Computers in Industry 98 (2018) 183–191

Contents lists available at ScienceDirect

Computers in Industry

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

Real-time object detection in agricultural/remote environments using the multiple-expert colour feature extreme learning machine (MEC-ELM)

Edmund J. Sadgrove*, Greg Falzon, David Miron, David W. Lamb

Precision Agriculture Research Group, University of New England, Armidale, NSW 2351, Australia

ARTICLE INFO

Article history:
Received 31 August 2017
Received in revised form 6 March 2018
Accepted 15 March 2018
Available online xxx

Keywords:
Extreme learning machine
Object detection
Machine vision
Unmanned aerial vehicle
Agriculture
Robotics

ABSTRACT

It is necessary for autonomous robotics in agriculture to provide real time feedback, but due to a diverse array of objects and lack of landscape uniformity this objective is inherently complex. The current study presents two implementations of the multiple-expert colour feature extreme learning machine (MEC-ELM). The MEC-ELM is a cascading algorithm that has been implemented along side a summed area table (SAT) for fast feature extraction and object classification, for a fully functioning object detection algorithm. The MEC-ELM is an implementation of the colour feature extreme learning machine (CF-ELM), which is an extreme learning machine (ELM) with a partially connected hidden layer; taking three colour bands as inputs. The colour implementation used with the SAT enable the MEC-ELM to find and classify objects quickly, with 84% precision and 91% recall in weed detection in the Y'UV colour space and in 0.5 s per frame. The colour implementation is however limited to low resolution images and for this reason a colour level co-occurrence matrix (CLCM) variant of the MEC-ELM is proposed. This variant uses the SAT to produce a CLCM and texture analyses, with texture values processed as an input to the MEC-ELM. This enabled the MEC-ELM to achieve 78–85% precision and 81–93% recall in cattle, weed and quad bike detection and in times between 1 and 2 s per frame. Both implementations were benchmarked on a standard i7 mobile processor. Thus the results presented in this paper demonstrated that the MEC-ELM with SAT grid and CLCM makes an ideal candidate for fast object detection in complex and/or agricultural landscapes.

1. Introduction

Agriculture systems require autonomous robotics for weed spraying, livestock detection and vehicle safety. The ability to detect and process objects quickly is a desire of many of these systems. Agricultural scenarios can be exceedingly complex as compared to other industrial based robotics systems. Object detection algorithms in agriculture may compete with any number of structurally similar or diverse objects. This makes for a complex environment with the potential for many false positives and false negatives. A potential solution to this problem is to adopt a cascading or multiple expert approach. These types of solutions have varying levels of success in both flora and fauna detection and there are numerous implementations, ranging from substratal to more complex. These approaches include simple colour and texture detection, as well as broad template matching and have been implemented for weed, horse and wildlife detection [1–3]. The more complex solutions use a wide range of techniques, which can include large data structures such as deep or exemplar neural networks and varying levels of texture and shape based analyses. This includes for the ripeness of bananas and for other generic object detectors [4–7].

The approaches discussed often rely on grey-scale images and in many cases, the detection of prominent features or stand out colour attributes. Processing speed is the key advantage in this case. Deep learning architectures are much slower and more complex and for this reason are often avoided in place of real time solutions. Notably, with a high-end GPU it is possible to process a large number of frames per second [8], but in remote mobile computing GPUs might not be attainable. A disadvantage of many approaches is the reliance on prominent key characteristics in the classification process. This often leads to poor overall classification accuracy, particularly in more complex scenarios, e.g., in weed detection there may not be any prominent features; in the case of cattle detection there may be variations in colour between the

* Corresponding author.

https://doi.org/10.1016/j.compind.2018.03.014

same breed and in other cases an object's features may appear different at different angles or rotations.

The goal of this research is then to explore methods that can be used to deliver both fast and accurate feature extraction and object classification. For this, an implementation of the Multiple-expert colour feature extreme learning machine (MEC-ELM) [9] is proposed. The MEC-ELM is a cascading implementation of the Colour Feature Extreme Learn Machine (CF-ELM) [10], which is itself an implementation of the Extreme Learning Machine (ELM) [11] for colour object detection. The ELM in part due to its efficient implementation and fast processing speeds has demonstrated suitability for computer vision based problems in the agricultural scenarios. Including for soybean classifications [12], unmanned aerial vision for palm tree detection [13] and has been benchmarked using notable feature extraction techniques [7]. The MEC-ELM can be used as both a feature extraction and classification technique. This can be achieved by adopting the summed area table (SAT) [14] (or integral image) and thereby reducing a landscape image (or video frame) to a grid of coloured blocks. The purpose of this is to provide a generic approach to HAAR features [15] and hence take advantage of the SATs fast, multi-scale feature extraction architecture. Fast processing may require low resolution image data and this can result in a potential loss of pixel based information. To meet this challenge, the output of the SAT grid is used to generate three colour level co-occurrence matrices (CLCM) [16], one for each colour band (red, green and blue). The outcome will be a texture based analysis, with the values provided as input of each CF-ELM. The two implementations of the MEC-ELM were implemented and processing speeds and overall accuracy compared. The algorithm is designed with the objective of fast object detection in complex and unpredictable terrain, making it an ideal candidate for use in the agriculture industry. The objective in this case is to implement the MEC-ELM as a weed detector for spraying, unobtrusive cattle tracking and as a vehicle interaction and avoidance tool. This paper will benchmark the MEC-ELM with pre-recorded video data, for eventual use in a remote laptop interface for interaction with an unmanned aerial vehicle (UAV or drone) and stationary surveillance devices.

2. Theory/calculation

2.1. Multiple-expert colour extreme learning machine (MEC-ELM)

The ELM is a single layer, feed forward neural network that is known for its fast and analytical training phase. In this phase the output of the neural networks hidden layer is stored in a matrix designated **H** the output weights are then determined analytically. This can be expressed [17,18]:

$$\mathbf{H}(\mathbf{W}_1\cdots,\mathbf{W}_{\tilde{N}},b_1\cdots,b_{\tilde{N}},\mathbf{x}_1\cdots,\mathbf{x}_N) = \begin{bmatrix} g(\mathbf{W}_1\cdot\mathbf{x}_1+b_1) & \cdots & g(\mathbf{W}_{\tilde{N}}\cdot\mathbf{x}_1+b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(\mathbf{W}_1\cdot\mathbf{x}_N+b_1) & \cdots & g(\mathbf{W}_{\tilde{N}}\cdot\mathbf{x}_N+b_{\tilde{N}}) \end{bmatrix}_{N\cdot\tilde{N}} \tag{1}$$

where **H** is the hidden layer output matrix, N is the number of samples used in the training phase and $\tilde{N}$ is the number of neurons in the hidden layer. In the activation function $g()$ [19], **W** is the input weight, **x** is the input sample pixels and b is the bias.

The Colour Feature Extreme Learning Machine (CF-ELM) is similar in architecture to the ELM, comprising of a single layer, feed forward, neural network with a partially connected hidden layer and a fully connected output layer. The hidden layer is divided into 3 sections and this gives the CF-ELM the ability to be used with different colour models, including red, green, blue (RGB), luminance, chrominance red, chrominance blue (Y'UV) and hue, saturation, value (HSV) [20], Y'UV (also known as YCrCb) is defined by the international telecommunications union as ITU-R B.601 [32]. Equivalent to the standard ELM it uses randomly assigned weights in the hidden layer and by using the pseudo inverse it can analytically determine the output weights from a fully connected output layer. By dividing the hidden layer into 3 sections, each colour attribute can be processed in a separate section of the hidden layer and it is for this reason that the number of neurons in the hidden layer must be a multiple of 3. By storing the output of these 3 sections into 3 sections of the **H** matrix it allows the matrix to be used to determined the output weights in the same way as the standard ELM. For Y'UV the CF-ELM hidden layer can be expressed [10].

$$\mathbf{H}(\mathbf{W}_1\cdots,\mathbf{W}_{\tilde{N}},b_1\cdots,b_{\tilde{N}},\mathbf{Y'}_1\cdots,\mathbf{Y'}_N,\mathbf{U}_1\cdots,\mathbf{U}_N,\mathbf{V}_1\cdots,\mathbf{V}_N) = \begin{bmatrix} g(\mathbf{W}_1\cdot\mathbf{Y'}_1+b_1) & \cdots & g(\mathbf{W}_{\tilde{N}}\cdot\mathbf{Y'}_1+b_{\tilde{n}}) \\ \vdots & \cdots & \vdots \\ g(\mathbf{W}_1\cdot\mathbf{Y'}_N+b_1) & \cdots & g(\mathbf{W}_{\tilde{N}}\cdot\mathbf{Y'}_N+b_{\tilde{n}}) \\ g(\mathbf{W}_1\cdot\mathbf{U}_1+b_1) & \cdots & g(\mathbf{W}_{\tilde{N}}\cdot\mathbf{U}_1+b_{\tilde{n}}) \\ \vdots & \cdots & \vdots \\ g(\mathbf{W}_1\cdot\mathbf{U}_N+b_1) & \cdots & g(\mathbf{W}_{\tilde{N}}\cdot\mathbf{U}_N+b_{\tilde{n}}) \\ g(\mathbf{W}_1\cdot\mathbf{V}_1+b_1) & \cdots & g(\mathbf{W}_{\tilde{N}}\cdot\mathbf{V}_1+b_{\tilde{n}}) \\ \vdots & \cdots & \vdots \\ g(\mathbf{W}_1\cdot\mathbf{V}_N+b_1) & \cdots & g(\mathbf{W}_{\tilde{N}}\cdot\mathbf{V}_N+b_{\tilde{n}}) \end{bmatrix}_{3N\cdot\tilde{N}} \tag{2}$$

where $\mathbf{Y}'$, **U** and **V** are equal to the individual colour pixel matrices for each image and are stored in the **H** matrix at the output of the hidden layer. The hidden layer process is repeated in the output layer, with the output β becoming the input multiplier for the output weights β. The output **T** of the CF-ELM is then the result of $\beta\cdot\mathbf{H}$.

$$\mathbf{T}=\beta\cdot\mathbf{H} \tag{3}$$

Here β can be expressed:

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N}\cdot m} \tag{4}$$

where m is the number of neurons in the output layer, which is equivalent to the number of outputs of the ANN. The matrix of target outputs **T** can be expressed as:

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N\cdot m} \tag{5}$$

where for each $\mathbf{T}_N$ the value is stored based on the input training sample and its desired output. This leaves β as the one unknown, by making β the subject we get:

$$\beta=\mathbf{H}^{-1}\cdot\mathbf{T} \tag{6}$$

where $\mathbf{H}^{-1}$ is the Moore-Penrose pseudo inverse of matrix **H**. The output values of this process are then stored in β and used as the weights in the output layer removing the need for a long gradient descent based training process.

The MEC-ELM is then a set of CF-ELMs, where each CF-ELM can be trained on a different set of sample images, different colour system or image analysis techniques. The MEC-ELM becomes a global consensus of all CF-ELMs or experts, the goal is then to find individual CF-ELMS of high classification accuracy but with varying consensus [21]. The method used in this paper was inspired by the Examplar SVM [22] and the Ensemble ELM (EN-ELM) [23], where training samples are divided among different instances of CF-ELMs. The training phase of the MEC-ELM is depicted in Fig. 1,