Knowledge-Based Systems 82 (2015) 128–138

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Learning compressive sampling via multiscale and steerable support value transform

Shuyuan Yang *, Min Wang, Li Jin, Shigang Wang, Fang Liu, Licheng Jiao

Key Lab of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, School of Electrical and Electronic Engineering, Xi'an, Shaanxi 710071, China

A R T I C L E I N F O

Article history:
Received 15 May 2014
Received in revised form 6 February 2015
Accepted 26 February 2015
Available online 7 March 2015

Keywords:
Learning compressive sampling
Multiscale and steerable support value transform
Biological visual saliency
Variable density function
Compressive imaging

A B S T R A C T

Considering that saliency areas in images play important roles in human visual perception, in this paper we propose a Learning Compressive Sampling (LCS) scheme and its physical implementation for more efficient image acquisition via a new Multiscale and Steerable Support Value Transform (MS2VT). The key idea is to learn geometric saliency maps of images by MS2VT, which is deduced from a mapped least square-support vector machine. Because MS2VT can produce a multiscale, multidirectional, undecimated, dyadic and aliasing transform with shift-invariant and anisotropy properties, the obtained support values can reveal the geometric and saliency information of images. The learned saliency map is then used to formulate a variable density compressive function, to realize a simple, fast and efficient sampling, which aims to allocate more sensing resources to saliency attention areas but fewer to non-salient regions. Several experiments are taken on some natural images and remote sensing images to compare our proposed LCS scheme with traditional samplings when saliency information is not used. Moreover, the performance of MS2VT based saliency detection scheme is also compared with other related saliency detection approaches. The experimental results indicate that it can obtain high quality images, especially in preserving more detailed edges, contours and complex structures, even at low sampling ratios.

1. Introduction

Compressive Sampling (or Compressed Sensing, CS) provides a new sampling framework to collect signals below the well-known Nyquist rate. The basic principle of CS is that sparse or compressible signals can be accurately recovered from a surprisingly small number of random measurements generated from those signals [1–4]. Although the fundamental theory of CS has been soundly developed, when dealing with high dimensional signals such as images and videos, there are still several challenges including huge memory required to store the random sampling waveforms and the subsequent computationally expensive recovery process. Some researchers have made a lot of outstanding work on compressive sampling of images, and in the available literatures[5–9], a novel idea of block compressive sensing of images has been advanced, where information acquisition is independently conducted in a block-by-block manner with the same sampling ratio. However, because contents of images can vary significantly across different regions or blocks, some blocks are more salient or important than the others in inciting visual responses of observers. Therefore, assigning the same sampling ratio for all the blocks seems not very reasonable in block compressive sampling of images.

It is well known that Human Visual System (HVS) exploits the visual attention mechanism by which only the saliency areas projected onto the retina is thoroughly processed by brain for semantic understanding. Visual attention is an effective perceptual behavior in sensing scenes, which aims to find some saliency objects from their surroundings by calculating a spatial saliency map [10–12]. It helps our brain to filter out excessive visual information and enables our eyes to focus on particular regions of interest, and nowadays visual attention has been a useful clue for image processing. Saliency areas in natural scenes are generally regarded as areas which the human eyes will typically focus on, and the human can effortlessly identify them in a complex scene. Therefore finding these saliency areas will be helpful for more efficient compressive sampling of images. In other words, more sampling resources on the saliency regions will obtain higher quality reconstructed images.

* Corresponding author.
E-mail address: syyang@xidian.edu.cn (S. Yang).

http://dx.doi.org/10.1016/j.knosys.2015.02.028
0950-7051/© 2015 Elsevier B.V. All rights reserved.

Over the past decade modeling visually saliency attention mechanism has attracted increasing interests [14–21,25,35–40]. For example, Itti [25] proposed a saliency model that calculated feature maps for luminance, color and orientation using a center-surround operator across different scales and combined them into a unified saliency map by normalization and summation. Navalpakkam [14] extended the Itti model by integrating the top down attention mechanism with a multiscale object representation. Hoya [15] also presented an attention model using a hierarchically arranged Generalized Regression Neural Network (GRNN). Recently, Hou [16] proposed a Spectral Residual (SR) approach based on Fourier transform, which can rapidly detect saliency objects. In paper [35], the authors analyzed the advantages and disadvantages of the existing saliency methods and summarized five principles for designing saliency detection algorithms. Several works have also explored the saliency map in terms of information maximization principle, for example, in paper [37], a bottom-up overt attention model was proposed based on the principle of maximizing information sampled from a scene. In paper [38], the authors proposed a computational model for visual saliency derived from the information maximization principle. In [39], a dynamic visual attention model was proposed, where an Incremental Coding Length (ICL) is defined to measure the perspective entropy.

Although these models explore the biological saliency attention mechanism from different aspects of biological visual perception, they are limited in capturing the geometric information of images, which is important in modeling images, images analysis and the subsequent processing. In this paper, we integrate the geometric information into the saliency map because saliency is a direct property of the geometry or morphology of images, to advance a Multiscale and Steerable Support Value Transform (MS2VT) based on a Least Square-Support Vector Machine (LS-SVM). The MS2VT can provide a multiscale, multidirectional and undecimated dyadic transform with the anisotropy and shift-invariant properties. Moreover, due to support values of pixels defined in the mapped neighborhood center instead of conventional intensity value, supported values of images can be taken as a measure of the geometric saliency features of the original image. Then we calculate a geometric saliency map of the image to construct a Learning Compressive Sampling (LCS) scheme via MS2VT. The key idea of LCS is to exploit the visual saliency information of images for more efficient image acquisition, by allocating more sensing resources to saliency attention regions but fewer to non-salient regions. Subsequently, both the quality of reconstructed images can be improved and the waste of sampling resources can be reduced, when only a few measurements are available.

As to the physical realization of our method, a Coded aperture snapshot spectral imager (CASSI) in [41] is employed. A coded aperture is added into the traditional imaging devices (between the imaging optics and relay optics), to acquire the block slices. The coded aperture can work on a low-ratio sampling state, to produce a pre-image. Then we perform the MS2VT on the pre-image to generate a geometric saliency map from multiscale support values. Using this process, both the salient and the other non-salient regions can be identified. As soon as this saliency map is learned, it is used to formulate a variable density compressive function, to realize a simple, fast and efficient compressive sampling, to allocate more sensing resources to saliency regions but fewer to non-salient regions. Finally, using those deliberately selected measurements, the imaging quality will be improved compared with uniform sampling. Some experiments are taken to compare the performance of our proposed LCS scheme with traditional samplings when saliency information is not used. Moreover, the performance of MS2VT based saliency detection scheme is also compared with other related saliency detection approaches. The experimental results show that our method can obtain higher quality images, especially in preserving more detailed edges, contours and complex structures at the same sampling ratio as that of uniform sampling.

Compared with other available works, the main contributions of our method are threefold: Firstly, new MS2VT is advanced to derive the geometric saliency features of images. Secondly, a variable density compressive function is designed to assign different number of measurements for salient and non-salient regions. Thirdly, a new LCS scheme and its physical realization is proposed for more efficient image acquisition. Consequently the quality of recovered images are improved and the waste of sampling resources are reduced. The rest of this paper is organized as follows. Section 2 introduces the proposed geometric saliency maps and the learned compressive sampling scheme. Section 3 presents some simulation experiments to illustrate the efficiency and superiority of our method to its counterparts. Finally in Section 4 some conclusions are drawn.

2. MS2VT based saliency maps for Leaning Compressive Sampling (LCS) of Images

In this section, we first discuss the deduction of multiscale and steerable support value filters, and then the multiscale and steerable support value transform. Based on it, the learning compressive sampling scheme via MS2VT is detailed.

2.1. Multiscale and Steerable Support Value Transform (MS2VT)

It is well known that Support Vector Machines (SVMs) can approximate function $f : \Re^d \to \Re$ through a hypothesis,

$$f(\mathbf{x}) = \omega^T \phi(\mathbf{x}) + b \tag{1}$$

where $\mathbf{x} \in \Re^d,\ b \in \Re,\ \omega \in \Re^l$ and $\phi(.) : \Re^d \to \Re^l (l > d)$ is a nonlinear mapping function. According to the represent theorem [28], ω can be linearly represented by a set of samples, i.e., $\boldsymbol{\omega} = \sum_{i=1}^{N} \alpha_i \phi(\mathbf{x}_i)$($N$ is the number of samples). So Eq. (1) can be reformulated as,

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{2}$$

where $K(\mathbf{x}_i, \mathbf{x}) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}), i = 1, \ldots, N$ is the kernel function satisfying the Mercer's condition; and α_i is the Lagrange multiplier of $\mathbf{x}_i$. Given a set of training samples $\{(\mathbf{x}_i, y_i)\}_{i=1,\ldots,N}$, a SVM seeks the solution to the following formula

$$\min_{\omega,b} \frac{\|\omega\|^2}{2} + \frac{\gamma}{2} \sum_{i=1}^{N} [f(\mathbf{x}_i) - y_i]^2 \tag{3}$$

where γ is a positive regularization constant. LS-SVM finds the optimal $\{\boldsymbol{\omega}, b\}$ by solving a set of equations instead of a convex quadratic programming problem for classical SVMs. The conditions for optimality can be written as the solution to the following set of linear equations,

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \boldsymbol{\Omega} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{Y} \end{bmatrix} \tag{4}$$

where $\mathbf{Y} = [y_1, \ldots, y_N]^T,\ \vec{1} = [1, \ldots, 1]^T,\ \boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^T,\ \boldsymbol{\Omega} = \mathbf{K} + \mathbf{I}/\gamma$($I_{N\times N}$ is an identity matrix) and $\mathbf{K}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$. Many types of kernel function K that satisfy Mercer's condition can be used in Eq. (2). It is well known that the geometric regularity of images is important in inciting the human visual attention when sensing scenes, so the geometric regularities along singularity of edges or contours should be emphasized for more accurate representations. We choose an