Pattern Recognition 60 (2016) 106–120

Contents lists available at ScienceDirect

Pattern Recognition

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

Learning feature fusion strategies for various image types to detect salient objects

Muhammad Iqbal*, Syed S. Naqvi, Will N. Browne, Christopher Hollitt, Mengjie Zhang

School of Engineering and Computer Science, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand

A R T I C L E I N F O

Article history:
Received 7 March 2016
Received in revised form
20 April 2016
Accepted 4 May 2016
Available online 21 May 2016

Keywords:
Object Detection
Saliency Map
Learning Classifier Systems
XCS
Pattern Recognition

A B S T R A C T

Salient object detection is the task of automatically localizing objects of interests in a scene by suppressing the background information, which facilitates various machine vision applications such as object segmentation, recognition and tracking. Combining features from different feature-modalities has been demonstrated to enhance the performance of saliency prediction algorithms and different feature combinations are often suited to different types of images. However, existing saliency learning techniques attempt to apply a single feature combination across all image types and thus lose generalization in the test phase when considering unseen images. Learning classifier systems (LCSs) are an evolutionary machine learning technique that evolve a set of rules, based on a niched genetic reproduction, which collectively solve the problem. It is hypothesized that the LCS technique has the ability to autonomously learn different feature combinations for different image types. Hence, this paper further investigates the application of LCS for learning image dependent feature fusion strategies for the task of salient object detection. The obtained results show that the proposed method outperforms, through evolving generalized rules to compute saliency maps, the individual feature based methods and seven combinatorial techniques in detecting salient objects from three well known benchmark datasets of various types and difficulty levels.

1. Introduction

Visual saliency has recently attracted much computer vision research, giving birth to a new sub-domain known as salient object detection [1]. For salient object detection, the task is to detect the salient, attention grabbing object(s) in a scene and subsequently segment it in its entirety [2,3]. It is similar to the problem of figure-ground segmentation [4–6], but differs from the traditional segmentation problem as the task is simply to find the most salient object rather than completely partitioning the image into perceptually homogeneous regions [7]. Salient object detection is actually the task of marking regions of interest in a scene, which facilitates various computer vision applications, e.g. image segmentation [8], image retrieval [9,10], picture collage [11,12], object recognition [13] or image compression [14].

Most methods specialized for the task of salient object detection concentrate on constructing deterministic tailor-made features [15,16] such as color or color gradient and apply heuristics to combine them. A class of models [17–19] use low, mid and high-level features to learn a single set of weighting parameters for combining features, but apply them across multiple types of images, e.g. images with cluttered backgrounds or multiple objects of interest. Therefore, such techniques inherently lose generalization when operated on test sets with different images having various properties and sets of features. An alternative approach is to learn model parameters using an assembly of weak learners, which increase generalization. However the quality of final solution depends upon the performance of individual learners and can be degraded if one of the learners is not optimal [20].

A learning classifier system (LCS) is a rule-based machine learning technique in which each rule relates sections of the feature space with a classification and a measure of accuracy [21,22]. To address the issue of loss in generalization on unseen image types and to make the system general for all image types, previously we utilized the strength of LCS to autonomously divide the feature space into niches and construct rules covering each image type [23]. The aim of this paper is to extend and

* Corresponding author.
E-mail addresses: muhammad.iqbal@ecs.vuw.ac.nz (M. Iqbal), syed.saud.naqvi@ecs.vuw.ac.nz (S.S. Naqvi), will.browne@ecs.vuw.ac.nz (W.N. Browne), christopher.hollitt@ecs.vuw.ac.nz (C. Hollitt), mengjie.zhang@ecs.vuw.ac.nz (M. Zhang).

http://dx.doi.org/10.1016/j.patcog.2016.05.020
0031-3203/

demonstrate the LCS technique proposed in [23] by fully investigating niching of image types and demonstrating performance on a wide range of domains and salient object detection benchmark techniques.

The rest of the paper is organized as follows. Section 2 briefly describes the related work in salient object detection. In Section 3 the proposed LCS technique to detect salient objects in an image is detailed. Section 4 introduces the datasets, parameter settings, and performance measures used in the experimentation. In Section 5 experimental results are presented and compared with existing state-of-the-art systems. Section 6 provides an analysis of the evolved classifier rules obtained using the proposed LCS system. In the ending section this work is concluded and the future work is outlined.

2. Background

This section introduces the necessary background in learning classifier systems, and the related work in salient object detection.

2.1. Learning classifier systems

Traditionally, an LCS represents a rule-based agent that incorporates a genetic algorithm (GA) and machine learning to solve a given task by evolving a population of interpretable classifiers. Each classifier covers a part of the feature space that may be overlapped with other classifiers. The LCS technique has been successfully applied to a wide range of problems including classification, data mining, control, modeling, image processing and optimization problems [24–31].

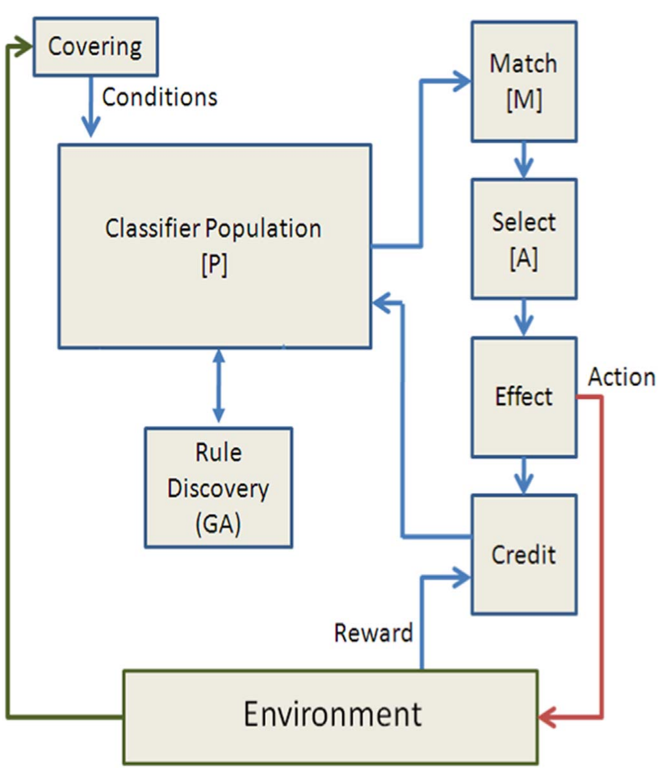

Fig. 1. Overview of a learning classifier system [28].

The proposed LCS method, to be presented in this study, to compute saliency maps is based on XCS [32], which is a well-tested LCS model. In XCS, the learning agent evolves a population [*P*] of classifiers, as depicted in Fig. 1, where each classifier consists of a rule and a set of associated parameters estimating the quality of the rule. Each rule is of the form 'if *condition* then *action*', where condition is used to match input observations, and the corresponding action predict the class label for a given observation. Commonly, the condition in a rule is represented by a conjunction of predicates using one predicate for each corresponding input feature; and the action is represented by a numeric constant.

In XCS, on receiving the environmental input state s, a match set [*M*] is formed consisting of the classifiers from the population [*P*] that have conditions matching the input s. For every action a_i in the set of all possible actions, if a_i is not represented in [*M*] then a covering classifier is randomly generated. After that an action a is selected to be performed on the environment and an action set [*A*] is formed, which consists of the classifiers in [*M*] that advocate a. After receiving an environmental reward, the associated parameters of all classifiers in [*A*] are updated. When appropriate, new classifiers are produced using an evolutionary mechanism, usually a GA. Additionally, in XCS overly specific classifiers may be subsumed by any more general and accurate classifiers in order to reduce the number of classifiers in the final population [33]. For a complete description, the interested reader is referred to the original XCS papers by Wilson [32,34], and to the algorithmic details by Butz and Wilson [35].

2.2. Salient object detection

Visual attention is a fundamental research problem in psychology, neuroscience, and computer vision literature. Researchers have built computational models of visual attention to predict where humans are likely to fixate [36]. Recently, this work has been expanded to identify salient objects in a scene for object detection and localization. Salient object detection is a difficult problem in computer vision as natural scenes can include objects with cluttered backgrounds (making it difficult to distinguish the object from background based on its features) and scenes containing multiple objects.

Deterministic methods to detect salient objects include fine human-constructed features, but they usually combine them linearly, thus neglecting the importance of individual features [16]. Machine learning approaches have the ability to learn feature importance during combination, which enhances their performance in challenging cases such as scenes with cluttered backgrounds and multiple objects [37].

Tong et al. [38] used 73 texture and color features by exploring both global and local cues to compute a saliency map. However, the simplistic nature of feature combination (i.e., the average of the local and global features) compromises the final saliency output on difficult cases of saliency detection. Judd et al. [17] learned a model of saliency from 33 features (including low, mid and high level features) to predict human eye fixations. They used support vector machines (SVMs) with linear kernels to learn feature weightings, while Zhao and Koch [19] used least square regression to learn weights for eye fixation prediction using basic saliency features (i.e., color, intensity and orientation). Both the discriminative approaches lose generalization on a subset of images due to a single weighting scheme being applied to features for all image types. Singh et al. [39] applied a constrained Particle