A Study of the Different Uses of Colour Channels for Traffic Sign Recognition on Hierarchical Temporal Memory

Wim J.C. Melis and Michitaka Kameyama
Graduate School of Information Sciences, Tohoku University
6-6-05, Aoba, Aramaki, Aoba-ku
Sendai, 980-8579, JAPAN

Abstract

When designing intelligence for a car many different tasks can be performed. Some of these tasks cannot easily be performed by conventional algorithms in comparison with the human brain. Recently, such intelligence has often been reached by using probability based systems. In this paper, Hierarchical Temporal Memory (HTM) is used to implement one of these tasks, namely traffic sign recognition. In implementing this traffic sign recognition task, it is noticed that the use of colour is of particular importance, and that colour information should be treated in a particular way to optimise the recognition. However it is also noticed that there are still a significant number of differences between the modelling of the brain and how the brain actually deals with colour and object recognition.

1 Introduction

Traffic sign recognition was previously based on object recognition in scenes [3], which in combination with the colour information could lead to recognition of the sign. Although real time performance could be achieved, these systems are still far from leading to an intelligent car. This would namely require other tasks to be implemented, like: driving distance and speed, recognition of other road users, anticipate the behaviour of other cars among many other. Although some of these tasks can be performed using conventional algorithms, not all of them can easily be implemented using these techniques. Therefore, the purpose of this research is to use a probability based system to perform these tasks. For this purpose, HTM [1, 2] was used, since it models the human brain. A further reason for this choice is in the fact that it is based on a hierarchical network, which would later allow for other tasks to be more easily combined into one system.

This paper starts with describing the HTM, and the implementation used for this specific application. The next section deals with the use of colour in HTM to improve the recognition accuracy, and shows results for representing the images in different ways. The differences in representation are in how the colour information is provided to the HTM network. Colour images can be split into different colour channels, e.g. R, G & B. Though other representations, like YUV might be more beneficial. These results are discussed together with a comparison on HTM's performance and its limitations in comparison to the human vision system. The paper finishes with conclusions and future work.

2 Hierarchical Temporal Memory

Hierarchical Temporal Memory models the functioning of the human brain. Its name refers to the main features, namely: the hierarchical structure, the large amount of memory it contains and the importance of time. This hierarchical structure can be seen for example in Figure 1 a) which represents a 2D network, generally applicable when the input data is one dimensional. For image processing, a 3D network as shown in Figure 2 is more commonly used. Any HTM network has its data fed in at the bottom of the hierarchy, and category data is output at the top node. This category data provides information on the classification performed by HTM. For example, when an HTM network has been trained with images of animals and their respective category information, then for each input provided, the system would output the category/animal which it considers to correspond with the provided input image.

Each HTM node within the hierarchy performs the same algorithm, but on different data. The node algorithm consists of two separate parts (see Figure 1 b)), namely a spatial and temporal pooler. These two parts correspond with the two main functionalities of the human brain. The spatial pooler is responsible for storing common input patterns in space. In case of an image application, the lower level nodes would remember adjacent pixels with a certain colour. Due to its hierarchical structure, the higher levels in the hierarchy also work on a higher level of abstraction. This allows for recognition of higher level objects. During the learning phase, patterns commonly provided to the spatial pooler are

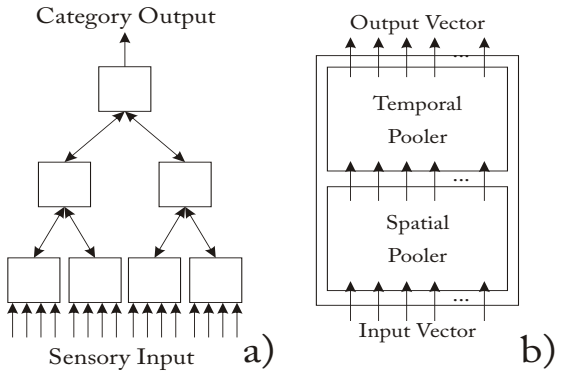

Figure 1: a) 2D Hierarchical Temporal Memory Structure; b) HTM Node Structure.

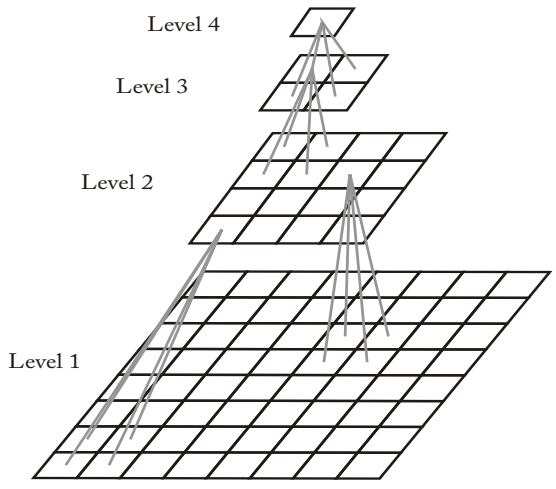

Figure 2: 3D Hierarchical Temporal Memory Structure

stored as coincidence patterns in the spatial pooler. During normal operation, the spatial pooler determines how closely the provided input matches with the stored patterns. The output of this spatial pooler forms input to the temporal pooler. This temporal pooler groups sequentially frequently occurring patterns. In the image application example, this part of the node would store whether a particular line moves from left to right or from bottom to top. During normal operation, the spatial pooler determines which group the coincidence from the spatial pooler belongs too.

3 HTM implementation using Colour Images

In daylight, the human eye has the capability of determining Red, Green and Blue separately. Though this does not seem to happen like in a digital camera. Depending on the consulted source a different grouping of colour is taken care of by one detector, though they all agree that there is not a single detector for a single colour. Although RGB is generally used in computer vision, it is known to suffer from light dependency.

Most applications implemented on HTM currently make use of greyscale images and/or in combination with a Gabor Filter. As will be seen later, using greyscale images only does not provide very good accuracy results. Furthermore, it was chosen not to use the Gabor filter due to its computational requirements. Since the eventual purpose of the application is hardware implementation, this computationally expensive pre-processing is considered as not beneficial. Instead colour information would be used, since traffic signs have quite some important information in the colours used. However, in computer vision using RGB leads to different colours being detected due to the light dependency. This in contrast to the human vision system that performs colour correction based on previous vision experience in combination with lighting information.

In order for HTM to deal with the different colour channels, multiple networks were fed with a greyscale reprensentation of a particular colour channel, each of these networks were then combined at the highest level leading to a combined decision on category. This combination could be performed in two ways, either the decision is based on a the highest input value or on the sum of results from the lower level networks. Although the provided images were of different size, before being provided to the HTM network, they were all scaled to the same size, namely 256 by 256.

Each of the networks were trained with the same data, which consisted of 12 different traffic signs. The number of signs was limited due to the significant increase in size of the networks when more signs were used, leading to extremely long simulation times. For each sign about 10 pictures were presented during the training phase (total 138 training images). The tests were then performed using the

Figure 3: Example images from: a) Training set; b) Test set; c) Difficult test set.

training data, as well as two other data sets not used during training. The first one having a similar size to the training set (134 images), and the second set containing only 64 images. The test with the training data allows to conclude on how well the network learned. The other data-sets give an indication of how well the network performs. Example pictures from the used data sets can be seen in Figure 3. Considering that HTM documentation advises to include all possible recognition angles in the training data, the currently used training set was very small. Since it was not the purpose to test the learning and recognition capability of HTM, but to focus on how to deal with colour, a larger training set was not generated. These results do however indicate another difference between HTM and the human vision system, which is that the human vision system can learn when seeing items in one particular angle, and then recognise from a different angle [4].

The HTM temporal pooler algorithm can make use of different algorithms, although these different algorithms were tested, the differences were small, and therefore only results for the Time Based Inference algorithm are presented, since this is the only algorithm that is further developed in the Numenta software.

Since these systems are probability based, and might therefore provide more reliable results when multiple networks are used, the greyscale approach was also repeated using multiple networks each time fed with the same data, in order to allow for a fair comparison with networks being fed with different colour channels.

As can be seen from Table 1 the results using the different colour approaches and algorithms does not lead to significant difference in accuracy when the training data is used, which implies that the learning is performed equally well for each of the networks. The differences become slightly more distinguishable when using the new data sets. The accuracy of the Difficult data set are particularly low. This is quite understandable taking that this data set consisted of a set of images, which would are extreme challenges for the network in comparison to the training data. This data set namely consists of images taken from extreme angles, as well as a large set of night images, which were then exposure corrected at different levels. The exposure correction was performed to model the human eye changing pupil size to capture more light in dark environments. However, no clear relation could be detected between the amount of exposure correction and the likelihood of correct classification.

In relation to using greyscale images as input, it is clear that using multiple networks can indeed improve the recognition accuracy. When using colour images, either based around RGB-channels or YUV-channels more differences can be noticed. Especially for the test sets, the RGB approach seems to perform slightly worse than the YUV based approaches, which is because the latter one separates luminance into one channel. Although YUV separates light into the Y channel, each of the channels are still calculated out of the RGB-channels, which makes them not completely independent of light changes.

As can be seen from the sample images in Figure 3, the sample images consisted of book type images as well as images of traffic sign along the roads and were gathered from various sources in the internet as well as personally taken images. In the latter case the background would clearly need to be eliminated. When converting the images to YUV, it seemed that the U and V channel already cancelled out quite some background (see Figure 4), which clearly aids the network in recognising the sign.

Although current studies indicate that the human eye registers closer to the RGB principle, it seems there are a large set of corrections performed on the data coming from the eye before object recognition takes place. The human brain being a system with a large amount of feedback, each of these processes are extensively linked with one another. To further distinguish between background and object, the human vision has two additional features. The first is that the eyes have one point of focus and a large pheripheral vision. This allows for detection of objects in the pheripheral vision followed by focusing on them later on. Secondly, the data from two eyes is combined to construct 3D information about the viewed scene. This allows for distance calculation, as well as object recognition in the sense of it being a flat surface, 3D object, which lighting is affecting the scene among others. This extra information clearly helps the recognition process.

Furthermore, the results with the difficult data set also indicate that greyscale images have better accuracy in dark

Table 1: Results for different colour HTM implementations of traffic sign recognition network

Data Set/Top Level Prob.	Greyscale	Double Greyscale	Triple Greyscale	GB	RB	RG	RGB	RGB Greyscale	UV	YU	YUV	YV
Train. Set, Sum	95.65%	95.65%	95.56%	93.47%	94.92%	93.47%	94.20%	94.20%	95.65%	94.92%	95.65%	94.92%
Train. Set, Max	94.20%	94.20%	94.20%	93.48%	94.93%	93.48%	94.20%	94.20%	95.65%	94.93%	95.65%	94.73%
New Set, Sum	14.93%	16.42%	16.42%	19.40%	17.16%	16.42%	17.92%	18.66%	19.40%	19.40%	20.15%	20.90%
New Set, Max	18.66%	18.66%	18.66%	20.15%	20.90%	20.15%	20.90%	20.15%	22.39%	19.40%	21.64%	20.90%
Diff. Set, Sum	7.81%	4.69%	4.69%	6.25%	4.69%	3.13%	3.13%	1.56%	0%	7.81%	4.69%	6.25%
Diff. Set, Max	1.56%	1.56%	1.56%	1.56%	1.56%	0%	0%	0%	0%	1.56%	1.56%	0%

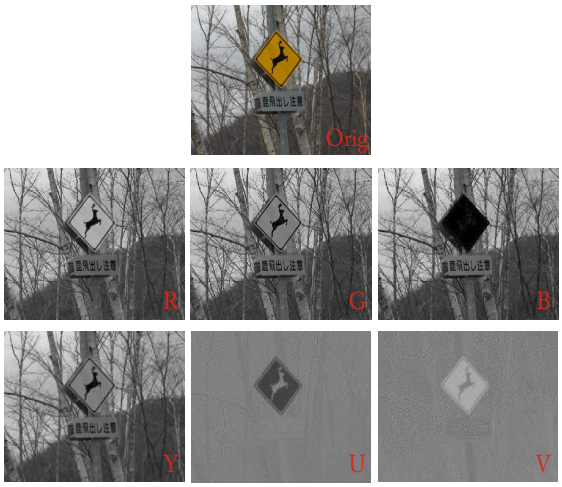

Figure 4: Original image and its R,G,B and Y,U,V components. The U & V components clearly cancel out background in favour of the traffic sign.

conditions. This is similar to the human eye, which in dark environments does not really use colour information. Since the eye and brain mainly focus on shapes instead of colour. Colour information is possibly added in the brain based on previous experience and knowledge.

As can be noticed, there are a significant number of differences between the computer vision with HTM and the human vision. These factors influence the accuracy results of the computer vision system. Most of these differences can be performed by modifying the images provided as input to the HTM network. Considering that the brain provides this functionality, there might also be a way to train some system how to deal with different type input and still present it in the same way to the network performing the object recognition.

4 Conclusions

It was shown that when implementing image recognition for HTM, colour information could lead to improved recognition in comparison to a single greyscale image. Further improvements are possible if lightning conditions could be taken into account.

Future work consists of two parts, firstly, steps will be taken to improve the accuracy of HTM, by bringing it closer to the human vision system. In the paper quite some differences were discussed. Overcoming these differences will not only improve accuracy, but should also reduce the storage requirements of HTM especially when the training data can be reduced to images from one angle only.

The second part of future work will focus on other tasks required for an intelligent vehicle. This includes tools to determine distance, as well as other road users. This could then be coupled with information available about the vehicles position, speed among other to improve the driving experience.

References

[1] D. George and B. Jaros. The HTM learning algorithms, 1/03/2007 2007. `http://www.numenta.com/for-developers/education/Numenta_HTM_Learning_Algos.pdf`.

[2] J. Hawkins and D. George. Hierarchical temporal memory: Concepts, theory, and terminology, 3/27/2007 2007. `http://www.numenta.com/Numenta_HTM_Concepts.pdf`.

[3] L. Priese, J. Klieber, R. Lakmann, V. Rehrmann, and R. Schian. New results on traffic sign recognition. In *Proceedings of the Intelligent Vehicles Symposium '94*, pages 249–254, 1994.

[4] D. Vishwanath, A. R. Girshick, and M. S. Banks. Why pictures look right when viewed from the wrong place. *Nature Neuroscience*, 8(10):1401–1410, 2005.