_i be the input feature vector.

Then, the radius of the enclosing ball is estimated as,

$$R=\sqrt{\zeta} \quad (8)$$

Where, $\zeta=\sqrt{2+1/C}$

Where, ζ is the square norm of the input vectors x_i. When the size of the dataset and the dimensionality of the feature space are large, the difference between R and $\tilde{R}$ is negligible. Similarly, two violating vectors are selected for each iteration. First, a random subset x_r of size N_r is drawn from the entire dataset. Then, from among the vectors x_i x_r a vector x_v is chosen, whose distance from the center of the ball C is greater than $(1+\tilde{\varepsilon})\tilde{R}$.Where $\varepsilon\in[0,1]$ and $\tilde{\varepsilon}$=0.5.

If such vector is not found in subset x_r, another random subset is selected and the search is performed again. Drawing the x_r subset may be repeated up to Na times. If after Na times there is no outlier satisfying the condition$\|C-xi\|>(1+\tilde{\varepsilon})\tilde{R}$, the $\tilde{\varepsilon}$ is decreased and the algorithm continues the next iteration. Finally, after violator x_v is selected, searching for another violator x_u begins. In other words, the algorithm searches for a support vector x_u that lies closest to the center of the ball. After the two violating vectors are selected, an update to the center of the ball is performed. The center of the ball is then shifted along the line connecting the two violating vectors. The modified center C' can be represented as,

$$C'=C+\beta(x_v-x_u) \quad (9)$$

The updated center [22] can be seen in fig.6,

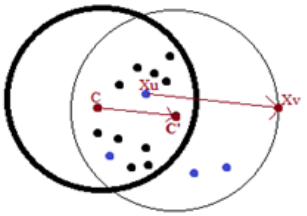

Figure 6.Updation of center in SphereSVM

The β coefficient is selected in such a way that the new sphere centered at C' touches the violator x_v where x_v must be lying on the boundary of the new enclosing ball. Specifically, the following condition must be satisfied,

$$\|C'-x_v\|=\tilde{R} \quad (10)$$

Substituting (9) into (10) will give,

$$\|C+\beta(x_v-x_u)-x_v\|^2=\tilde{R}^2 \quad (11)$$

The above equations can be reduced as,

$$\tilde{\beta}=\rho-\sqrt{\rho^2-\frac{\|x_v-C\|^2-\tilde{R}^2}{\|x_v-x_u\|^2}} \quad (12)$$

Where ρ is

$$\rho=\frac{(x_v-x_u).(x_v-C)}{\|x_v-x_u\|^2} \quad (13)$$

In the dual space, it is equivalent to the increase of α_v by β and the decrease of α_u, also by β. In particular, the non-negativity condition of the α_i weights must be fulfilled. Therefore, $\beta \leq 1-\alpha_v$ and $\beta \quad \alpha_u$ must be true. The first of these requirements is always fulfilled. However, one must assure non-negativity of all α_i. For this reason, the β coefficient must be limited from above by the weight α_u.

$$\beta=\min\{\tilde{\beta},\alpha_u\} \quad (14)$$

The center of the Sphere is updated by changing the value β. This approach can improve the accuracy of the classifier. The main advantage of the SphereSVM is that it can improve the accuracy of the classifier as well as it can carry large class of dataset which will not affect the accuracy of the classifier.

C. SphereSVM as Classifier

In final step in scene classification the H-Descriptor with LBP and H-Descriptor with LGP are feed to the system based on supervised learning. The Sphere Support Vector Machine is a supervised classifier which looks for an optimal hyper plane as a decision function. In the training process, firstH-Descriptor with LBP and H-Descriptor with LGP, i.e., the structural and textural features are extracted to obtain the image representation. When the structural feature is extracted, it is represented by using the H-Descriptors. Texture feature are represented by using 3D-LBP and 3D-LGP. Regarding structural and textural features as two independent feature channels, then combine them to arrive at the final decision. Finally, the multiple category scenes are classified with a SphereSVM. Once the trained images are effectively trained means, the SphereSVM classifier can make decisions regarding the presence of scene, such as a forest, city, buildings etc. in additional to the test images.

V. IMPLEMENTATION AND RESULTS

The performance of the shape and texture based scene classification system is evaluated by using the publically available OT (Oliva and Torralba) [10] dataset. The OT dataset includes 2,688 images of eight different categories. Total number of images in each category as, coast category contains 360 images, forest category contains 328 images, mountain category contains 374 images, open country category contains 410 images, highway category contains 260 images, inside city contains 308 images, street category contains 292 images and tall building category contains 356 images are collected to form OT dataset. The image belongs to each category of same size is 256×256. By proper selection of testing and training image, better recognition rate can be achieved.Some of the example image from each category of OT dataset as follows,

Figure 7. Example images from OT data sets

In the eight categories of OT dataset 25 images from each category is taken as training image and another 35 images of each category is taken into test image. After selecting the test and train images randomly, the shape and textural features are extracted from test and train images separately. Then by combiningH- Descriptor with LBP and the H- Descriptor with LGP as two independent feature channels to categories using anSphereSVM classifier. Then the performance was evaluated by selecting test image of different random split, and the average for the multiple experiments was taken as the final result.

The performance is depicted using a confusion table with all performance values quoted as the average of the diagonal entries in the confusion table. In the confusion table, the *y*-axis represents the ground truth categories of scenes and the *x*-axis represents the category of scenes obtained using our method. The scene categories are consistently ordered on both axes. Each value on the diagonal of the confusion matrix is the classification accuracy for a scene category. Each value on the diagonal are also called as the correctly classified image. And the value except the diagonal are called as the misclassified image. The classification accuracy is defined as the ratio between the total images which is correctly classifiedfrom the test image to the total number of test image selected. The classification accuracy is shown in figure 8.

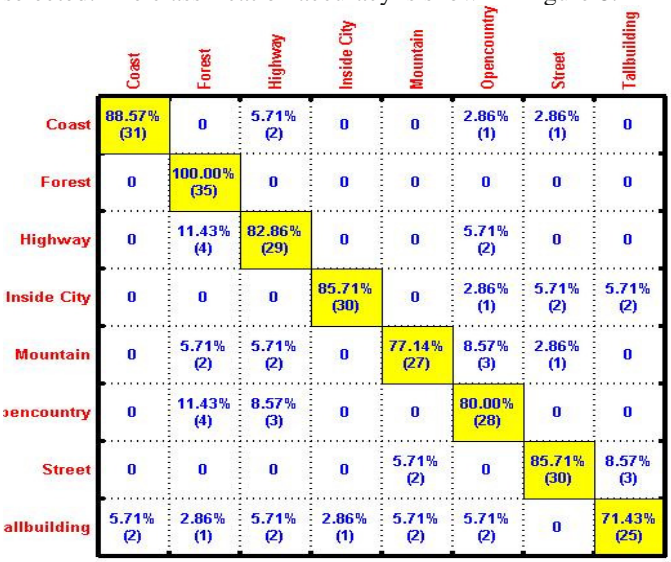

Figure 8. Confusion Matrix for OT Dataset

From the above fig. 8 the performance of the scene recognition system using SphereSVM classifier with 25 training and 35 testing images can be easily visualized. In the category of Coast about 31 images out of 35 are correctly classified and 3 images are misclassified. And in the case of Forest category all images are correctly classified. Only 29 images from the Highway and 30 images from Inside city categories are correctly classified. Similar to that Mountain, open country, Street and Tall building categories are correctly classified as 27, 28, 30 and 25, and other images from above categories are misclassified. Finally 235 images out of 280 images are correctly classified and only 45 images are misclassified. Thus the overall accuracy of 83.92% can be achieved by the SphereSVM classifier by combining H-Descriptor with LBP and H-Descriptorwith LGP features.

A. PERFORMANCE COMPARISON

The performance comparisons are summarized in Table 4.1. The table comprises three different enhancement schemes. By comparing with the performance of the SphereSVM classifier with different descriptors, the performance achieved by the WHOG+Gabor feature along with SphereSVM classifier achieves the recognition rate of 68%. LGP+H-Descriptor along with SphereSVM classifier achieve the recognition rate of 82.3%. Finally theH-Descriptor with LBP+H-Descriptorwith LGP along with SphereSVM classifier achieve the recognition rate of 83.92%. The overall recognition rate of the scene classification can be calculated by the scene categories which are correctly classified to the total number of the test images applied to categories.

TABLE 1. Performance Comparison

Enhancement Scheme	Recognition Rate of SphereSVM Classifier(%)
WHGO+Gabor Filter	68
H-Descriptor+LGP	82.85
H-Descriptor with LBP+H-Descriptor with LGP	83.92

The proposed scene recognition system consists of two major feature extraction namelyH-Descriptorwith LBP andH-Descriptor with LGPare finally classified by using SphereSVM classifier. Structural features are represented by using H-Descriptor and the textural features are represented by using Local Gradient Pattern and Local Binary Pattern. After extracting the two features and classified by using SphereSVM gives better recognition rate which is about 83.92% and the misclassification rate is less compared with other methods.

VI. CONCLUSION

This shape and texture based natural scene classification is simple and yet effective method of scene classification. This method combines structural and textural features. H-Descriptor can effectively extract structural arrangement of pixels in an image and textural representation. First, the original image is splitted into three different slices and LBP, LGP are applied for each slices separately to obtain 3D-LBP and 3D-LGP.Then Haar wavelet is applied for the input image and three different slices then HOG is applied for each Haar wavelet transformed images to produce H-Descriptorwith LBP andH-Descriptor with LGP. Then by taking theH-Descriptor with LBP andH-Descriptor with LGPas two independent feature channels, and combined them to arrive at a final decision using SphereSVM. It is experimentally concluded that theH- Descriptor with LBP and the H- Descriptor with LGP has superior scene recognition performance for the SphereSVM classifier. Experimental results for the publicly available OT datasets including indoor and outdoor scenes show that our method performs well against previous methods across all datasets.

Shape and texture based features are efficient for scene classification. The accuracy or the recognition rate of the scene classification can be effectively improved by extracting some low level features combined it by using H-Descriptor and Extended HOG at the output of the H-Descriptor is the future work. Then SphereSVM classifier can be employed to achieve better recognition rate in shape and texture based scene classification.

REFERENCES

[1] Freeman W.T and Roth M, "Orientation histograms for hand gesture recognition",Intl. Workshop on Automatic Face and Gesture- Recognition, IEEE Computer Society, Zurich, Switzerland, Page No(296-301), 1995.

[2] Szummer M and Picard R, "Indoor-outdoor image classification", In: IEEE Workshop on Content-based Access of Image and Video Databases, Bombay, India, Page No(42–51),1998.

[3] Schmid C, "Constructing models for content-based image retrieval", In: IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, Page No (39–45),2001.

[4] Oliva A and Torralba A," Modeling the shape of the scene: a holistic representation of the spatial envelope", IJCV 42(3),PageNo (145–175),2001.

[5] Porwik P and Lisowska A, "The haar wavelet transform in digital image processing: its status and achievement", Mach. Graph. Vis. 13, Page No (79–98),2004.

[6] Koch C, Li Fi and Van Rullen R, "Why does natural scene categorization require little nattention? Exploring attentional requirements for natural and synthetic stimuli", Visual Cognition, Page No (893–924), 2005.

[7] Lazebnik S, Ponce J and Schmid C, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories", In: IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, Page No (2169–2178),2005.

[8] Dalal N and TriggsB,"Histograms of oriented gradients for human detection", In: IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA, Page No (886-893), 2005.

[9] Sun N, Zheng W, Sun C, Zou L, and Zhao C, "Gender Classification Based on Boosting Local Binary Pattern", Proc. Int'l Symp. Neural Networks, Page No (194-201), 2006.
[10] Oliva A and TorralbaA,"The role of context in object recognition", Trends in Cognitive Sciences, 11(12), Page No (520–527), 2007.
[11] Schiele B and Vogel J," Semantic model of natural scenes for content-based image retrieval", International Journal for Computer Vision, Page No (133–157), 2007.
[12] Zhang J, Marszalek M, Lazebnik S, "Local features and kernels for classification of texture and object categories:a comprehensive study", International Journal for Computer Vision, Page No (213–238),2007.
[13] Huang K.Q, Huang Y.Z and Tao D.C,"Enhanced biologically inspired model", In: IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, USA, Page No (1–8),2008.
[14] Bosch A, Munoz X and ZissermanA,"Scene classification using a hybrid generative/discriminative approach", IEEE Transaction on Pattern Analysis Machine Intelligence, Page No (712–727), 2008.
[15] X. Wang, T.X. Han, and S. Yan, "An HOG-LBP Human Detector with Partial Occlusion Handling", Proc. DAGM Symp. Pattern Recognition, Page No(82-91),2008.
[16] H. Bay, A. Ess, T. Tuytelaars, and L. Gool, "Surf: Speeded Up Robust Features", Computer Vision and Image Understanding, vol. 110, no. 3, Page No (346-359),2008.
[17] Beck M,Caddigan E, Fei-Fei, Walther D," Natural scene categories revealed in distributed patterns of activity in the human brain", Journal of neuroscience, 29, Page No (73-81),2008.
[18] Hebert,Pantofaru C and Schmid C , " Object recognition by integrating multiple image segmentations", In: Proceedings of the European Conference on Computer Vision, Morseille, France, Page No (481–494),2009.
[19] Verma A, Banerji S and Liu C , "A new color SIFT descriptor and methods for image category classification", in: Proceedings of the International Congress on Computer Applications and Computational Science, Singapore, Page No (819–822),2010.
[20] Tanveer Syeda-Mahmood, David Beymer, and Fei Wang, "Shape-based Matching of ECG Recording", (IJCSE) International Journal on Computer Science and Engineering Vol.02, No.07,Page No (2502-2505),2010.
[21] Hu DeWen,Zhou Li and Zhou ZongTan ,"Scene recognition combining structural and textural features", In proceedings of the National University of Defense Technology, China, Science china, Page No (1-14),2012.
[22] BeataStrack, Qi Li,RobertStrack and Vojislav Kecman," Sphere Support Vector Machines for large classification tasks", Neurocomputing, Elsevier. Vol 101, Page No (59-67), 2013.
[23] BongjinJun, DaijinKim and Inho Choi , " Local Transform Features and Hybridization for Accurate Face and Human Detection", IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 35, No. 6, Page No (1423-1436),2013.
[24] Atreyee Sinha, ChengjunLiu and SugataBanerji, "New image descriptors based on color, texture, shape, and wavelets for object and scene image classification", Neurocomputing, Elsevier. Vol 117, Page No (173-185), 2013.

AUTHORS PROFILE

BibinPrasad.M currently Doing M.E Applied Electronics Final year in the Department of ECE at Regional Center of Anna University, Tirunelveli Region, Tirunelveli, Tamil Nadu. The B.E degree in ECE from Sun college of Engineering and Technology, Nagercoil, Tamilnadu in 2012. His research interests include Digital Image processing, Computer Vision, Object Recognition.

Usha Kingsly Devi K is an Assistant professor in Regional Center Anna University: Tirunelveli region. Her research interests include computer vision, image analysis and recognition and pattern recognition.