ARTICLE IN PRESS

Computers and Electrical Engineering xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Multi-scale discriminant saliency with wavelet-based Hidden Markov Tree modelling ☆

Anh Cat Le Ngo [a,b,*], Kenneth Li-Minn Ang [c], Jasmine Kah-Phooi Seng [d], Guoping Qiu [b]

[a] School of Engineering, The University of Nottigham, Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia
[b] School of Computer Science, The University of Nottingham, Jubilee Campus, Nottingham NG8 1BB, United Kingdom
[c] Centre for Communications Engineering Research, Edith Cowan University, 270 Joondalup Dr, Joondalup, WA 6027, Australia
[d] Department of Computer Science & Networked System, Sunway University, Jalan Universiti Bandar Sunway, 46150 Petaling Jaya, Selangor, Malaysia

ARTICLE INFO

Article history:
Available online xxxx

ABSTRACT

Supposed saliency is a binary classification between centre and surround classes, saliency value is measured as their discriminant power. As the features are defined by sizes of chosen windows, a saliency value at each location is varied accordingly. This paper proposes computing saliency as discriminant power in multiple dyadic scales of Wavelet Hidden Markov Tree (HMT), in which two consecutive dyadic scales provide surrounding and central features, organized in a quad-tree structure. Their discriminant power is estimated as maximum a posterior probability (MAP) by Expectation-Maximization (EM) iterations. Then, a final saliency value is the maximum discriminant power generated among these scales. Standard quantitative tools and qualitative assessments are used for evaluating the proposed multi-scale discriminant saliency (MDIS) against the well-know information based approach AIM on its image collection with eye-tracking data. Simulation results are presented and analysed to verify the validity of MDIS as well as point out its limitation for further research direction.

1. Introduction

Visual attention is a psychological phenomenon in which human visual systems are optimized for capturing scenic information. Robustness and efficiency of biological devices, the eyes and their control systems, visual paths in the brain have amazed scientists and engineers for centuries. From Neisser [1] to Marr [2], researchers have put intensive effort in discovering attention principles and engineering artificial systems with equivalent capability. For decades, this research field has been dominated by saliency concept, which focuses on quantitatively measuring differences between attractive and irrelevant features. It also promotes existence of saliency maps, which contain saliency scores at every spatial location on perceptual reigns and theory that human attention is respectively driven to the spatial location with the maximum saliency score. Besides influence of spatial features on saliency values, scale-space features also have decisive roles in visual saliency computation since centre or surround environments are basically processing features with different scales. In signal processing, scale-space and spectral space are two sides of a coin; therefore, there is a strong relation between scale-frequency-saliency in visual attention problem. Several researchers [3–6] have outlined that fixated regions have high spatial contrast

☆ Reviews processed and recommended for publication to Editor-in-Chief by Guest Editor Dr. Jing Tian.

* Corresponding author at: School of Engineering, The University of Nottigham, Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia. Tel.: +60 016 2586964.

E-mail addresses: keyx9lna@nottingham.edu.my, lengoanhcat@gmail.com (A.C. Le Ngo).

http://dx.doi.org/10.1016/j.compeleceng.2014.01.012
0045-7906/© 2014 Elsevier Ltd. All rights reserved.

Please cite this article in press as: Le Ngo AC et al. Multi-scale discriminant saliency with wavelet-based Hidden Markov Tree modelling. Comput Electr Eng (2014), http://dx.doi.org/10.1016/j.compeleceng.2014.01.012

or showed that high-frequency edges allow stronger discrimination fixated over non-fixated points. In brief, they all come up with one conclusion: increment in predictability at high-frequency features. Although these studies emphasizes a greater visual attraction to high frequencies (edges, ridges, other structures of images), there are other works focusing on medium frequency as well. Perhaps, that attention system may involve different range of frequencies for optimal eye-movements. In other words, the diversity of attention in spectral spaces needs necessary utilization of scale-space theory.

Though multi-scale nature has been emphasized as the implicit part of human visual attention, it is often ignored in several saliency algorithms. For example, DIS approach [7] considers only one fixed-size window; hence, it may lead to inconsideration of significant attentive features in a scene. Therefore, DIS approach needs constituting under the multiscale framework to form multi-scale discriminant saliency (MDIS) approach. This is the main motivation as well as contribution of this paper which are organized as follows. Section 2 reviews principles behind DIS [8] and focuses on its important assumption and limitation. After that, MDIS approach is carefully elaborated in Section 3 with several relevant contents such as multiple dyadic windows for binary classification problem in Section 3.1, multi-scale statistical modelling of wavelet coefficients and learning of parameters in Sections 3.2, 3.3, maximum likelihood (MLL) and maximum a posterior probability (MAP) computation of dyadic sub-squares in sub Sections 3.4, 3.5. Then, all MDIS steps are combined for final saliency map generation in sub Section 3.6. Quantitative and qualitative analysis of the proposal with different modes are discussed in Section 4; moreover, comparisons of the proposed MDIS and the well-known information-based saliency method AIM [9] simulation data are presented with several interesting conclusions. Finally, main contributions of this paper as well as further research direction are stated in the conclusion Section 5.

2. Visual attention – discriminant saliency

In decision theory, saliency is regarded as power for distinguishing salient and non-salient classes. Such theory allows integrate the classical centre-surround hypothesis into a statistical architecture. At each spatial location, saliency score is measured as discriminant power of two feature sets with respect to centre and surround classes or two hypotheses in the binary classification problem defined as follows.

- Interest hypothesis: observations within a central neighborhood W_l^1 of visual fields location l.
- Null hypothesis: observations within a surrounding window W_l^0 of the above central region.

Within DIS, feature responses within the windows are randomly drawn from the predefined sets of features X. Since there are many possible combinations and orders of how such responses are assembled, the observations of features can be considered as a random process of dimension d. $X(l) = (X_1(l), \ldots, X_d(l))$. Each observation is drawn conditionally on the states of hidden variable $Y(l)$, which is either centre or surround state. Feature vectors $\mathbf{x}(\mathbf{j})$ such that $j \in W_l^c,\ c \in \{0, 1\}$ are drawn from classes c according to conditional densities $P_{X(l)|_{Y(l)}}(x|c)$ where $Y(l) = 0$ for surround or $Y(l) = 1$ for centre. The saliency S at location l, $S(l)$, is equal to the discriminant power of X for classifying observed feature vectors, which can be quantified by the mutual information between features, X, and class labels, Y.

$$S(l) = \sum_c \int p_{X,Y}(x, c) log \frac{p_{X,Y}(x, c)}{p_X(x) p_y(c)} dx \tag{1}$$

$$= \frac{1}{|W_l|} \sum_{j \in W_l} \left[H(Y) + \sum_{c=0}^{1} P_{Y|X}(c|x_j) log P_{Y|X}(c|x_j) \right] \tag{2}$$

Given a location l, there are corresponding centres W_l^1 and surround W_l^0 windows along with a set of associated feature responses $x(j),\ j \in W_l = W_l^0 \cup W_l^1$.

3. Multiscale discriminant saliency

Expansion from a fixed window-size to multi-scale processing is commonly desired in development of computer vision algorithms [10,11]. In long literature of computer vision research fields, there are many multi-scale processing models [11–14] which can be used as references for developing a so called Multi-scale Discriminant Saliency (MDIS). Any selected model has to adapt binary classification in multi-scale stages. Put differently, it requires efficient classification of imaging data into a class at a particular scale with prior knowledge from other scales. With respect to these requirements, a multi-scale image segmentation framework should be a great starting point for MDIS since DIS can be considered as simplified binary image segmentation with only two classes. However, the binary classification is only an intermediate step to measure discriminant power of centre-surrounding features, and accuracy of segmentation results does not really matter in this case.

Typical algorithms employ a rigid classification window in a vague hope that all pixels in region of interest (ROI) belong to the same class. Obviously, DIS has similar problems of choosing suitable window sizes as well. Clearly, the size is crucial to balance between its reliability and accuracy. A large window usually provides rich statistical information and enhance reliability of the algorithm. However, it also risks including heterogeneous elements in the window and eventually loses segmentation accuracy. Therefore, appropriate window sizes are equivalently vital in avoidance of local maxima in calculation