ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

SALIENCY BASED IMAGE CATEGORY RECOGNITION

Parvathy Ashok and Vineetha K. V.
Model Engineering College, Ernakulam, Kerala, India
E-Mail: parvathyashokmec@gmail.com

ABSTRACT

A novel approach using visual attention technique is proposed for classification that categorises objects in challenging conditions. 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 labelling of images into one of the predefined categories. In image classification it is very difficult to deal with background information. 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 resulting scale space in general preserves important structures such as edges, lines in the foreground, and inhibits and smoothes clutter in the background.

Keywords: saliency, nonlinear diffusion filtering, classification.

INTRODUCTION

Image classifications have achieved significant progress in recent years. By modelling the spatial configurations among local features, extracting informative image representations, modelling contextual relationships among attributes or objects, and learning powerful classification models, Image classification has exhibited significant progress. Image classification is one of the most important problems for computer vision and machine learning. This covers a wide variety of application areas such as handwritten digit recognition, face recognition, scene recognition and even human computer interaction. It has stimulated researches in many areas including feature extraction, feature fusion, visual code book and classifiers. In image classification, it is an important but difficult task to deal with the background information. Sometimes the spatial context information may help to detect object. Previous approaches for image classification were not considered the background information as relevant 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 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 drove nonlinear diffusion filtering to generate a multi-scale space.

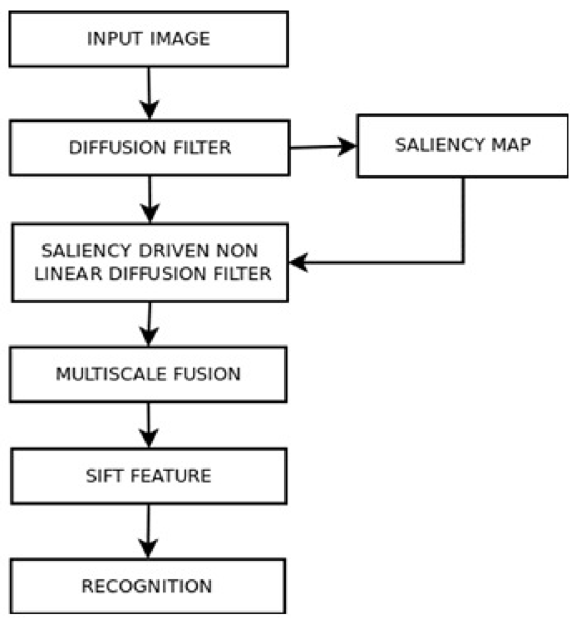

Figure-1. Structure of the proposed method.

This saliency driven nonlinear multi-scale image representation has several advantages.

- The In the nonlinear scale space, important image structures are enhanced or preserved at large scales. After diffusion at any scale, the locations of the important image structures are not shifted. This differs from the Gaussian scale space in which important image structure parts may be smoothed and shifted from their true locations after Gaussian convolution.
- The background image regions can be partly dealt with by fusing information from different scales, no matter whether the background is a context for the foreground
- This saliency driven multi scale representation can be easily given as input to any existing image classification algorithms.

BACKGROUND

The intent of the classification process is to categorize all pixels in a digital image into one of several land cover classes, or "themes". This categorized data may then be used to produce thematic maps of the land cover present in an image. Normally, multispectral data are used to perform the classification and, indeed, the spectral pattern present within the data for each pixel is used as the numerical basis for categorization. The objective of image classification is to identify and portray, as a unique gray level (or color), the features occurring in an image in terms of the object or type of land cover these features actually represent on the ground. Image classification is perhaps the most important part of digital image analysis. It is very nice to have a "pretty picture" or an image, showing a magnitude of colors illustrating various features of the underlying terrain, but it is quite useless unless to know what the colors mean. Two main classification methods are Supervised Classification and Unsupervised Classification.

The diffusion process can be seen as an evolution process with an artificial time variable t denoting the diffusion time where the input image is smoothed at a constant rate in all directions. Starting from the initial image $u_0(x)$, the evolving images $u(x, t)$ under the governed equation represent the successively smoothed versions of the initial input image f (x), and thus create a scale space representation of the given image f , with $t > 0$ being the scale. As we move to coarser scales, the evolving images become more and more simplified since the diffusion process removes the image structures at finer scales.

In multi-scale image representation for each scale t, scale invariant feature transform (SIFT) features, which are widely used to represent image regions, are extracted, and the bag-of-words model is used to generate a word frequency histogram [1]. Recently, several local features, e.g., SIFT and HOG, are quite popular in representing images due to their ability to capture distinctive details. After that Harris corner detector has been the first widely used detection algorithms which addressing the limitations of this operator through improve its capability of noise immunity and invariance to intensity change and rotation. SIFT algorithm proposed by Lowe has been considered as one of the most robust approaches [2]. The SIFT features are invariant to:1) scale; 2) rotation; and 3) illumination, which are essential in many applications like robot navigation, image stitching for panoramic photographs, stereo matching, image registration, object detection, and recognition [3].

In fact, the notion of scale is an essential part of early visual processing, where the main task is to separate the image into relevant and irrelevant parts.

For image classification, single scale method is not good to detect the responses and orientation of pixels. Because different images have different widths and there are complicated intersections between lines. Normally the feature includes both the background and fore ground and the back-ground information was not relevant for image classification. So this kind of Recognition yields less accurate result than using the foreground feature alone.

The linear SVM is adopted to classify the images based on individual image representation, or their concatenated image representation. In machine learning, support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier [4].

METHODOLOGY

The global structure of the proposed method is shown in Figure-1. This figure shows different modules and the procedures 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.

Saliency Detection

The image pixels in the significant 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.

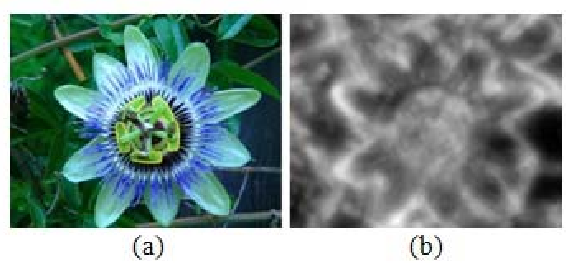

Figure-2. (a) Input image. (b) Saliency map.

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 ^{Pi} centered at ^{i} and patch ^{Pi} centered at ^{j} as follows:

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

Where c is a constant and set to c=3.0 in the implementation. The term $d_{color}(p^i, p^j)$ normalized

VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

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 p^i , K smallest distance patches are considered $(q_k)_{k=1}^{k}$. Set K = 60 in this implementation. So, we can calculate the saliency of pixel i at scale r as follows:

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

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 p^i and all of other patches ^{j} in the whole image. Here, for the sake of computing the patch dissimilarity effectively, we limit the search range of K most similar patches into the local neighboring of patch ^{i} .Figure-2. Shows a saliency map obtained for an input image.

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 equation which blurs non-salient regions 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

$$\delta_t u = div(D.\nabla u) = div(g(\nabla u)\nabla u)$$

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

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

The regions in which ∇ u < 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. However, there may be large image gradients in the background. Thus, here propose a saliency driven nonlinear diffusion equation which blurs non-salient regions and preserves salient regions.

Saliency Driven Nonlinear Diffusion

In this saliency map is taken as priori knowledge with nonlinear diffusion filtering. Let Is be the saliency map for the diffusion process. Then combine Is into D in (6) and define D as a function g of ∇ u and Is . Then, the diffusion equation becomes

$$u(x;\, y;\, t) = f(x;\, y) \qquad \text{if } t = 0$$
$$\delta_t u = div(g(u,\, Is\,)u) \qquad \text{if } t > 0.$$

Figure-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 evolution of scales preserves or even enhances semantically important structures, such as edges and lines. The images produced by our saliency driven nonlinear dif-fusion are more suitable for image classification than those produced by normal nonlinear diffusion. Although these examples are taken from static background and still images, our work can be adapted to time varying background from moving platforms, because our saliency driven nonlinear diffusion filtering can effectively deal with the background information.

Figure-3. Multiscale space of an image.

Feature Extraction and Classification

From the multi-scale image representation foregrounds are clearer than their backgrounds are more likely to be correctly classified at different scale. Then, information from different scales can be fused to acquire more accurate image classification results. Each image is represented by its multi-scale images and for each scale t, scale invariant feature transform (SIFT) features, which are widely used to represent image regions, are extracted, and the bag-of-words model is used to generate a word frequency histogram.

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

CONCLUSIONS

Saliency driven multi-scale nonlinear diffusion filtering, to enhance the classification of images using nonlinear diffusion filtering, and determining 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 fused using a weighted function of the distances between images at different scales. The saliency driven multi-scale representation can include information about the background in order to improve image classification results.

REFERENCES

[1] Weiming Hu, Ruiguang Hu, Nianhua Xie, Haibin Ling and Stephen Maybank. 2014."Image Classification Using Multiscale Information Fusion Based on Saliency Driven Nonlinear Diffusion Filtering", IEEE Transaction on Image processing, Vol. 23, No. 4.

[2] Weisheng Li, Rui Liu and Ying Huang. 2012. "SIFT features of fusion region information entropy in Bag-of-Words", International Conference on Fuzzy Systems and Knowledge Discovery.

[3] Hua Cao and Jiazhong Chen. "Multicore computing for SIFT Algorithm in MATLAB Parallel Environment", in Proc. IEEE 18th International Conference on Parallel and Distributed Systems, pp. 2083-2090.

[4] Jucheng Yang, Min Luo and Yanbin Jiao. 2013."Face Recog-nition Based on Image Latent Semantic Analysis Model and SVM", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 6, No. 3.

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

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

[7] K. Grauman and T. Darrell. 2007. The pyramid match kernel: Efficient learning with sets of features, J. Mach. Learn. Res., Vol. 8, No. 4, pp. 725-760.

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

[9] K. Grauman and T. Darrell. 2007. The pyramid match kernel: Efficient learning with sets of features, J. Mach. Learn. Res., Vol. 8, No. 4, pp. 725-760.

[10] J. Harel, C. Koch and P. Perona. 2007. Graph-based visual saliency, in Proc. Annu. Conf. Neural Inf. Process. Syst., pp. 545-552.

[11] L. Itti, C. Koch and E. Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254-1259.

[12] Yongwei Miao, Ruifeng Han and Huahao Shou. 2012."A Fast Algorithm for Content-aware Saliency Detection and Stylized Rendering", The 2nd International Conference on Computer Application and System Modeling.

[13] J. Weickert. 1997. A review of nonlinear diffusion filtering, in Scale-Space Theory Compute Vision (Lecture Notes in Computer Science).New York, NY, USA: Springer-Verlag, pp. 1-28.

[14] Kumar G. and Bhatia P.K. 2014. "A Detailed Review of Feature Extraction in Image Processing Systems", Fourth International Conference on Advanced Computing & Communication Technologies (ACCT).

[15] Girish Kulkarni, Visruth Premraj, Vicente Ordonez, Sagnik Dhar, Siming Li, Yejin Choi and Alexander C. Berg. 2013. "BabyTalk: Understanding and Generating Simple Image Descriptions", IEEE Transactions on Pattern Analysis and Ma-chine Intelligence, Vol. 35, No. 12.

[16] A. Aker and R. Gaizauskas. 2010. "Generating Image Descriptions Using Dependency Relational Patterns", Proc. 28th Ann. Meeting Assoc. for Computational Linguistics, pp. 1250-1258.

[17] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng and S. Li. 2013. Salient object detection: A discriminative regional feature integration approach, "in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Jun, pp. 2083-2090.