Improved Visual Perception for Conceptual Colour Understanding in ROS

Kalp Garg and Xie Ming

Abstract— **Colour is essential element in visual language. Early stages of vision requires popping it out at viewers. It is essential in various areas such as computer graphics, image processing etc. Understanding color enhances visual search, improve object recognition, convey structures, improve usability, show associations, communicate mood. Since it is important and powerful form of communication among human beings, robots employing in real world that coexist with human beings should also have a good perception of colors. Many work that occurs in this area rely upon modelling of human color perception and mathematical complexities. This work demonstrates machine learning approaches to make computer learn colors through interaction from the humans. Finally this algorithm has been implemented into NTU Singaboat, which are Unmanned Surface Vehicle (USV) developed for Maritime RobotX challenge.**

I. INTRODUCTION

Colours play a vital role in the world we live in. This is the reason that in early childhood colours and shapes are being taught first to identify things around us. Our mind doesnt know the name of green trees, blue sky, square windows, etc., but it is noticing them. This is the reason why the red colour is generally associated with danger sign in some countries. This is evident from the fact that even when you are not directly looking at the red object from your eyes but you can notice it with the corner of your eyes much sooner than any other different colour object. Colour also plays an important part in recognising various objects and classifying them apart from other attributes such as name, shape, etc.

In human society, sociable robot aimed to live to coexist with the human [1]. As a new member in human society, it is expected to become a coworker and actively participated in social participation with humans. Thus this robot, to be able to interact properly with the human on the ones related to colours, must learn colours from the society it belongs [2].

Huge research is going on the development of Unmanned Surface Vehicle(USV) in many areas such as military, commercial, industrial, because of its ability to free human from hazardous works. In regard to this, an autonomous vehicle is being built by Nanyang Technological University(NTU) for Maritime RobotX Challenge which is conducted by Association for Unmanned Vehicle Systems International (AUVSI) Foundation, USA. There are different tasks to be performed by the vehicle on water. Some of the tasks such as 'find totems', 'scan the code', 'identify symbols and dock', 'detect and deliver' require correct colour identification by the vehicle to properly complete its task [3].

Kalp Garg is with Department of Electrical Engineering, Indian Institute of Technology Roorkee, India kalpgarg.12@gmail.com

Xie Ming is with the Faculty of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore mmxie@ntu.edu.sg

II. PREVIOUS WORK

Generally, clustering algorithms are being developed to achieve colour learning by computer. Artificial Neural Network (ANN) techniques are used because of its learning capabilities, by which network adapt itself to the input-to-category connection so that when input arrives, a correct category can be invoked [4]. As it also posses hidden layers between input and output layers, non-linearity between the output categories can also be learned and these layers can be interpreted as an implicit representation of input [5][6]. RCE neural network for colour categorization for hand segmentation is being used by Yin [7], which further can be extended for colour-based image segmentation. A combination of SOM(Self-organising Map) and ART(Adaptive resonance theory) neural network is being used by Yeo[8]. Dowman[9] uses a fuzzy Bayesian system to simulate colour acquisition and evolution of colour categorization. Steels and Belpaeme [10] developed various colour categorization systems, modelling them as well as give comparison among them. Works such as [11] [12] [13] also use ANN for colour learning. Sometimes, mathematical complexity on clustering algorithm alone doesn't necessarily result in proper colour learning shared with the human.

III. CONCEPTUAL LEARNING OF COLORS

A. Perception Space

Perception Space (or in simple words input space) is defined as how colours are perceived and how are they organised based on which colour categorization is made. This is why it is an important factor to achieve colour learning. HSV (Hue, Saturation, Value), RGB (Red, Green, Blue), CIELAB (Colour space defined by International Commission on Illumination, L- lightness, A-green-red, B-blue-yellow component) are commonly used perception spaces. RGB colour space is being used in this work. When red, green and blue light is added together in various ways, it produces a broad array of colours. This shows that RGB colour model is an additive type of colour model [14]. Therefore every colour seen in the world belongs to RGB cube [15]. Usually, 8-bit binary values are used to define the scale of the cube. Thus, this scale allows representation of any colour in three bytes, one for red, green, blue component. The major reason for choosing RGB colour space over others is that retina of the human visual system contains three types of cone cells. Their sensitivity to light varies across the visible spectrum. Cone cells known as rho receptor system are very sensitive to red light (around 600nm), whereas gamma receptor type cone cells are sensitive to medium wavelengths (around 550nm), which appears greenish. Beta receptors are sensitive to short wavelengths (around 450nm) of the visible spectrum, which appears bluish. To build all colours, the eye sees all the colours in the visible spectrum in a continuous way from red to violet, so it is the mixture of these colours that the brain is able to produce. Thus RGB colour model is invented to match the functioning of the human eye. Even the monitor only see in RGB.

B. Approach

Learning of colours doesn't completely rely on mathematics in this work. Rather a new approach, that is, making the computer learn colours through interaction with humans has been adopted so that an effective and efficient social interface between human and computer can be established. Therefore, in this work, in the beginning, the computer doesn't know any colour, but eventually through interaction, it learns numerous colours which indicate shred learning of colours with humans in society.

K-means clustering is perhaps most widely used clustering algorithm that minimizes the cost function. It is more robust and versatile than another classifier such as Artificial Neural networks(ANN) and Support Vector Machine(SVM). The procedure of it is very simple and aims at minimizing cost (objective) function, in this case, a squared error function. Generally, KNN is non-parametric, instance-based and used in a supervised learning setting. In the classification task, KNN algorithm calculated the majority vote between the k-most similar instances to a given "unseen" observation. The similarity is defined by Euclidean distance metric between two data points. Euclidean distance is given by

$$d(x, x^{'}) = \sqrt{(x_1 - x_1^{'})^2 + (x_2 - x_2^{'})^2 + ... + (x_n - x_n^{'})^2}$$

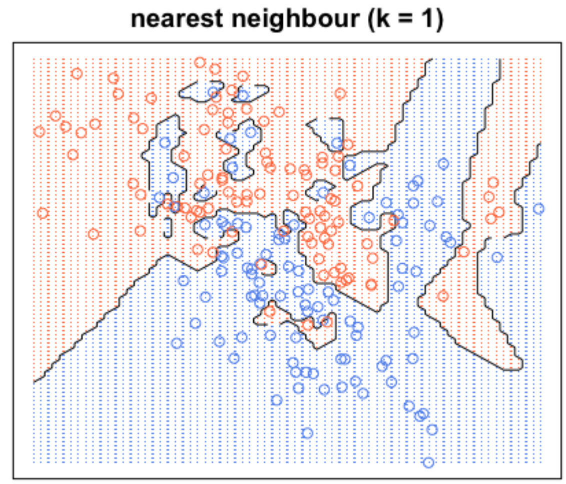
Therefore, given a positive integer K, an unseen observation (test data) x and a similarity metric d, KNN performs following steps:

1. We'll first find K points in the training data set that are closet to x in the set A by computing d between x and each training observation. To prevent odd situations, usually, K is chosen to be odd.

2. It then estimates the probability for each class, that is the fraction of the points in A with that given class label.

$$P(y = j | X = x) = \frac{1}{K} \sum_{i \epsilon A} I(y^{(i)} = j)$$

where $I(x)$ is an functions whose value is 1 when $x = 0$ and 0 otherwise.
Finally, new observation x is being assigned to class with largest probabiltiy.



thpb

Fig. 1. Classification of dataset for $K = 1$

C. Learning through interaction

ROS (Robot Operating System) provides tools, libraries and some other functionalities to develop software related to robot applications. By using KNN and integrating it image processing platform OpenCV in ROS, we can make the computer learn colours through human interaction and this work can be directly integrated with any physical robot. This GUI can be implemented on any visual device which is publishing its ROS messages as *camera image*. This GUI is made as a ROS node which subscribes to the *camera image* messages. For the purpose of colour learning, a total of 614 instances has been produced for the creation of training dataset. Laptop webcam is used as an input device for camera messages to ROS node. Therefore each time an coloured image is being shown to the computer through webcam and colour name is being inputted by the user. Using OpenCV, the computer calculates average Red, Green and Blue value of the image and finally, this instance is being saved into the dataset with the label of the colour (inputted by the user). This RGB value is being divided into nine big classes of colour namely grey, blue, brown, green, orange, red, violet, white and yellow.

D. Choosing best K

K in KNN is a parameter which should be chosen to obtain best possible fit for the data set. Intuitively, K controls the shape of decision boundaries. When K is small, the classifier will be "blinder" for overall distribution providing a flexible fit. It will have a low bias but high variance. Our decision boundary will be more broken if looked graphically as shown in Figure 1 [16].
On the other hand, a higher K will have smoother decision boundaries which mean lower variance but increased bias as shown in Figure 2 [16].

In this work, best K is being selected by dividing the training dataset into training and validation dataset in the

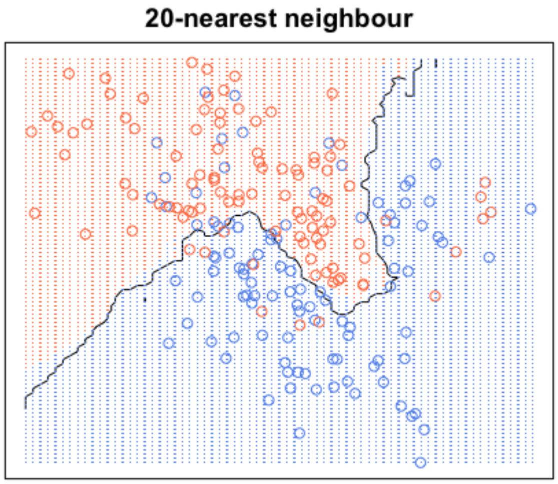


Fig. 2. Classification of dataset for $K = 20$

ratio of $3:1$ in the random fashion. Whenever dataset is divided into training and validation set, for each value of K from 1 to 10, accuracy in the validation set is being determined. Every time, best K which gives best possible fit for our dataset comes out to be $K = 2$ which gives an accuracy of over 90%. Since the odd number is usually chosen (due to previously cited reasons), therefore K is chosen to equal to 3.

E. Decision Boundary

Data with 614 instances is as shown in Figure 3. To interpret decision boundaries, we can imagine a 3-dimensional spheres surrounding classes of each colour. These spheres can be non-linear in a way so to improve generalisation and incorporate real-world behaviour of colour learning. This spheres can be assumed to have a centre and radius which is represented by mean and standard deviation quantitatively. Every time we add an instance to dataset or computer trying to learn a new colour, new centre and radius is being determined by the computer and accordingly new decision boundaries have been made.

F. GUI

Figure 4 shows a GUI developed is ROS to incorporate colour learning through human interaction. After our node subscribes to *camera image* messages, a camera display window opens at the bottom. Now suppose is we want our computer to learn blue colour, then by showing the colour in front of camera, we can make him learn that colour by writing /textitblue in the colour name text box. After adding any number of observations you want, finally to train the network, click on train button and computer will find the best possible K for given dataset. This K will be used further during the test.

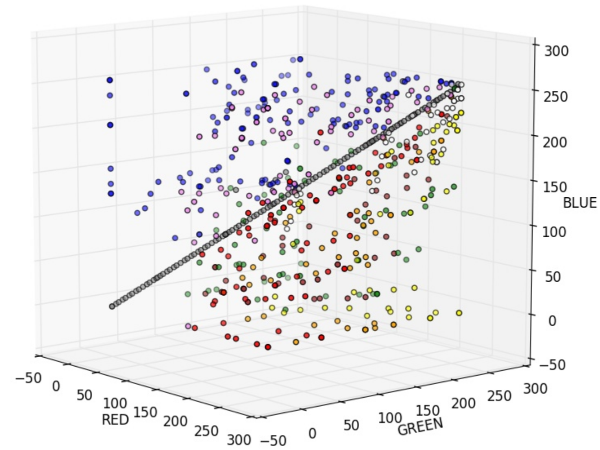


Fig. 3. Plot for RGB values and Colour Name

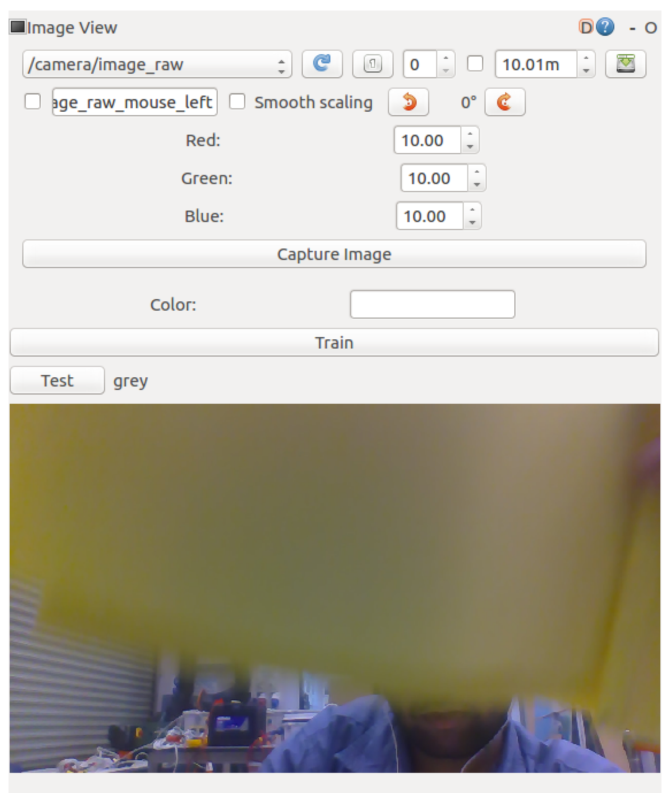


Fig. 4. GUI made for color learning in RQt

TABLE I

CONFUSION MATRIX

	grey	blue	brown	green	orange	red	violet	white	yellow
grey	12	0	0	0	0	0	0	0	0
blue	0	0	0	0	0	0	0	0	0
brown	0	0	0	0	0	0	0	0	0
green	0	0	0	0	0	0	0	0	0
orange	0	0	0	0	0	1	0	0	0
red	0	0	1	0	0	0	0	0	0
violet	0	0	0	0	0	0	0	0	0
white	0	0	0	0	0	0	0	0	0
yellow	0	0	0	0	0	0	0	0	0

IV. RESULTS AND DISCUSSIONS

A confusion matrix is generally used to describe the performance of a classifier. Training data is employed on the model to obtain it. A confusion matrix as shown in I is obtained out of testing dataset containing 30 instances of RGB colour values along with their colour name. As we can see, our trained classifier gives good results with success rate of above 90% for test data. One of many implementations of this work is in USV (Unmanned Surface Vehicle) built by NTU (Nanyang Technological University) for Maritime RobotX Challenge. As stated previously, there are many tasks to be performed by the vehicle which requires proper colour recognition. One of it is 'detect and deliver', in which vehicle has to insert objects on one of the faces of the four-sided floating platform. Since all faces have the different colour, therefore vehicle has to detect right colour autonomously and then has to complete his task. There were several cameras mounted on the USV to gather visual information. The image information from the camera is gathered by the onboard computer which sends it directly to the master computer using ROS (Robot Operating System).

REFERENCES

[1] C. Breazeal, *"Toward Sociable Robot, Robotics and Autonomous Systems"*, vol. 42, pp. 167-175, 2003.

[2] Rinaldo Christian Tanumara, Xie M., Au Chi Kit, 2006, *Learning Human-like Color Categorization through Interaction*, International Journal of Computational Intelligence, Vol. 3, No. 4, pp. 338-345.

[3] http://robotx.org/images/files/RobotX_2018_Task_Summary.pdf

[4] S. Harnad, S.J. Hanson, and J. Lubin,*Learned categorical perception in neural nets: Implication for symbol grounding, in Symbolic Processors and Connectionist Network Models in Artificial Intelligence and Cognitive Modeling: Steps Toward Principled Integration*, V. Honavar and L. Uhr, Eds. Boston: Academic Press, 1995, pp. 191-206.

[5] S. Usui, S. Nakauchi, and M. Nakano,*Reconstruction of Munsell Color Space by a Five Layered Neural Network*, presented at IEEE International Conference on Neural Network, 1990.

[6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams,*Learning internal representation by back-propagating errors*, Nature, vol. 323, pp. 533536, 1986.

[7] X. Yin, D. Guo, and M. Xie, *Hand image segmentation using color and RCE neural network, Robotics and Autonomous Systems*, vol. 34, pp. 235-250, 2001.

[8] N. C. Yeo, K. H. Lee, Y. V. Venkatesh, and S. H. Ong, *Colour image segmentation using the self-organizing map and adaptive resonance theory*, Image and Vision Computing, vol. 23, pp. 1060-1079, 2005.

[9] M. Dowman, *"Colour terms, syntax and bayes modelling acquisition and evolution"*, PhD Thesis, University of Sydney, 2004.

[10] L. Steels and T. Belpame, *"Coordinating perceptually grounded categories through language: A case study for colour,"* Behavioral and Brain Sciences, vol. 28, pp. 469-529, 2005.

[11] A. H. Dekker, *Kohonen neural network for optimal color quantization, Network: Computation in Neural System*, vol. 5, pp. 351-367, 1994.

[12] T. Uchiyama and M. A. Arbib, *Color Image Segmentation Using Competitive Learning*, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 16, no. 12, pp. 1197-1206, 1994.

[13] A. Verikas, K. Malmqvist, M. Bachauskene, L. Bergman, and K. Nilsson, *Hierarchical neural network for color classification*, presented at IEEE International Conference on Neural Network, 1994.

[14] Poynton, Charles A, *Digital Video and HDTV: Algorithms and Interfaces. Morgan Kaufmann*. ISBN1-55860-792-7,2003.

[15] Alejandro C. Frery, Carlos Alberi dos Santos Melo, Rodrigo C. Fernandes, *Web-Based Interactive Dynamics for Color Models Learning, Color Research and Application*, vol. 252, issue 6, 2000.

[16] https://kevinzakka.github.io/2016/07/13/k-nearest-neighbor/

[17] D. Boley and R. Maier, "A Parallel QR Algorithm for the Non-Symmetric Eigenvalue Algorithm", *in Third SIAM Conference on Applied Linear Algebra*, Madison, WI, 1988, pp. A20.