Image Classification based on Saliency Driven Nonlinear Diffusion and Multi-scale Information Fusion

Ms. Swapna R. Kharche[1], Prof.B.K.Chaudhari[2]

[1]*M.E. student, Department of Computer Engg, VBKCOE, Malkapur (M.S.), India*
[2]*Assistant Professor, Department of Computer Engg, VBKCOE, Malkapur (M.S.), India*

---***---

Abstract: *The saliency driven nonlinear diffusion and multi-scale information fusion preserves or even enhances semantically important structures such as edges, lines, or flow-like structures in the foreground, inhabits and smoothed out the background. Our algorithm considers foreground features; backgrounds which contain only clutter provide no information need to filter out to increase the performance of image classification. The background image regions, whether considered as contexts of the foreground or noise to the foreground, can be globally handled by fusing information from different scales. We propose a multi-scale approach based on Gaussian pyramid representation, which drives the sampling process to ensure separability of the variance from the background clutter.*

Key words: **Image classification, Saliency detection, Nonlinear diffusion, Feature extraction, Information fusion.**

1. INTRODUCTION

Image classification is one of the most challenging problems in computer vision, especially in the occurrence of intra-class variation, clutter, occlusion and pose changes. Image classification refers to the labelling of images into one of the predefined categories. There are various approaches for solving this problem such as k nearest neighbor (KNN), Adaptive boost (Adaboosted), Artificial Neural Network (NN), Support Vector Machine (SVM).

Classification includes image sensors, image preprocessing, object detection, object segmentation, feature extraction and object classification. Classification system consists of database that contains predefined patterns that compares with detected object to categorize in to proper category. Image classification has exhibited considerable progress. Image classification covers a wide variety of application areas such as handwritten digit recognition, face recognition, scene recognition and even human computer interaction. It has motivated researches in many areas including feature extraction, feature fusion, visual code book and classifiers.

Classification process consists of following steps:

A. **Pre-processing-** Atmospheric correction, noise removal, image transformation, main component analysis etc.

B. **Detection and extraction of a object-** Detection includes detection of position and other characteristics of moving object image obtained from camera and in extraction from the detected object estimating the trajectory of the object in the image plane.

C. **Training-** Selection of the particular attribute which best describes the pattern.

D. **Classification of the object-** Object classification step categorizes detected objects into predefined classes by using suitable method that compares the image patterns with the target patterns.

In image classification it is very difficult to deal with background information. Sometimes the spatial context information may help to detect object. Previous approaches for image classification were not considered the background information as significant for the classification. The background of the image gives the context information. Spatial contexts were used to correct some of the labels in classification based on object co-occurrence. The background image regions, whether considered as contexts of the foreground or noise to the foreground, can be globally handled by fusing information from different scales. Saliency driven image multi-scale nonlinear diffusion filtering can be used for this classification process. The background clutters are to be filtered out and background context are used for improving the image classification. For the effective classification the proposed system deals with background information, which uses a saliency driven nonlinear diffusion filtering to create a multi-scale space.

The saliency driven nonlinear multi-scale representation has several advantages. First, the nonlinear diffusion-based multi-scale space can preserve or enhance semantically important image structures at large scales.

Second, this saliency driven multi-scale representation can deal with the background information no matter whether it is a context or noise, and then can be tailored to backgrounds which change over time. Finally, this saliency driven multi-scale representation can be easily united with any existing image classification algorithms (e.g. bag-of-words).

2. RELATED WORK

Image classification is a very active research topic which has accelerated researches in many important areas of computer vision, including feature extraction and feature vision, the generation of visual vocabulary, the quantization of visual patches to generate visual words pooling methods and classifiers.

Liu, Yuan, Sun et al [2] formulates salient object detection problem as a binary labelling task where we separate the salient object from the background. Here they proposed a set of novel features, including multiscale contrast, center-surround histogram, and color spatial distribution, to describe a salient object locally, regionally, and globally. Further, extend a proposed approach to detect a salient object from sequential images by introducing the dynamic salient features.

Zhang et al. [3] experimentally analyzed the influence of the background may have correlations with the foreground objects, using both the background and foreground features for learning and recognition yields less precise results than using the foreground features alone. Overall, the background information was not significant to image classification.

Galleguillos et al. [4] proposed an algorithm that uses spatial context information to classify image. The input image was first segmented into regions and each region was labeled by a classifier. Then, spatial contexts were used to correct some of the labels based on object co-occurrence. The result shows that combining co-occurrence and spatial contexts improves the classification performance.

Nilsback et al. [9], proposed a method to scope "bag of visual words" models to differentiate categories which have significant visual similarity. They demonstrated that by developing a visual vocabulary that explicitly represents the various aspects of color, shape, and texture that differentiate one flower from another, can overcome the ambiguities that exist between flower categories. This method is based on nearest neighbor classifier architecture. Experiments showed that the color feature has a performance of 73.7%, and the shape features attained a performance of 71.8%.

M. Varma and D. Ray[10] investigated the problem of learning optimal descriptors for a given classification task. They focused on learning the optimal trade- off for classification given a particular training set and prior constraints. The problem is posed in the kernel learning framework. They learn the optimal, domain-specific kernel as a combination of base kernel corresponding to base features which attain different levels of trade-off (such as no invariance, rotation invariance, scale invariance, affine invariance, etc.) This leads to a convex optimization problem with a unique global optimum which can be solved for efficiently.

N. Xie, H. Ling, W. Hu, and X. Zhang [11] proposed using bin-ratio information, which is composed from the ratios between bin values of histograms, for scene and category classification. To use such information, new histogram dissimilarity, bin-ratio dissimilarity (BRD), is calculated. They show that BRD provides several attractive advantages for category and scene classification tasks: First, BRD is robust to cluttering, partial occlusion and histogram normalization; Second, BRD captures rich co-occurrence information while enjoying a linear computational complexity; Third, BRD can be easily united with other dissimilarity measures, such as L1 and x^2, to gather complimentary information. They apply the proposed methods to category and scene classification tasks in the bag-of-words framework.

3. METHODOLOGY

The aim of preprocessing is an improvement of image data that smother unwanted distortions or enhance some image features essential for further processing. Pre-processing could be a method to get free of noises from the linear image.

The global structure of the proposed method is shown in Fig-1. This figure shows different modules and the actions of the proposed work. Image input is given to a saliency map detection method which is proposed then follows a classification System. Then it is used for feature extraction and final classification. The saliency detection carried out to locate the interesting area in target image and feature calculation by SIFT. The multi-scale information fusion carried out to correctly recognize image.

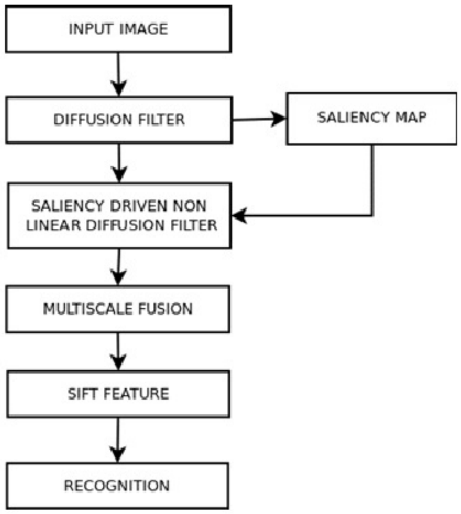

Fig-1: Structure of the proposed method.

1.1 Saliency Detection

The image pixels in the important regions should get high saliency values, and those pixels in the unimportant areas should get lower saliency values. Various image features such as brightness, color may contribute to the saliency measure of pixels. The degree of dissimilarity between the pixels patch and other patches is calculated. Greater dissimilarity generates higher saliency value and vice versa.

Goferman's method[5] is based on context aware saliency detection where the image is separated into patches of 7×7.The color difference between the patches are taken as the saliency value. The color distance between the pixels are calculated in the CIE L*a*b color space instead of the RGB color space, so need translation of RGB to L*a*b color space. This dissimilarity degree can be defined by the patch distance, and the saliency value of pixel can be determined by the dissimilarity degree. Content aware saliency detection consists of the following steps.

1) Patch distance computation: For two image pixels i and j, we first define the patch distance between patch P^i centered at i and patch P^j centered at j as follows:

$$d(p^i,p^j) = \frac{d_{color}(p^i,p^j)}{1+c.d_{position}(p^i,p^j)}$$

(1)

The term $d_{color}(p^i,p^j)$ normalized Euclidean distance between two patches P^i and P^j in CIE L*a*b color space, which is calculated by a quadratic sum of the color differences between the corresponding pixels of two patches. The $d_{color}(p^i,p^j)$ term is the Euclidean distance between the positions of patches i and j, normalized by the larger image dimension.

2) Saliency estimation: For a patch pi, K smallest distance patches are measured $(q_k)_{k=1}^{K}$. So, we can calculate the saliency of pixel i at scale r as follows:

$$s_i^r = 1 - \exp\left(d\left(p_i^r, q_k\right)\right)$$

(2)

The K most similar patches are selected from the whole image, which will lead to the distance $d(p^i,p^j)$ computation between patch pi and all of other patches j in the whole image. Here, for the sake of computing the patch dissimilarity effectively, we border the search range of K most similar patches into the local neighboring of patch i. Fig-2. Shows a saliency map obtained for an input image.

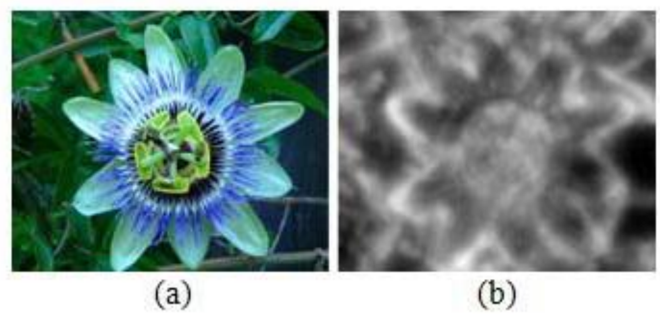

Fig-2: (a) Input image. (b) Saliency map.

Cheng et al. [6] proposed a histogram-based contrast method to measure saliency. Their algorithm separates a large object from its surroundings, and enables the assignment of similar saliency values to homogenous object regions, and highlights entire objects. An image histogram is created by color quantization (median cut algorithm). In the image histogram the color difference between the pixels are computed. The sum of the color difference is taken as the saliency value and each of the pixels are later on replaced by this saliency value. Our framework uses combination of both methods for saliency map detection.

1.2 Nonlinear Diffusion filtering

The nonlinear diffusion preserves and enhances image structures defined by large gradient values. If image structures with large gradients are all in the foreground, nonlinear diffusion filters out the background. However, there may be large image gradients in the background. Saliency driven nonlinear diffusion is process which blurs non-salient region and preserves salient regions.

Let *u(x, y, t)* be the grey value at position *(x, y)* and scale *t* in the multi-scale space. The image diffusion filtering is defined by the diffusion equation [7]:

$$\partial_t u = div\,(D \cdot \nabla u) = div \cdot (D \cdot \nabla u) \quad (3)$$

where ∇ is gradient operator: $\nabla = (\partial/\partial x, \partial/\partial y)$, "*div*" is the divergence operator, and *D* is the diffusion tensor which is a positive definite symmetric matrix.

If the *D* in (3) is a function $g(\nabla u)$ of the gradient ∇u of the evolving image *u* itself, then Equation (3) defines a nonlinear diffusion filter [7], [8].

The nonlinear diffusion filtering is represented as:

$$\partial_t u = div\,(D \cdot \nabla u) = div\,(g\,(\nabla u)\,\nabla u). \quad (4)$$

The D in equation is a function $g(\nabla u)$ of the gradient of the evolving image u itself. The function $g(\nabla u)$ is usually defined as:

$$D = g(\nabla u) = \frac{1}{1+\frac{(\nabla u)^2}{\lambda^2}} (\lambda > 0) \quad (5)$$

The regions in which $\nabla u < \lambda$ are blurred, while the other regions are sharpened. If image structures with large gradients are all in the foreground, nonlinear diffusion filters out the background.

1.3 Saliency Driven Nonlinear Diffusion

In this saliency map is taken as prior knowledge with nonlinear diffusion filtering. Let I_s be the saliency map for the diffusion process. Then combine I_s into D in (2) and define D as a function g of ∇u and I_s. Then, the diffusion equation becomes defined by [1]

$$U(x, y, t) = f(x, y) \text{ if } t = 0$$
$$\partial t\, u = div\,(g\,(\nabla u, Is\, \nabla u) \text{ if } t > 0. \quad (6)$$

Saliency driven nonlinear diffusion preserved foreground regions and largely smoothed the background regions.

1.4 Feature Extraction and Classification

Images whose foregrounds are clearer than their backgrounds are more likely to be correctly classified at a large scale, and images whose backgrounds are clearer are more likely to be correctly classified at a small scale. So, different scales information can be combined to obtain more efficient results of image classification. Fig.3 shows an image at different scales. It is seen that our saliency driven nonlinear diffusion leads to image simplification in the non-salient region, most of the structures in this region are blurred and smoothed. In the salient region, the progression of scales preserves or even enhances semantically important structures, such as edges and lines. The images produced by our saliency driven nonlinear diffusion are more appropriate for image classification than those produced by standard nonlinear diffusion.

Each image is represented by its multi-scale images. Then, for each scale t, scale invariant feature transform (SIFT) features, which are widely used to signify image regions, are extracted, and the bag-of-words model is used to create a word frequency histogram h_t. The dissimilarity between images 1 and 2 at scale t is represented by the $\chi 2$ distance d (h_t^1, h_t^2) between histograms h_t^1 and h_t^2.

Fig-3: Multiscale space of an image.

The distances $\{d(h_t^1, h_t^2)\}t\epsilon T$ between images 1 and 2 obtained at different scales are united to yield the final distance $d(h_1, h_2)$ between images 1 and 2 [1]:

$$d(h1,h2) = \frac{\sum_{t\epsilon T} w_t d(h_t^1, h_t^2)}{\sum_{t\epsilon T} w_t} \quad (7)$$

where w_t is a weight for scale t, and T is a chosen set of scales. Weighted averaging, which is a general way for information fusion, is used to combine information from different scales. The final distance d (h1, h2) between images obtained by combining the distances at the three scales, it transformed to a kernel which is used by an SVM for classification. Used to the extended Gaussian kernels:

$$K(h1, h2) = \exp\left(-\frac{1}{A} d(h1, h2)\right) \quad (8)$$

where A is a scaling parameter that can be determined by the cross-validation. An SVM classifier is trained using the kernel matrix of the training images.

Support vector machine is applied to the images. The feature values calculated between the test feature and the train feature. The image corresponding to the feature having minimum distance is collected from the database. In addition to performing linear classification, SVMs can efficiently execute a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

4. EXPERIMENTAL RESULT

The performance of Image Classification system can be determined in terms of its Recognization Rate.

$$Recognization\ Rate = \frac{Number\ Of\ images\ correctly\ classified}{Total\ number\ of\ images}$$

Tested the image classification algorithm on flower category by different methods and the proposed saliency driven nonlinear diffusion filtering and multiscale information fusion:

TABLE-I: Recognization Rates for flower category by different methods

Methods	Recognization Rate (%)
Nilsback and Zisserman [9]	71.76±1.76
Varma and Ray [10]	82.55±0.34
Xie[11]	89.02±0.60
Khan[13]	89
Gehler and Nowozin [14]	85.5±1.2
Y. Chai [15]	90.40±2.3
Scale 0	87.45±1.13
Scale T_m	87.69±1.61
Scale T_M	88.21±1.19
Proposed Method	93.75±2.0

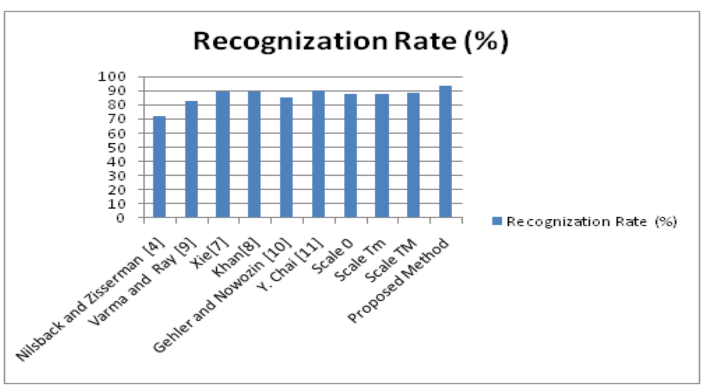

Chart-1: Graphical Representation of results on flower category

Table-II: Recognization Rates for bike, cycle person and car categories by different methods

Methods	Bike	Cycle	Person	Car	Average
Winner(x^2[12]	79.8	72.8	71.9	72.0	74.1
Winner(EMD)[3]	79.7	68.1	75.3	74.1	74.3
PDK[16]	76.9	70.1	72.5	78.4	74.5
Xie[11]	79.1%	75.4	73.9	78.2	76.7
Proposed Method	77.4	77.7	77.2	80.0	78.1

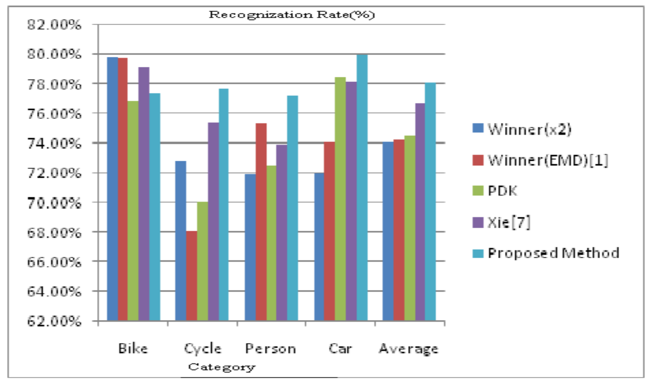

Chart-2: Graphical Representation of results on bike, cycle, person and car categories

Below Fig.6 show diffusivity function value with respect to different values of parameters c, m and λ. When λ is too small, neither the foreground nor the background are filtered; when λ is too large, both the foreground and the background are filtered; when λ is appropriately chosen, the background is smoothed and the foreground is conserved. If *m* is small, semantically important structures are filtered out, no matter whether λ is suitable or not. When *m* is small, there is a broad transitional zone from 1 to 0 in the value of the diffusivity function. Those edges, whose gradients magnitudes are in the transitional zone, are partially filtered out. Consequently, the transitional zone should be narrow, and *m* should be large. When λ is chosen correctly and *m* is large enough, *C* has little effect. As a result, *C* is considered as constant.

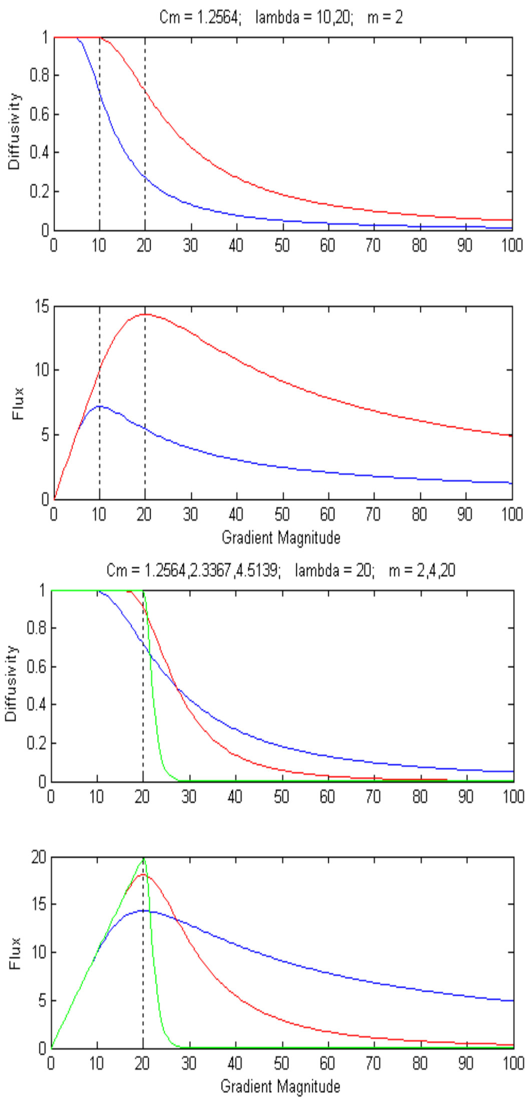

Chart-3: Diffusivity Function value with respect to different values of c, λ and m.

5. CONCLUSION

Saliency driven multi-scale nonlinear diffusion filtering enhance the classification of images using nonlinear diffusion filtering and help to find out the diffusion parameters using the saliency detection results. We have further applied this new method to image classification. The saliency driven nonlinear multi-scale space preserves and even enhances important image local structures, such as lines and edges, at large scales. Multi-scale information has been combined using a weighted function of the distances between images at different scales. The saliency driven multi-scale representation can incorporate information about the background in order to progress image classification results. The experimental results have demonstrated that proposed system based saliency driven multi-scale information fusion improves the accuracy of image classification. Video classification can be carried out using the proposed method so that system can be used for video surveillance application.

REFERENCES

[1] W. Hu, R. Hu, H. Ling, N. Xie, and S. Maybank, Image Classi_cation Using Multi-scale Information Fusion Based on Saliency Driven Nonlinear Diffusion Filtering, IEEE Trans.on Image Processing, vol. 23, no. 4,Apr. 2014,pp. 1513-1526.

[2] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, et al., "Learning to detect a salient object," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33,no. 2, Feb. 2011, pp. 35–367.

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

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

[5] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 2376–2383.

[6] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. L. Huang, and S. M. Hu, "Global contrast based salient region detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Dec. 2011, pp. 409–416.

[7] J.Weickert, A review of nonlinear diffusion filtering, in Scale- Space Theory Compute Vision (Lecture Notes in Computer Science). New York, NY, USA: Springer-Verlag, 1997, pp. 128

[8] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 7, Jul. 1990, pp. 629–639.

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

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

[11] N. Xie, H. Ling, W. Hu, and X. Zhang, "Use bin-ratio information for category and scene classification," in Proc. IEEE Conf. Comput. Vis.Pattern Recognit., Jun. 2010, pp. 2313–2319.

[12] M. Everingham, A. Zisserman, C. K. I. Williams, L. van Gool, M. Allan, C. M. Bishop, et al., "The 2005 PASCAL

visual object classes challenge," in Proc. 1st PASCAL Challenges Workshop, 2006, pp. 117–176.

[13] S. Khan, J. van de Weijer, and M. Vanrell, "Top-down color attention for object recognition," in Proc. IEEE 12th Int. Conf. Comput. Vis., Oct. 2009, pp. 979–986.

[14] P. Gehler and S. Nowozin, "On feature combination for multiclass object classification," in Proc. IEEE 12th Int. Conf. Comput. Vis., Oct. 2009, pp. 221–228.

[15] Y. Chai, V. Lempitsky, and A. Zisserman, "BiCoS: A Bi-level cosegmentation method for image classification," in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 2579–2586.

[16] H. Ling and S. Soatto, "Proximity distribution kernels for geometric context in category recognition," in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2007, pp. 1–8.