

A Cognitive Framework for Object Recognition with Application to Autonomous Vehicles

Jamie Roche, Varuna De Silva, Ahmet Konoz

Institute for Digital Technologies
Loughborough University London
London

A.J.Roche@lboro.ac.uk, V.D.De-Silva@lboro.ac.uk

Abstract—Autonomous vehicles or self-driving cars are capable of sensing the surrounding environment so they can navigate roads without human input. Decisions are constantly made on sensing, mapping and driving policy using machine learning techniques. Deep Learning – massive neural networks that utilize the power of parallel processing – has become a popular choice for addressing the complexities of real time decision making. This method of machine learning has been shown to outperform alternative solutions in multiple domains, and has an architecture that can be adapted to new problems with relative ease. To harness the power of Deep Learning, it is necessary to have large amounts of training data that are representative of all possible situations the system will face. To successfully implement situational awareness in driverless vehicles, it is not possible to exhaust all possible training examples. An alternative method is to apply cognitive approaches to perception, for situations the autonomous vehicles will face. Cognitive approaches to perception work by mimicking the process of human intelligence – thereby permitting a machine to react to situations it has not previously experienced. This paper proposes a novel cognitive approach for object recognition. The proposed cognitive object recognition algorithm, referred to as Recognition by Components, is inspired by the psychological studies pertaining to early childhood development. The algorithm works by breaking down images into a series of primitive forms such as square, triangle, circle or rectangle and memory based aggregation to identify objects. Experimental results suggest that Recognition by Component algorithm performs significantly better than algorithms that require large amounts of training data.

Keywords—Object recognition; recognition by component; deep learning; one short classification; intelligent mobility; autonomous vehicles

I. INTRODUCTION

Worldwide there are an average 3,287 road deaths a day. In the UK alone, from 1999 to 2010, there were more than 3 million road casualties [1]. Most recently Transport for London (TfL) data shows that in 2015 25,193 casualties took place at signal-controlled urban junctions [2], [3]. Increasing degrees of automation - from semi-manual to fully computer-controlled vehicles - are already being added to vehicles to improve safety, reduce the number of driver tasks and improve fuel efficiency. It is expected that as autonomous vehicles become more ubiquitous traffic incidents and fatalities will reduce [4], [5].

This research aims to reduce incident and causality numbers on roads using smart autonomous systems. By applying an alternative classification method, this research will develop an improved method of computer perception and spatial cognition. The approach takes inspiration from mammal vision, early childhood development, multimodal sensor data, spatial and social cognition. Actions such as vision, perception, and motivation assist in human development. They are derived from the internal mapping of external stimuli and the internal mapping of internally perceived stimuli [6]. Commonly referred to as Spatial Cognition and Perception, they are tools humans use to understand and navigate the world [7].

Currently, Deep Nets are a popular method of classification [8]. Their main shortcoming is the limited knowledge about the internal workings of artificial networks and the large quantities of training data required to classify objects accurately (i.e. supervised learning) [9]. As objects become more complex the quantity of training data needed increases. Only after supervised learning has taken place can unsupervised learning begin. Unsupervised learning occurs when a machine relates an object to an event it has previously encountered before learning the new image [10]. Although improved learning methods and modern Graphic Processing Units (GPU) can bring forward the point when unsupervised learning begins, the process is slow and relies heavily on resources.

To address the above issues this paper proposes a novel method for object classification utilizing Recognition by Component (RBC). RBC explains the cognitive approaches to perception that humans rely on to understand the surrounding world. RBC denotes an ability to break complex images into a series of primitive forms such as square, triangle, circle or rectangle. Cross-correlating this information with the arrangement of the geometric primitive improves the accuracy [11], [12]. This classification method - best viewed as reducing equivocation - relies on increasing the verity of data the machine is exposed to.

The rest of this paper is organized as follows: Section II describes relevant literature and related work, Section III presents the proposed object recognition framework, Section IV discusses the results, and Section V concludes the paper, with some references to future work.

II. LITERATURE REVIEW AND RELATED WORK

A review of the relevant literature and related work is presented in this section. It is organized into four sections: Perception and Cognition, childhood development, Deep Learning, One Shot Classification and Geometric Based Recognition.

A. Perception and Cognition

Cognition, or cognitive development, is how humans develop and understand their environment and the things they encounter. Researchers from disciplines in neuroscience, cognition, and sociology, have learned a great deal about how humans sense, interpret and communicate information about the things they see [13], [14]. Revision of cognitive information is an on-going process and adapts as humans age [15]. This infers that memories can be changed, some forgotten, with others out of reach at a certain time. How well humans utilize memories affects how they think and understand their surroundings.

Many cognitive pathways are employed to survey a visual scene before a judgment and associated action are made. During this period, objects in sensory range are classified and identified. Posterior and occipital lobes are critical in linking the visual map with reality, and therefore to determining the location of objects [16]. The sound, neck, and extra-ocular muscle contribute to this ability to geo-locate [17]. These muscles and auditory ability are responsible for maintaining the link between reality and visual perception [18]. For example, when the ears hear a sound, the head and eyes move with respect to the body, synchronous input from both eyes is required to locate the object of interest. Once an object has been located, the visual input is compared to memories stored in the temporal lobes - bringing about recognition of the objects humans see.

Clearly perception is a result of many senses working in parallel. For example, humans do not solely rely on vision to recognize people. Shape, movement and other characteristics, that are equally human, contribute to classification [19]. Obstacles can be easily perceived and quickly acted on because of the importance humans assign to a response [20]. For example, when scared, humans experience an increase of adrenaline and nano-adrenaline into their body and brain. Brain activity increases, and oxygen flow allowing muscles and organs to function at a faster rate so that people can move away from the feared object [21].

Not all mammals localize and identify objects in the same manner. Dolphins, Whales and Bats, for example, use echolocation in conjunction with their extra-ocular and ocular muscles to locate and identify objects they encounter [22]. Echolocation is listening for reflected sound waves from objects. The sounds generated are used to determine the position, size, structure and texture of the object [23]. While the process is similar to the echo humans hear, the term echolocation is mostly used for a select group of mammals that employ it on a regular basis [24] for environmental perception, spatial orientation and hunting [22], [25].

B. Childhood Development

Research shows that approximately 50% of a person's intellectual potential is reached by age four and that early life events have an extended effect on intellectual capacity, personality, and social behavior [26]-[29]. How children perceive and make sense of the world can be explored through the interpretation of objects they encounter. Although conditioning plays a crucial role in development, motivations such as 'hunger', a 'desire to understand' and other 'basic instincts' are equally important [30], [31].

Lowenfield's and Edwards describe the stages of development in their research - as per Fig. 1 [32]. Lowenfield's and Edwards hypothesized that children initially portray the world in a series of scribbles, enjoying kinesthetic activities, that are merely manipulation of the environment [32], [33]. After passing through different iterations of the scribble stage a child enters the schematic or landscape stage before eventually progressing onto the realism stage [32], [34].





			
2 years	3 years	4 years	6 years
Scribbling stage	The preschematic stage		The schematic stage
The scribbling stage	The stage of symbols	Pictures that tell stories	The Landscape

Fig. 1. Lowenfield's and Edwards stages of early childhood development [33].

The schematic or landscape stage of development is particularly useful for this research. During this time, a child uses shapes to describe complex images while only starting to discover perspective [35]. This is of vital importance and marks the point where a child arrives at a definitive way of portraying an object. Although the object will be modified with the addition of features when the child is trying to describe something familiar, the structure of the object will largely remain the same. This stage represents active knowledge of the subject and will contain order along a single line upon which all images sit [32].

C. Deep Learning

The structure of a Deep Net is largely the same as a Neural Net; there is an input layer, an output layer and connected hidden layers - as per Fig. 2 [36], [37]. The main function of such a Deep Network is to perceive an input before performing more complex calculations, resulting in an output that can be used to classify and solve a problem. Image based

classification is predominately used to categorize groups of objects using features that describe them [38]. There are many types; Logistical Regression, Support Vector Machines, Naive Bayes and Convoluted Neural Nets. When a classifier is activated it produces a score that is dependent on the weight and bias [39].

When a string of classifiers are placed in a layered web they can be viewed as a Neural Net [40]. Each layer can be broken down into nodes that produce a score, which is passed onto the next layer, diffusing through the network before reaching the output layer. At this point the score generated by the nodes of each layer dictates the result of the classification. This process is repeated for each input into the net and is commonly referred to as feed forward propagation [40].

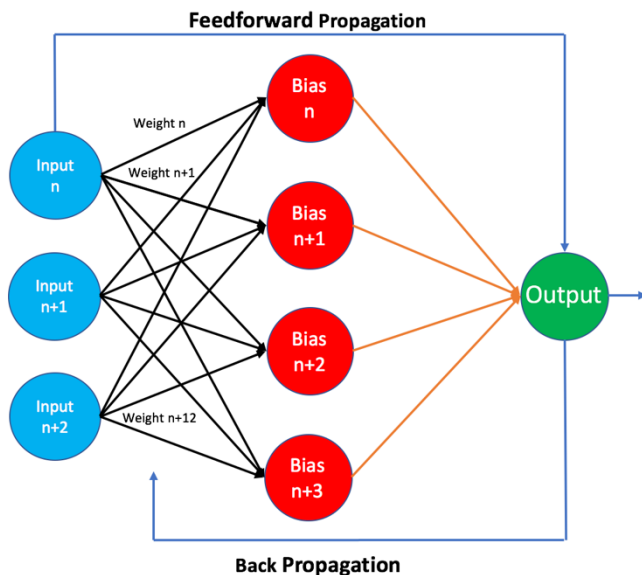


Fig. 2. A typical deep network layout showing, weights, bias, feedforward and back propagation.

When a neural network is faced with a problem the weights and bias that affect the output enable a prediction. Weights and bias are generated and influenced during training. When the output generated during feed forward propagation does not match the output that is known to be correct, the weights and bias change. As the net trains, the difference – often referred to as cost – is constantly reducing [41].

This is the whole point of training. The net gets familiar with the features of the training data before adjusting the weights and bias until the predictions closely match the inputs that are known to be correct. As the problem to be solved becomes progressively complex, Deep Neural Nets start to outperform standard classification engines. Inauspiciously, as the problem become increasingly complex, the number of nodes within the layers grows, and the training becomes more expensive [42]. Deep Nets work around this issue by breaking objects down into a series of simpler patterns [36], [37], [41].

This important aspect of using features, edges and pixels to identify more complex patterns is what gives Deep Learning its strength and its vulnerability. For example, when learning

human faces the Deep Net passes a large region of an image - onto a smaller output region - until it reaches the end. The net result is a small change in output even though a large change in input has occurred. Networks that only ever make small changes do not have the opportunity to learn and never make that giant change to the network that is required for autonomous decisions [43]. Consequently “the gradients of the network’s output with respect to the parameters in the early layers become extremely small” [44].

Commonly referred to as the vanishing gradient problem, it is largely dependent on how the activation function passes inputs into a small output range in a non-linear manner [45]. In 2006 Hinton, Osindero, and Yee-Whye Teh published breakthrough work on the vanishing gradient point problem [37], [46]. Thinking of the gradient as a hill and the training process as a wheel rolling down the hill, the wheel rolls fast along a surface with a large gradient and slow along the low gradient. The same is true of a Deep Net; at the early stages of the net when there is a small learning curve the progress of the net is quite slow. Towards the end where there is a much larger learning curve the net learns at a much quicker rate [46].

This gives way to a singularity since the layers at the start of the net are responsible for identifying the simpler patterns and laying the building blocks of the image. If the layers at the start of the net perceive things incorrectly then the later layers will also get things wrong. To overcome this problem when a Deep Net wants to learn it starts looking at the error to identify the weights that are affecting the output. After this the Net attempts to reduce the error by changing the specific weights [47]. This process, known as back propagation, is used for training Deep Nets and removes the issues created by the vanishing gradient problem [37], [41], [46].

D. One Shot Classification

Humans demonstrate a strong ability to recognize many different types of patterns. Humans, in particular, have an innate ability to comprehend foreign concepts and many different variations on these concepts in future perception [48]. Unlike humans, machine learning is computationally expensive, and although it has proven to be successful in a variety of applications – spam detection, speech and image recognition – the algorithms often falter when forced to make decisions with little supervised information.

One particularly difficult task is classification under restriction – where predictions are made having only observed a single example [48]-[51]. Commonly referred to as One Shot Classification, this form of machine learning identifies “domain specific features” or “inference procedures” that have extremely discriminative properties for the classification task [52]. Subsequently, machines that feature One Shot Classification excel at similar tasks, but fall short at providing reliable results to unfamiliar types of data.

E. Geometric Based Recognition

Falling somewhere between One Shot Learning and Deep Learning, Geometric based Recognition uses pre-defined metrics and some knowledge about the subject before making a decision about the objects perceived. To function, effectively Recognition by Component (RBC) requires an image to be

segmented at regions of deep concavity - as per Fig. 3. This allows an image to be broken into an arrangement of simple geometric components - cubes, cylinders, prisms, etc. The theory, first proposed in 1987 by Irving Biederman, makes the fundamental assumption that humans segment objects of any form into 36 generalized components, called primitives [53].

For true identification, the position of the primitive is the key relationship between perceptual order and object recognition. This enables humans to reliably perceive an image at an obscure angle and still understand what is being observed [53]. If the image can be viewed from any orientation, the projection at that time can be regarded as two-dimensional. Objects; therefore, do not need to be presented as a whole, but can be represented as a series of simplified shapes, even if some parts are occluded [54], [55].

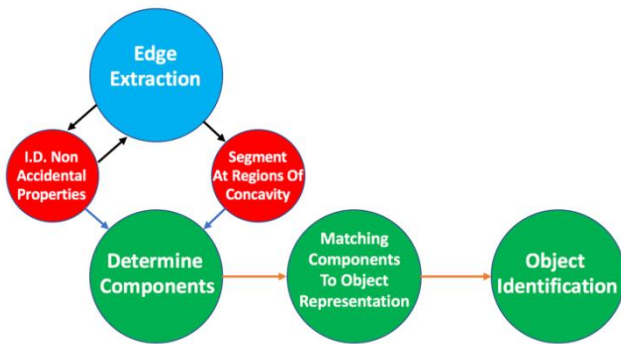


Fig. 3. A schematic of the processes used to recognize an object, as proposed by Biederman [53].

In addition to filling in the blanks for occluded sections of an object, humans are excellent at trying to make sense of the unknown. For example, when presented with unfamiliar objects humans easily recognize the primitives of which the image is composed, even if the overall image is not recognized [54], [55]. Biederman and others believed that humans perform this process on a regular basis [53], [55]-[58]. Therefore, humans rely on what the image is composed of rather than the familiarity of the image as a whole. This is a representational system that identifies elements of complex images to assist in human understanding and development [53].

This phenomenon of RBC allows humans to rapidly identify objects from obscure scenes, at peculiar angles and under noisy conditions [53], [55]-[58]. Deep concavities between primitives are identified using the surface characteristics of the overlapping parts. Non-accidental properties - shapes that look alike from certain angles - are distinguished by co-linearity and symmetry of the primitive being observed [59]. Co-linearity and symmetry play a vital role in identifying components, as does the orientation of the components. For example, a triangle on top of a square bears a striking resemblance to a house, whereas a square on a triangle makes little sense. Just like Lowenfield and Edwards schematic stage of development, the components need to match the representation of the memory both in shape and orientation [33].

III. PROPOSED OBJECT RECOGNITION FRAMEWORK

The previous sections discussed a commonly used method for object detection - Deep Learning. To address the identified problems, an alternative and novel approach of object recognition - RBC - as depicted in Fig. 4 is proposed.

The process of RBC requires the decimation of complex patterns into basic geometric shapes or primitive forms. Difficulties arise when shapes are occluded or overlapping as in Fig. 5. These issues can be resolved by identifying the watershed ridge line at areas of deep concavity between the individual components, as in Fig. 5. Once identified the Euclidean distance between the geometric shapes can be computed before applying the individual component metrics.

Fig. 6 shows the corresponding watershed ridge lines for the same pixels displayed in Fig. 5. From Fig. 6, it is possible to determine the areas of concavities and the geometric shape catchment basin. Images must be of binary form to prepare for boundary tracing and prevent inner contours from being identified. Boundary tracing of eroded images facilitated the identification of object properties - ratio of dimensions, roundness, area, etc. - before classification can occur.

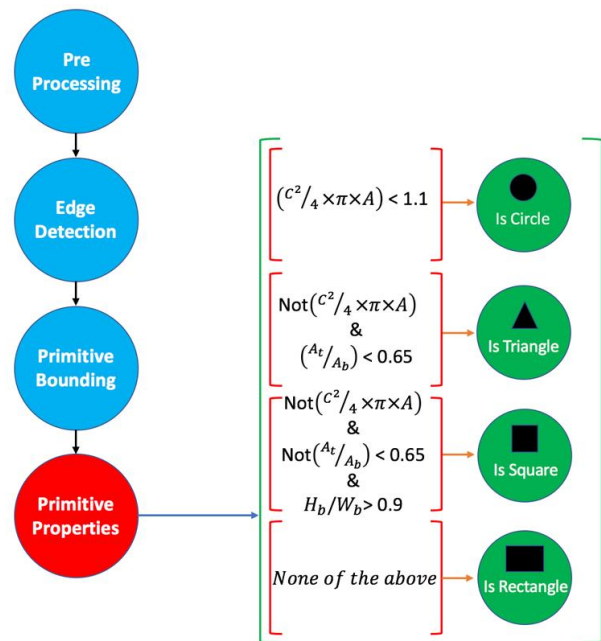


Fig. 4. Proposed framework for RBC algorithm to classify circles, triangles, squares and rectangles.

The segmented image output is equivalent to a book where each page has a geometric shape in the corresponding location to Fig. 5. It should be noted that the original image shown in Fig. 5 contained only four geometric shapes, yet there are 13 layers. These additional layers arise from irregularities within the original image. The additional layers can be viewed as valuable information if looking to identify the background or noise to be filtered at a later point in the process.

The function “minboundrect” - developed by John D'Errico - generates the smallest rectangular bounded box that contains the primitive perimeter [60]. Orientation of the primitives influences the size of the bounding box and affects the accuracy in identifying the primitive in question. Consequently the bounded box surrounding the primitive needs to be rotated by some angle to make the object axis parallel to the horizontal axis of the image [61]. Only then is it possible to calculate the metrics of the shape bounded by the box.

There are a variety of image quantities and features that facilitate the identification of the geometric shape. One of the returning properties - centroid - generates a 1-by-Q vector that specifies the center of mass for each primitive. Additional useful properties for identifying shapes are area and perimeter. When used in conjunction with the function regionprops, area returns a scalar that specifies the number of pixels inside the region of interest. Perimeter is determined in a similar manner to the area, but focuses on the individual pixels around the shape rather than what's inside the boundary. The distance between each adjoining pixel of the primitive boundary is calculated and returns a single value similar to the area scalar returned previously. For two-dimensional shapes area and perimeter provide vital information. These functions allow for the calculation of certain metrics that distinguish the different geometric shapes and facilitate recognition.

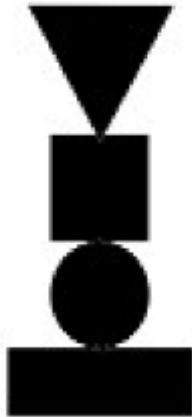


Fig. 5. Image composed of multiple primitives overlapping and touching.

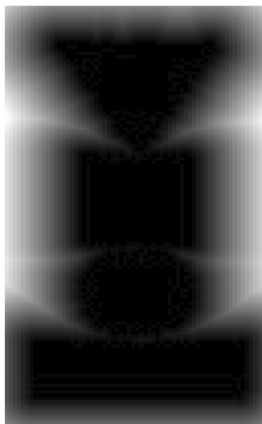


Fig. 6. Image composed primitives showing the watershed ridge lines.

A. Component Metrics

Humans recognize that simple geometric shapes are often categorized into basic classes such as square, rectangle, circle or triangle. However, most shapes frequently encountered are more complex, and typically composed of a combination of the 36 primitives forms [53], [55]-[58]. If machines are to develop cognition to perceive the world as a whole, they first need to be able to understand the simple primitives.

To date our research has focused on the process of classification of four different primitives - circle, triangle, square and rectangle. To classify a circle there are many methods, of which the majority rely on the radius being calculated. Since the method stated above does not produce a radius for any of the primitives, we need to find an alternative method of describing a circle based solely on area and perimeter:

$$A = \pi \times r \tag{1}$$

$$\tag{2}$$

Looking for a solution that utilizes area and perimeter independently of radius, equation 1 and 2 must be rewritten to isolate radius, as follows:

$$\sqrt{A/\pi} \tag{3}$$

$$C/(2\pi) \tag{4}$$

The resulting is two expressions equal to r, but neither containing r in the body of the equation. Since both equations equal to r, (3) can be substituted for r, and rewritten as:

$$C/(2\pi) = \sqrt{A/\pi} \tag{5}$$

It is possible to further simplify by squaring both sides to get:

$$C^2/4\pi^2 = A/\pi \tag{6}$$

Cross-multiplying and divide by π (6) returns:

$$\tag{7}$$

Since C will have units in length and A will have units of area, C will need to be squared. To confirm this, we can set the radius in (1) and (2) to (1) before substituted into (7) to find:

$$(2\pi)^2 \tag{8}$$

Which can be simplified as:

$$— \tag{9}$$

For a triangle, the process relies on comparing the box bounding area and the primitive area. If the ratio between the two is approximately half, the identified shape is as a triangle and not a circle:

$$A \tag{10}$$

Focusing solely on the aspect ratio (width to height) a square is distinguished from all other primitive forms when the ratio is equal to one (= 1.0):

$$(11)$$

Although relatively simple, identifying shapes in this manner raises a complex problem. If the square metrics and the circle metrics return a value close to 1 the shapes could be classified incorrectly. This problem can be addressed using hierarchal conditions with the added benefit of classifying a rectangle at the same time.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we evaluate the performance of the proposed RBC methodology against the more common method of recognition - Deep Learning. It should be noted that data used to test the different recognition methods was of different types.

A. Recognition by Deep Networks

The Rasmus Berg Palm Deep Learning Toolbox in MatLab was used to train a Deep Net with the MINST dataset. The Modified National Institute of Standards and Technology (MNIST) database consists of centered and normalized handwritten digit images - 60,000 examples for the training and 10,000 examples for testing - measuring 28 pixels wide by 28 pixels high. Each pixel of each image is represented by a value between 0 and 255, where 0 is black, 255 is white and anything in between is a different shade of grey.

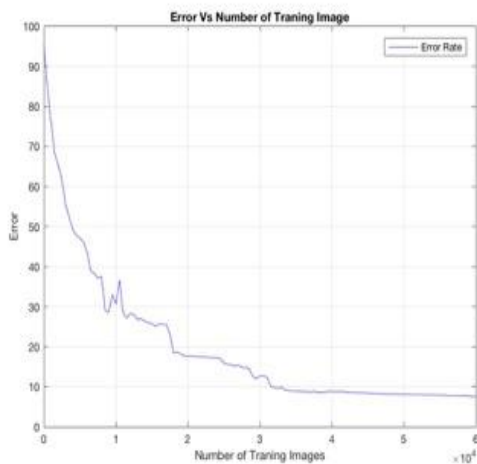


Fig. 7. The rate plotted against the number of images used to train the deep network.

When a machine must decide whether an image contains a digit number of interest - a Deep Net uses features and edges to detect different parts of the number - the whips, curve, length, crown. The accuracy of the Deep Net was proven to be dependent on the total number of images presented to the Deep Net during training - as per Fig. 7.

During training, 60,000 images are broken down into a feature, edge and pixel layer. To test the Deep Net accuracy the above process is repeated with a further 10,000 images from the same dataset. An arduous process that only returns positive

results when the training set is of certain quantity. It was found that the error fell below 10% when the Deep Net was presented with 32,000 or more training images. Below this the error increased and the accuracy of the Net dramatically reduced. To achieve accuracy in the 99th percentile a total of 60,000-digit images were required to train the Deep Net (see Table I).

Observations of the Deep Nets response to hand drawn digit images that were not part of the MNIST dataset were made (see Table I). Typically, the expected output matched the users input with accuracy between of 70% and 80%. For example, a predicted output of 5 was generated with 90% confidence, when the Deep Net was presented with a hand drawn digit image of the number 8 - as per Fig. 8. In another example, a predicted output of 1 was generated with 50% confidence, when the Deep Net was presented with a hand drawn digit image of the number 2 - as per Fig. 8. The results vary depending on the person drawing the digit image, the image type and how closely they match the training data. Table II shows the observations for the predicted value and the corresponding accuracy.

TABLE I. A CONFUSION MATRIX SHOWING THE ACCURACY OF THE TRAINED DEEP NET

1	511 10.2 %	0 0.0 %	1 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	1 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	99.6 %
2	0 0.0 %	501 10.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	1 0.0 %	0 0.0 %	1 0.0 %	99.6 %
3	0 0.0 %	0 0.0 %	495 9.9 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	100 %
4	0 0.0 %	0 0.0 %	0 0.0 %	493 9.9 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	100 %
5	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	508 10.2 %	1 0.0 %	0 0.0 %	3 0.0 %	0 0.0 %	0 0.0 %	99.2 %
6	0 0.0 %	0 0.0 %	0 0.0 %	1 0.0 %	0 0.0 %	492 9.8 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	99.8 %
7	2 0.0 %	0 0.0 %	0 0.0 %	1 0.0 %	0 0.0 %	0 0.0 %	496 10.0 %	0 0.0 %	0 0.0 %	0 0.0 %	99.4 %
8	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	2 0.0 %	0 0.0 %	497 9.9 %	0 0.0 %	0 0.0 %	99.6 %
9	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	494 9.9 %	0 0.0 %	100 %
10	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	1 0.0 %	0 0.0 %	0 0.0 %	0 0.0 %	493 9.9 %	99.8 %
	99.6 %	100 %	99.8 %	99.6 %	100 %	99.2 %	99.8 %	99.2 %	100 %	99.8 %	99.7 %
	0.4 %	0.0 %	0.2 %	0.4 %	0.0 %	0.8 %	0.2 %	0.2 %	0.0 %	0.2 %	0.3 %
	1	2	3	4	5	6	7	8	9	10	

Target Class

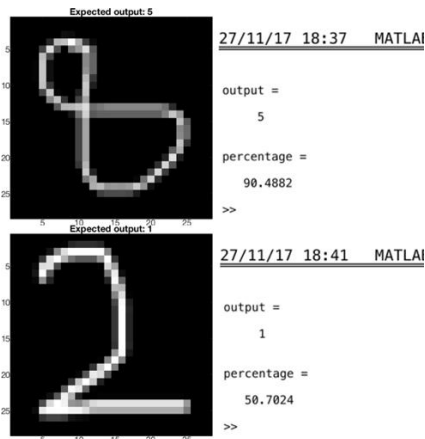


Fig. 8. Hand drawn digit images not contained in the MNIST dataset that were presented to the trained Deep Net. Note the expected output did not always match the hand drawn number correctly.

TABLE II. OBSERVATIONS OF THE DEEP NETS RESPONSE TO HAND DRAWN DIGIT IMAGES THAT WERE NOT PART OF THE MNIST DATASET

Dataset	Digit Numbers not part of the MNIST dataset	
	Predicted Output	Accuracy
Digit 1	1	94.64%
Digit 2	1	50.70%
Digit 3	7	62.36%
Digit 4	4	96.60%
Digit 5	5	98.58%
Digit 6	7	29.56%
Digit 7	7	89.77%
Digit 8	5	90.48%
Digit 9	7	52.18%
Digit 0	0	97.87%

B. Recognition by Component

To test the accuracy of the RBC algorithm, a dataset of 8 synthetic primitives was generated. The dataset consisted of five triangles, one square and one rectangle – as per Fig. 9. In all cases the RBC algorithm accurately identified all synthetic shapes.



Fig. 9. A synthetic dataset developed for testing the RBC algorithm.

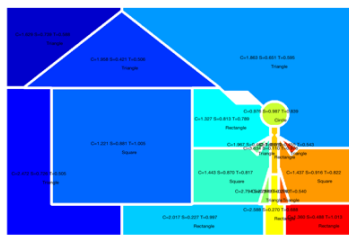


Fig. 10. Primitives recognized by the RBC algorithm using a synthetic dataset.

Fig. 10 shows the results of the combined watershed and shape recognition RBC algorithm. The resulting matrix contains integers of different values, displayed as different colors. The accurately identified shapes were labelled with the class and metrics that define the primitive – C = circle metrics, S = square metrics and T = triangle metrics.

Note the irregularities within the original image generate additional artefacts that the RBC algorithm attempts to classify as shapes. When faced with Fig. 9 the RBC algorithm returned 8 positive results and 9 false positives. As noted, the false positives identified are an unwanted quantity that can be used to identify the background or filtered out depending on requirements. Table III shows the confusion matrix for the RBC algorithm when tested with 22 different primitives of four different classes. In all cases the geometric components were accurately identified with high levels of accuracy.

TABLE III. CONFUSION MATRIX FOR PROPOSED RBC ALGORITHM. A TOTAL OF 22 PRIMITIVES WERE USED TO TEST THE ALGORITHM

Class	Square	Circle	Triangle	Rectangle
Square	4 94.30%	4.60%	1.10%	0.00%
Circle	4.02%	9 88.20%	7.78%	0.00%
Triangle	6.90%	4.10%	4 88.90%	0.00%
Rectangle	0.00%	0.00%	0.00%	5 100%

Observations of the RBC algorithm’s response to data captured by monocular imaging sensor were made. The dataset consisted of one triangle, one square and one rectangle – as per Fig. 11. The triangle, square and triangle were identified with 98%, 94% and 100% accuracy. Irregularities within the original image generate additional artefacts that the RBC algorithm classifies as shapes. The additional artefacts and the components of interest can be seen in Fig. 12.



Fig. 11. A triangle, square & rectangle captured by monocular imaging sensor used to test the RBC algorithm.



Fig. 12. The additional artefacts and the components of interest identified by the RBC algorithm.

C. Limitations of the Current Method

During the course of this research a number of limitations of the proposed framework were identified. Firstly, the proposed method works only on a single frame, and does not utilize the diversity offered by temporal redundancies. Secondly, RBC is best applied to fused perception data captured using Light Detection and Ranging (LiDAR) and a monocular image sensor. Although it is possible to apply RBC to monocular sensor data alone, information redundancy removes the possibility of irregularities being falsely identified as a component of interest. Finally, the proposed framework is the first step towards object recognition using RBC. General density estimation models for a fixed set of fundamental objects will need to be developed before objects other than basic primitives can be recognized. Once fundamental models have been identified it is envisaged that the RBC algorithm will adapt original density estimation models to form new classes. We will consider the above problems and address the limitations in our future work.

V. CONCLUSION AND FURTHER WORK

This paper illustrates some of the shortcomings of Deep Learning methods when applied on systems that do not have the luxury of massive amounts of training data. To address the situational awareness of autonomous vehicles (or similar systems) - which require algorithms to react to new situations to which they were not trained - we propose a novel cognitive approach for object recognition. The proposed method, named Recognition by Components (RBC) - is inspired by early childhood psychology - is shown to be more practical to use, without the need for large amounts of training data.

To facilitate RBC a method identifying the watershed ridge line between adjoining primitive forms was explored. Unlike traditional methods of machine learning, this approach mimics early childhood development and the multi sensing methods mammals frequently use to recognize objects. The preliminary results presented in this paper indicate that the proposed method is capable of learning with small amounts of data. The future work planned, includes the development of a sensor fusion framework to include multiple cameras, radar scanners and ultra sound scanners. Furthermore, methods for robust free space detection based on the data fusion framework will be investigated and refined using our RBC algorithm to identify objects in real world scenarios.

REFERENCES

- [1] BBC. (2011, April 2). Every death on every road in Great Britain from 1999 to 2010. Available: <http://www.bbc.co.uk/news/uk-15975564>
- [2] O.N.S., "Cycling to work in London," G. L. Authority, London, 2011.
- [3] S. Copsey, "A review of accidents and injuries to road transport drivers", EU-OSHA, Luxembourg, 2012.
- [4] N. Bernini, M. Bertozzi, L. Castangia, M. Patander, and M. Sabbatelli, "Real-time obstacle detection using stereo vision for autonomous ground vehicles: A survey," in 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2014, pp. 873-878.
- [5] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," IEEE Transactions on Intelligent Transportation Systems, vol. 14, pp. 1773-1795, 2013.
- [6] F. Dolins and R. Mitchell, "Spatial cognition, spatial perception: Mapping the self and space," Cambridge: Cambridge University Press, 2010.
- [7] R. Fleming, "Visual perception of materials and their properties," Vision Research, vol. 94, pp. 62-75, 2014.
- [8] J. H. Kim, W. Yang, J. Jo, P. Sincak, and H. Myung, Robot Intelligence Technology and Applications: Springer International Publishing, 2015.
- [9] S. Suresh, N. Sundararajan, and R. Savitha, Supervised Learning with Complex-valued Neural Networks: Springer Berlin Heidelberg, 2012.
- [10] B. Baruque, Fusion Methods for Unsupervised Learning Ensembles: Springer Berlin Heidelberg, 2010.
- [11] Y. Chen, M. Jahanshahi, P. Manjunatha, W. Gan, M. Abdelbarr, S. Masri, et al., "Inexpensive Multimodal Sensor Fusion System for Autonomous Data Acquisition of Road Surface Conditions," IEEE Sensors Journal, vol. 16, pp. 7731-7743, 2016.
- [12] VANTSEVICH AND M. BLUNDELL, ADVANCED AUTONOMOUS VEHICLE DESIGN FOR SEVERE ENVIRONMENTS: IOS PRESS, 2015.
- [13] J. Burack, The Oxford handbook of intellectual disability and development, 2nd ed. Oxford: Oxford University Press, 2012.
- [14] D. Waller and L. Nadel, Handbook of Spatial Cognition: American Psychological Association, 2013.
- [15] R. Chowdhury, T. Sharot, T. Wolfe, E. Düzel, and R. Dolan, "Optimistic update bias increases in older age," Psychological Medicine, vol. 44, pp. 2003-2012, 2014.
- [16] J. Badcock, "The Cognitive Neuropsychology of Auditory Hallucinations: A Parallel Auditory Pathways Framework," Schizophrenia Bulletin, vol. 36, pp. 576-584, 2010.
- [17] N. Wade, "Pioneers of eye movement research," i-Perception, vol. 1, pp. 33-68, 2010.
- [18] G. Dutton, "Cognitive vision, its disorders and differential diagnosis in adults and children: knowing where and what things are," Eye, vol. 17, pp. 289-304, Apr 2003.
- [19] J. Yun and S. Lee, "Human Movement Detection and Identification Using Pyroelectric Infrared Sensors," Biomedical Sensors and Systems, vol. 14, p. 24, 2014.
- [20] S. Monaco, G. Buckingham, I. Sperandio, and J. Crawford, Perceiving and Acting in the Real World: From Neural Activity to Behavior, Frontiers in Human Neuroscience 2016.

- [21] E. Martin, Concise colour medical dictionary, 3rd ed. Oxford: Oxford University Press, 2002.
- [22] J. Thomas, C. Moss, and M. Vater, Echolocation in bats and dolphins. Chicago: The University of Chicago Press, 2004.
- [23] T. Gudra, J. Furmankiewicz, and K. Herman, Bats Sonar Calls and its Application in Sonar Systems, Sonar Systems: InTechOpen 2011.
- [24] Surlykke, P. Nachtigall, R. Fay, and A. Popper, Biosonar: Springer, 2014.
- [25] Akademiya-nauk, S. O. B. Airapetyants, E. S. Konstantinov, and A. Ivanovich, Echolocation in animals. Jerusalem: IPST, 1973.
- [26] UNICEF, "Early Childhood Development: The key to a full and productive life," UNICEF, 2014.
- [27] H. C. Council, Supporting Children with Dyslexia: Taylor & Francis, 2016.
- [28] R. Arden, M. Trzaskowski, V. Garfield, and R. Plomin, "Genes Influence Young Children's Human Figure Drawings and Their Association With Intelligence a Decade Later," Psychological Science, vol. 25, pp. 1843-1850, 2014.
- [29] S. Miles, P. Fulbrook, and D. Mainwaring-Mägi, "Evaluation of Standardized Instruments for Use in Universal Screening of Very Early School-Age Children," Journal of Psychoeducational Assessment, 2016.
- [30] R. Beck, Motivation: theories and principles, 5th ed. ed. Upper Saddle River: Pearson Prentice Hall, 2004.
- [31] M. Ford, Motivating humans: goals, emotions, and personal agency beliefs. Newbury Park: Sage Publications, 1992.
- [32] D. Twigg and S. Garvis., "Exploring Art in Early Childhood Education," The International Journal of the Arts in Society, vol. 5, p. 12, 2010.
- [33] V. Löwenfeld and W. Brittain, Creative and Mental Growth: Macmillan(N.Y.), 1964.
- [34] Edwards, Drawing on the Right Side of the Brain: A Course in Enhancing Creativity and Artistic Confidence: Souvenir Press, 2013.
- [35] R. Siegler and E. Jenkins, How Children Discover New Strategies: Taylor & Francis, 2014.
- [36] Arel, D. Rose, and T. Karnowski, "Deep machine learning-a new frontier in artificial intelligence research," IEEE Computational Intelligence Magazine, vol. 5, pp. 13-18, 2010.
- [37] J. Heaton, Artificial Intelligence for Humans: Deep learning and neural networks: Heaton Research, Incorporated, 2015.
- [38] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network," in Neurocomputing: Algorithms, Architectures and Applications, F. F. Soulié and J. Héroult, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 41-50.
- [39] L. Iliadis, H. Papadopoulos, and C. Jayne, Engineering Applications of Neural Networks: Springer, 2013.
- [40] F. Roli and J. Kittler, Multiple Classifier Systems: Springer Berlin Heidelberg, 2003.
- [41] M. Nielsen, Neural Nets and Deep Learning: Determination Press, 2017.
- [42] V. Sgurev and M. Hadjiski, Intelligent Systems: From Theory to Practice: Springer Berlin Heidelberg, 2010.
- [43] G. Raidl, Applications of Evolutionary Computing. Essex: Springer, 2003.
- [44] Graves, Supervised Sequence Labelling with Recurrent Neural Networks: Springer Berlin Heidelberg, 2012.
- [45] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE Transactions on Neural Networks, vol. 5, pp. 157-166, 1994.
- [46] G. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," Neural Comput, vol. 18, pp. 1527-54, Jul 2006.
- [47] Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems: O'Reilly Media, 2017.
- [48] Lake, R. Salakhutdinov, J. Gross, and J. Tenenbaum, "One-shot learning of simple visual concepts," in Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Boston, Massachusetts, USA., 2011.
- [49] L. Fei-Fei, R. Fergus, and P. Perona, "A Bayesian Approach to Unsupervised One-Shot Learning of Object Categories," presented at the Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2, 2003.
- [50] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 594-611, 2006.
- [51] M. Palatucci, D. Pomerleau, G. Hinton, and T. M. Mitchell, "Zero-shot learning with semantic output codes," presented at the Proceedings of the 22nd International Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2009.
- [52] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese Neural Networks for One-shot Image Recognition," in International Conference on Machine Learning, Lille, France, 2015.
- [53] Biederman, "Recognition-by-components: a theory of human image understanding," Psychol Rev, vol. 94, pp. 115-47, Apr 1987.
- [54] E. Ronald, "Patterns of identity: hand block printed and resist-dyed textiles of rural Rajasthan," Ph. D, De Montfort University, 2012.
- [55] Tversky and K. Hemenway, "Objects, parts, and categories," J Exp Psychol Gen, vol. 113, pp. 169-97, 1984.
- [56] T. Binford, The vision laboratory. Cambridge: M.I.T. Project MAC Artificial Intelligence Laboratory, 1970.
- [57] R. A. Brooks, "Symbolic reasoning among 3-D models and 2-D images," Artificial Intelligence, vol. 17, pp. 285-348, 1981.
- [58] Guzman, Analysis of curved line drawings using context and global information. University of Edinburgh Press., 1971.
- [59] Marr and H. Nishihara, Representation and recognition of the spatial organization of three dimensional shapes. Cambridge: Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1977.
- [60] J. D'Errico, "A suite of minimal bounding objects," in Tools to compute minimal bounding circles, rectangles, triangles, spheres, circles., [Program], Matworks, 2014.
- [61] S. Rege, R. Memane, M. Phatak, and P. Agarwal, "2d geometric shape and color recognition using digital image processing," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering vol. 2, p. 8, 2013.