Salient Map Creation and Database Retrieval Based on Salient Features

Nejma M A  
Department of Computer Science and Engineering  
KMEA Engineering College  
Aluva, Ernakulam  
nejmal61@gmail.com

Smita S  
Department of Information Technology  
KMEA Engineering College  
Aluva, Ernakulam  
smithasathyana2@yahoo.com

Abstract—Features of an image can be extracted with the help of salient maps. Feature extraction can be either in spatial domain or in transform domain. In this work, a novel salient map creation model based on wavelet transform domain. At first, the wavelet transform is employed to create the multi-scale feature maps which can utilise low-level features. This saliency detection model is based on high-pass coefficients of the wavelet decomposition after eliminating some high-frequency components of the image. The idea is to create the feature maps by Inverse Wavelet Transform (IWT) on the multi-level decomposition. Then, a model for the saliency map from these features is proposed. Based on these saliency maps common saliency from multiple images is extracted and the matching images based on input images are retrieved. For this matching purpose weakly supervised learning technique will be used. Support Vector Machine (SVM) is actually designed for binary classification. In this paper, Multi-SVM is used for one against all implementation.

Keywords— Gaussian filter; Multi-SVM; Wavelet.

I. INTRODUCTION

Automatically detecting salient regions in images is the basis of many different applications in image processing. There have been all kinds of research for detecting salient regions in still images and videos. The goal of image saliency detection is to determine regions which are particularly noticeable to human observers. Most approaches make the assumption that an object is salient if it significantly differs from other objects in its local neighbourhood. It is common to distinguish between top-down and bottom-up approaches [5]: Top-down attention is straightforward and efficient for task accomplishment. But bottom up attention is inspired by a neuronal architecture of early primate vision. Bottom up is induced by stimuli regarding color, intensity and orientation on several hierarchy levels. Pure bottom-up is the only way to select potentially important information of the environment for further processing rather than concrete top-down information. Top-down and bottom-up selection always work together in human attention to determine the attention allocation and control the human gaze behaviour. If there is a unique target object, a top-down biased bottom-up strategy can help. But, it fails, if a group of objects is searched, presence of those object cannot be uniquely described by using a bottom-up computation model, which are based on low-level features. Consider different traffic signs are all salient in color but in different geometry and with different text on them [3]. They are not distinguishable from each other with the help of bottom up attention. So a bottom-up visual attention mechanism is used to focus on interesting object in the environment.

This paper presents an efficient salient-region extraction algorithm based on the significance of accumulated wavelet coefficients. Wavelets can be used to extract information from many different kinds of data. Wavelet transforms are broadly divided into three. They are continuous, discrete and multi-resolution based [9]. Haar, Biorthogonal, Coiflets, Morlet, Daubechies, Mexican Hat, Symlets, and Meyer wavelets are included in Mat lab wavelet toolbox.

Lab color space is one of the important color spaces like RGB and CMYK color space. Lab color space contains all perceivable colors. One device independent color space is Lab color space. That means, the colors are defined independent of device they are displayed and the nature of creation. Lab gamut includes CMYK and RGB gamut. So Lab used as an interchange format between different devices because of its device independency. Lab color space is designed for approximate human vision [1]. Here L component matches human perception of lightness, but it do not take the Helmholtz–Kohlrusch effect into account. Lab color space can make accurate color balance corrections by modifying a, b components or L (lightness contrast) components.

The Saliency Map is a topographically arranged map that represents visual saliency of a corresponding visual scene. The saliency map combines information from each of the feature maps into a global measure where points corresponding to one location in a feature map project to single units in the saliency map.

Detection of salient regions is useful for applications like object recognition, object segmentation and compression. Now a day, full-resolution salient maps can retain have attracted attention with well-defined boundaries.


II. Motivation

One of the earliest algorithms obtains the saliency map based on the intensity, color, and orientation conspicuity maps. These conspicuity maps are attained by across-scale addition of feature maps, while the feature maps capture the center-surround differences between various Gaussian pyramid and oriented pyramid scales. Since the saliency map is computed in coarser scales, local information loss is unavoidable in this algorithm.

In other method, saliency map is processed by the mean-shift segmentation algorithm to extract the regions of interest in the input image for better performance. This approach is mostly good for images with large and homogeneous object with clear boundaries. Hence, it has limitations on applications because of its dependency on object size and uniformity.

Recent studies have tried to obtain the saliency map for images in the transform domain. The Fourier Transform (FT) can be expressed with the polar form as two different components: phase and amplitude spectrums. Here phase gives information concerning the form and the position of local image structure. While the latter holds the information of the global composition of the image that relates to the overall scene layout [1]. One disadvantage of this model is the high down-sampling requirement for images, which would yield spatial information loss.

Another method based on Wavelet Transform (WT) based salient detector for image retrieval depends on local and global variations of wavelet coefficients at multi-scale levels [9]. The idea is to account for both global and local features in the WT space: the points with higher global variation based on the absolute wavelet coefficients at coarser scales are selected, and these selected points are tracked along the finer scales to detect the salient points. The salient points will be the sum of the wavelet coefficients at tracked points along the multi-scale WT.

SVM is the abbreviations of Support Vector Machine, which used for binary classification. But extension of SVM is Multi-SVM, assign labels to instances by using support vector machines, where the labels are obtained from a finite set of elements. One-against-all approach constructs M binary SVM classifiers, which separates one class from the rest.

III. Overview

In this work, a novel saliency detection model is proposed, based on high-pass coefficients of the wavelet decomposition after eliminating some high-frequency components of the image. The idea is to create the feature maps by IWT on the multi-level decomposition. The advantage of the proposed model is that we create more detailed feature maps (edge to texture) by applying IWT on various decomposition levels. This helps to observe the irregularities with different bandwidths. Then, two saliency maps are created: local and global saliency maps [1].

Here, we create these two saliency maps for two reasons: i) to avoid the normalization of each feature map separately which is not efficient for considering the statistical relation among the feature maps for the saliency in a global perspective; ii) to incorporate local and global saliency as two different maps to make sure of taking both local and global contrast into consideration sufficiently. Finally, the local and global maps are combined to yield the final saliency map.

The proposed saliency detection model includes both local and global saliency information by integrating local feature differences with the global distribution of these features.

The global saliency is based on the likelihood of local features for a given location. The proposed model can be used for images with different object sizes and extents of uniformity. It also generates the saliency map with the same resolution of the input image. Another advantage is that the wavelet decomposition in the proposed model continues until the possible coarsest scale is reached, in order to attain the feature maps. By this way, obtain the salient regions independent from their sizes or uniformity due to the features with high contrast from edge to texture since the salient points can be defined as the feature variations of the location with its surroundings based on the WT coefficients. The first step of the proposed model starts with the computation of feature maps. First of all, instead of using RGB color space for saliency detection, an image is converted to the CIE Lab color space. The conversion is needed due to the fact that the Lab color space is uniform and similar to the human perception, with a luminance and two-chromatic channels. To remove noise, apply a 2D Gaussian low-pass filter to the input color image. The wavelet coefficients representing the details of the image at various scales are used to create several feature maps with increasing frequency bandwidths. The feature maps can be calculated by IWT. This creates 3XN feature maps for an input color image, and each feature map’s resolution is equal to the size of the input image.

Feature maps of L, a, and b channels for each sample images are constructed. There are N feature maps representing the reconstruction results from wavelet coefficients of the 1st-level decomposition to the Nth-level decomposition for the given input images. It should be noted that 8-level reconstruction consists of details from the coarsest scale (the 8th level) to the finest scale (the 1st level) wavelet coefficients, the 7-level reconstruction consists of details from the 7th level to the finest scale (the 1st level) wavelet coefficients, and so on [1]. Then Local and Global saliency map will be created. After that combine these local and global saliency maps, to get final salient map [1]. With the saliency map, the auto-context model is learned and can be used to recognize the same type of objects in the challenging images. In this work, use SVM for learning and predicting the object.

The main application of saliency map is to give attention to special area of images, i.e., to identify the most important information in visual inputs. Image classification and scene recognitions are other applications.

IV. System environment

Here propose saliency map creation model and database retrieval based on this saliency map. The simulation tool used here is mat lab 7.14.0.739. Here introducing the image salient
The proposed saliency map creation model consists of feature maps and saliency maps. Here local and global saliency map will be constructed to get the final saliency map. The entire saliency detection models consist of following section.

A. Pre-processing Step

First of all, instead of using RGB color space for saliency detection, an image is converted to the CIE Lab color space. The conversion is needed due to the fact that the Lab color space is uniform and similar to the human perception, with a luminance and two-chromatic channels (RG and BY). A filter can be applied after changing the color space. A 2D Gaussian low-pass filter is used to remove noise of the input image \( g^c \).

\[
 g'^c = g^c \ast I_{m \times m} \tag{1}
\]

Where \( I \) is the \( m \times m \) 2-D filter. \( g'^c \) is noise-removed version of \( g^c \). \(-\ast I\) denotes the convolution operation. For noise reduction, a small filter size is selected to filter very high frequency noise. Then, each channel is normalized to range of \([0 \ 255]\).

Figure 1. (A) Shows the input image. (B) Shows RGB to Lab conversion of input image A. (C) Shows the filtered images of (B).

A Gaussian filter is a filter, which uses a Gaussian function. A Gaussian filter modifies the input signal by convolution with a Gaussian function; this transformation is also known as the Weierstrass transform. A Gaussians function in two dimensions can be represented as,

\[
 g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}} \tag{2}
\]

Where \( x \) and \( y \) shows distance from the origin in the horizontal axis and vertical axis, \( \sigma \) is the standard deviation of the Gaussian distribution.

B. Feature Map Generation

The sub-bands of the image will be formed by WT for a number of levels. Discrete Wavelet Transform (DWT) wavelets are chosen for this purpose.

\[
 [A_N^C, H_S^C, V_S^C, D_S^C] = WT_N(g'^c) \tag{3}
\]

Where \( N \) is the maximum number of the scaling for WT decomposition process, i.e., the resolution index \( s \in [1, \ldots, N] \) and the \( N^{th} \) level corresponds to the coarsest resolution. \( C \) is the channels of \( g'^c \) as \( C \in \{L, a, b\} \). There is \( A_N^C, H_S^C, V_S^C \) and \( D_S^C \) components. \( A_N^C \) is the approximation output at the coarsest resolution for each channel; \( H_S^C, V_S^C \) and \( D_S^C \) are the wavelet coefficients of horizontal, vertical and diagonal details for the given \( c \) and \( s \), respectively. The wavelet coefficients representing the details of the image at various scales are used to create several feature maps with increasing frequency bandwidths. The feature maps can be calculated by IWT. Since here already apply the Gaussian filter, create feature maps by neglecting approximation data during the IWT process.

\[
 f_s^C(x, y) = \frac{(IWT_s(H_S^C, V_S^C, D_S^C))^2}{\eta} \tag{4}
\]

Where \( f_s^C(x, y) \) is the feature map generated for the \( s^{th} \) level decomposition for each image sub-band \( C \). \( \Pi \) is the scaling factor (since the range of the Lab input image for each channel is \([0 \ 255]\), there is a large range for feature values in the above equation; therefore, an appropriate value of \( \eta \) is the scaling factor to limit the feature maps, and \( \eta = 10^4 \) in above equation, where this scaling is necessary to avoid the huge variation in the covariance matrix among the feature maps during the computation of global saliency map). \( IWT_s(.) \) is the reconstruction function here neglecting the \( A_N^C \). Thus, for \( s \) and \( C \) above equation creates \( 3 \times N \) feature maps for an input image. In this work, for each channel \( C \), extract 8 feature maps, which used for reconstruction purpose, from wavelet coefficients of the 1st-level decomposition to the 8th-level decomposition for the given input image. It should be noted that 8-level reconstruction consists of details from the coarsest scale (the 8th level) to the finest scale (the 1st level) wavelet coefficients, the 7-level reconstruction consists of details from the 7th level to the finest scale (the 1st level) wavelet coefficients. In this way feature maps are created. Feature maps of Fig.1 shown in Fig.2.

C. Saliency Map Generation

From the feature maps local and global saliency computations are carried out. From \( f_s^C(x, y) \) in (4), a location \((x, y)\) can be represented as a feature vector \( f(x, y) \) with a length of \( 3 \times N \) (3 channels \( L, a \) and \( b \), and \( N \)-level wavelet-based features for each channel) from all feature maps.

Regarding the feature maps, the features at a given location can be defined by the probability density function (PDF) with a normal distribution. So, the Gaussian PDF can be written as

\[
 p(f(x, y)) = \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2} \sum_{i=1}^{N} |f(x, y) - \mu|^2} \tag{5}
\]

\[
 \sum_i E[(f(x, y) - \mu)(f(x, y) - \mu)^T] = \mu \tag{6}
\]
Where \( \mu \) is the mean vector, which consists of the mean of each feature map. \( T \) is the transpose operation and \( \Sigma \) in (6) is the n x n covariance matrix, \( n=3 \times N \), the number of the feature vector including 3 color channels and \( N \) feature maps for each color channel, and \( |T| \) is the determinant of the covariance matrix. Using the PDF in (5), the global saliency map can be computed as (7) below. Obtained result in (7) is filtered with a \( k \times k \) 2-D Gaussian low-pass filter.

\[
S_G(x, y) = \left( \log \left(p \left(f(x, y) \right) \right)^{-1} \right)^{1/2} \ast I_{k \times k} \tag{7}
\]

Where \( S_G \) give global saliency map. The result from (7) may generate a saliency map with small salient regions, and thus causes some loss in local saliency information.

Local saliency is created by fusing the feature maps at each level linearly without any normalization operation. The feature maps obtained in (4) are used for calculating the local saliency map as:

\[
S_L(x, y) = \left( \sum_{s=1}^{N} \arg \max (f_s^L(x, y), f_s^a(x, y), f_s^b(x, y)) \right) \tag{8}
\]

Where \( f_s^L(x, y), f_s^a(x, y) \) and \( f_s^b(x, y) \) are the feature maps at scale \( s \) for \( L, a, b \) channels respectively. \( S_L(x, y) \) is the local saliency map. Based on (7) and (8), create the global final saliency map. The integration of local and global is performed to modulate the local saliency map with its corresponding global saliency map. Final saliency calculated by using,

\[
S'(x, y) = M \left( S'_L(x, y) \times e^{S_G(x, y)} \right) \ast I_{k \times k} \tag{9}
\]

Where \( S'(x, y) \) is the final saliency map. Here local saliency map \( S'_L(x, y) \) and global saliency map \( S'_G(x, y) \) are used to get final saliency map. The modulation is applied by the multiplication of local saliency and the exponential value of the global saliency, \( M(\cdot) \) is used as the non-linear normalization function. From this (9), obtain a saliency map which considers local features at a location with its respective global feature distribution. Therefore, the global relation between local feature maps is established without the need of any complex feature map normalization process for enhancing each feature map. In addition, enhance the final saliency map. The idea is that the locations around the focus of attention (FoA) have to be more attentive than those away from the FoA. So, saliency values around the most salient points are increased to enhance the performance of the saliency map.

\[
S(x, y) = S'(x', y') \left( 1 - d_{\text{FoA}(x, y)} \right) \tag{11}
\]

Where \( S(x, y) \) is the saliency value at point \((x, y)\) and \( S'(x', y') \) is the salient value of the most salient points at the location \((x', y')\) extracted from the saliency map in (9). \( d_{\text{FoA}} \) represented as the distance between the location \((x, y)\) and its closest FoA at the location \((x', y')\). With the help of (11) enhancement of saliency map will be carried out. So, the saliency values around the salient regions will increase in the final saliency map.

Fig.2. Shows the feature maps of each channel such as \( L, a, b \).

Fig.3. Shows the local, global and final saliency maps of Fig.1

**D. Principal Component Analysis**

Principal component analysis (PCA) used for dimensionality reduction, to get a low dimensional feature space for data representation. It will also give shape feature of an image. PCA is a mathematical technique; it transforms the original image data, which are highly correlated, to principal components, which is a new set of uncorrelated variables. These new components are linear combinations of the original image bands and are derived in decreasing order of importance. Each principal component is called an eigenchannel.

For computing the principal components, calculating the eigenvectors and eigenvalues of the data covariance matrix.
Direction of greatest variation will be the eigenvector with the largest eigenvalue, the direction with the next highest variation will be second largest eigenvalue and so on. Let A be an n x n matrix. Eigenvalues of A are defined as the roots of:

\[ \text{determinant}(A - \lambda I) = |(A - \lambda I)| = 0 \quad (12) \]

Where I is the n x n identity matrix. Equation (12) is called the characteristic equation or characteristic polynomial and has n roots. Let \( \lambda \) be an eigenvalue of A. Then there is a vector \( x \) such that,

\[ Ax = \lambda x \quad (13) \]

Where \( x \) is called an eigenvector of A associated with the eigenvalue \( \lambda \).

For dimension reduction, consider N images each with n pixels. It can represent as n x n data matrix. The first step in PCA is to move the origin to mean of the data. Here finding a mean image \( \mu \) by averaging the columns of matrix. Then subtract the mean image from each image of the dataset, i.e., each row of matrix to create the mean centered data vector. From this covariance matrix will be calculated. Then eigenvectors and eigenvalues are calculated. Finally get the data in a compact form.

E. Support Vector Machine

Firstly, the training images are selected by weakly supervised learning. Next, proposed saliency detection method provides the saliency map as the full labeling map. With the saliency map, the auto-context model is learned and can be used to recognize the same type of objects in the challenging images. In this work, use support vector machine (SVM) for learning and predicting the object. A common task in machine learning is classifying data. Suppose some data which belong to one of two classes, the main goal is to decide which class a new data point will be in. Multiclass SVM aims to assign labels to instances by using support vector machines, in which the labels are drawn from a finite set of several elements. Here SVM is used to retrieve more matching images from database.

V. System Evaluation

Evaluated the performance of local saliency, global saliency, and final saliency in our model to demonstrate how they affect the overall performance. The quantitative performance for the different images evaluate based on overall precision P, recall R and F-Measure. Precision is the fraction of the documents retrieved that are relevant to the user's information need. Similarly Recall can also be calculated, that is the fraction of the documents that are relevant to the query that are successfully retrieved. The F-measure is the weighted harmonic mean of precision and recall. For the evaluation (14), (15) and (16) is used.

\[ P = \frac{\sum_x \sum_y (t(x,y) \times s(x,y))}{\sum_x \sum_y s(x,y)} \quad (14) \]

\[ R = \frac{\sum_x \sum_y (t(x,y) \times s(x,y))}{\sum_x \sum_y t(x,y)} \quad (15) \]

\[ F_\beta = \frac{(1 + \beta^2) \times P \times R}{\beta^2 \times P + R} \quad (16) \]

Where \( t(x,y) \) is ground truth map and \( s(x,y) \) is the saliency map.

VI. Experimental Result

Overall performance of the proposed system evaluated for local saliency maps, global saliency maps and final saliency map with the help of precision (P), recall (R) and F-Measure. The experimental results of P, R and F-measure shows the final saliency map have better performance. Here 50 images of MSRC dataset are used for overall performance evaluation. Experimental results shown in Fig.5.
VII. CONCLUSION

Saliency map for images are created based on wavelet coefficients. In images various feature maps are generated by IWT with the band-pass regions of the image in various scales. Using these feature maps, the local and global saliency maps are introduced to form the final saliency map. The final saliency map represents both the local contrast of each location on the scene and the global distribution of the features as an amplifier for local saliency values. The local saliency map is calculated based on the linear combination of each level’s maximum value in the feature maps within L, a, b channels, while the global saliency map is computed based on the normal distribution of the local features. At last based on this saliency map matching images can be retrieving from database with the help of Multi-SVM according to the input images. In Multi-SVM classifiers to differentiate each class from the rest. At the classification stage, the decision function produce the largest value will be the final output.

REFERENCES
