Development of a NAO Humanoid based Medical Assistant

Aditya Kumar
Department of Mechanical Engineering
Indian Institute of Technology, Guwahati
P.O. 781039, Assam, India
aditya.kumar@iitg.ernet.in

Atman Patel
Department of Mechanical Engineering
Indian Institute of Technology, Guwahati
P.O. 781039, Assam, India
p.atman@iitg.ernet.in

Santosha K. Dwivedy
Department of Mechanical Engineering
Indian Institute of Technology, Guwahati
P.O. 781039, Assam, India
dwivedy@iitg.ernet.in

ABSTRACT

In this paper[1], socially assistive human-robot interaction has been explored on a NAO Humanoid Robot in order to automate the pharmacy and bio-medical sector, with a broader aim of addressing all similar tasks. The problem has been divided into three sub-segments viz. pick and place operation with smooth gripping mechanism, reading printed and handwritten text from prescriptions, and use of smart detection technology, with a focus on barcode detection to locate target objects (medicine flaps) in real time. Iterative Jacobian Pseudo inverse kinematics algorithm is implemented to calculate the joint angles. To account for the poor performance of Google Tesseract for handwritten text, the image contrast is enhanced for histogram equalization and fed to maximally stable extremal regions (MSER) algorithm in a combination with Stroke Width Transform (SWT) to make text detection more robust even in presence of blur, before feeding it to Tesseract. Lastly, two techniques are developed to incorporate barcode integration with NAO, first using ALBarcodeReader API, the limitations of which are solved using vertical descent sobel scharr operator to attain real time barcode scanning for all types of barcodes.

CCS CONCEPTS

•**Computing methodologies ~Object identification** •*Computing methodologies~Motion path planning* • *Computing methodologies~Image processing* • **Computing methodologies~Perception** • *Computer systems organization~Robotic autonomy*

KEYWORDS

NAO Humanoid Robot, MSER, Barcode, text detection

ACM Reference format:

A. Kumar, P. Atman, and S.K. Dwivedy. 2017. SIG Proceedings Paper in word Format. In *Proceedings of ACM Advances in Robotics conference, Indian Institute of Technology, Delhi, Delhi India, June 2017 (AIR'17)*, 6 pages.
DOI: https://doi.org/10.1145/3132446.3134899

1 INTRODUCTION

Role of robotics is substantial in rehabilitation [1], neuroscience [2], social assistance [3] and medical treatments like autism [4] in children, where robots are found extremely useful in investigating the emotional, social and communication deficits. For example, Kismet [5] developed by MIT can produce a range of facial expressions, and NAO humanoid robot [6] developed by the French company Aldebaran Robotics can help in assistive therapy of autistic children [7].

Robots can greatly enhance healthcare automation, to attain low labor cost, higher throughput, and improved quality and consistency. Focus is cost reduction and efficiency. In a pharmacy, the robot can help the operator by scrutinizing the prescriptions for the medicine name and then supplying the same to the customer. Inside a hospital, in the patient's cabin, a caretaker say nurse may not be present at all times, but the patient can have any urgent need that has to be immediately fulfilled. While it's understandable that humans can have no complete replacement, the robots can still help in maximum of the tasks such that the human can use that time efficiently in developing new technology, than being stuck up with redundant tasks. Also, it's not advisable for the caretakers to be present in ICUs where high degree of purity is required. Robots will be most efficient as medical assistants in these scenarios.

In this work, NAO Humanoid is being used as the test bed platform, as it is affordable and has great performance and modularity [6]. NAO is trained to perform three basic operations i.e., pick and place operation using gripping mechanism, text recognition, and identification of the detected medicine using automated detection techniques. After a survey on RFID (Radio Frequency Identity Detection), Barcode scanning, OCR (Optical character recognition), MICR (Magnetic Ink character

1.

AIR '17, June 28-July 2, 2017, New Delhi, India

ACM ISBN 978-1-4503-5294-9/17/06…$15.00
https://doi.org/10.1145/3132446.3134899

recognition), it was found out that barcode would be the most efficient to implement as it is much smaller, lighter, less expensive, more accurate and ubiquitous. OCR can detect the medicine name in the prescription very efficiently, if it is not handwritten text i.e. it is printed text. Identifying name from handwritten prescription and complex background uses machine learning advanced algorithms like use of classifiers and support vector machines [8], for handwriting identification in noisy documented images [9]. In this paper, text recognition approach has been made simple using MSER algorithms (textural features) due to its invariance to affine transformation (similar to linear transformations, but not exhaustively) of image intensities, along with Stroke Width Transform (SWT) and other geometric features. This combination reduces the problem of blur in images (which is a limitation of MSER) and also deals with the relatively low accuracy [10] and higher time complexity of support vector machines (SVM).Barcode scanning has been integrated with NAO using two techniques. First is using library ALBarcodeReader in upgraded NAO robot version 2.1, and second is training NAO's vision system to use vertical descent algorithms to attain relevant information.

2 LITERATURE REVIEW

Human-robot interaction has been quite an interesting and engaging research field for scientists all across the globe, especially in the previous decade. Some relevant breakthroughs have been covered as under.

Kwang et.al [8] presented a unique texture-based method for detecting texts in images, without the use of any feature extraction module. A support vector machine (SVM) was trained to analyze the textural properties of texts, where text regions were identified by invoking a continuously adaptive mean shift algorithm (CAMSHIFT). Yefeng et.al [9] addressed the problem of the identification of text in noisy document images, by modeling the noise separately and then using classifiers for segmenting and identifying between handwritten and typewritten text. Chen et.al [11] performed text detection using edge enhanced maximally stable extremal regions algorithm for natural images. Matas et.al [12] worked on establishing correspondence between two images using maximally stable extremal regions, due to their linear complexity to affine transformations. Li et.al [13] performed scene detection using Stroke Width Transforms (SWT) by incorporating binary thinning operations together with distance transforms.

Laurent George et.al [14] proposed a method for humanoid robot navigation, using 2D barcodes. They used embedded hardware to realize that not only was this technique usable in domestic environments but could also navigate a complex outdoor environment. Jan et. al [15] worked on experimental evaluation of vision based on-map localization procedures that apply QR-codes or NAO marks, as implemented in service robot control systems, chiefly to detect obstacles. They developed a vision based detection system which treats the NAO marks and the QR codes as landmarks. Sherin et. al [16] proposed a smart barcode detection and recognition system which can detect barcodes even in low quality images, based on fast hierarchical Hough transform (HHT), where the recognition is performed by training a back propagation neural network.

The main contribution of this paper are as follows:

- Complete end-to-end process is presented for a fully automated functional medical assistant, ranging from kinematics, grasping to recognition and perception.

- Maximally stable extremal regions (MSER) algorithms are implemented prior to the K-means and adaptive recognition algorithms for Optical Character recognition, as MSER worst-case time complexity is $O(n)$, in contrast to Support vector machines with $O(n^3)$ time complexity [17] , where n is the number of pixels in the image. This is simpler than SVM and produces almost similar results for handwritten text, if the image quality is high.

- Barcode scanning algorithm is developed and fully integrated in NAO for all types of barcodes, using Vertical descent sobel Scharr operator. The API present in NAO SDK can only detect QR codes, and so poses constraints, since majority pharmaceutical companies use linear barcodes.

3 PICK AND PLACE OPERATION

The fact that the location of medicines inside a pharmacy are approximately known is used for assistance in reaching the target. Thus, after feeding the location of different medicines, finding the target medicine becomes an easier task once the name is precisely identified from the prescription. The humanoid can be guided to an estimated position in front of the medicine, from where barcode reading algorithm is used to identify the desired medicinal flap from the bundle.

To grab the medicine from the bundle, the location of the flap or bottle needs to be determined. Using the Microsoft Kinect, the precise coordinates of the barcode (attached to the target object) can be determined easily. In our experiments, a fixed location of the target object is assumed for simplicity. In real world frame, the relative coordinates of target object with respect to the humanoid keeps on changing as the humanoid moves closer to it. As the initial relative coordinates are known, the final relative coordinates, when the humanoid is in front of the target object, can be found out by calculating the shift of the humanoid from initial position using magnetic rotary encoders present in hip, knee and ankle joints. To finally grab the target object, Jacobian Pseudo Inverse method [18] has been used to calculate the required joint values. Although the Jacobian Pseudo Inverse method is prone to poor performance due to instability near singularities, it shows good accuracy in our case because of the presence of only a single end-effector and a stationary target. Fig. 1 shows the sequential order of pick and place operation using the Pseudo inverse algorithm briefly described below.

Let the end effector position in Cartesian coordinates be defined by $X = [x, y, z]$ and the angles be defined by $\theta = [\theta_1, \theta_2, \ldots \theta_n]$where n is the degree of freedom of the manipulator (in our case, *n=25*). It is known that$X = f(\theta)$.

$\frac{dX}{d\theta} = J(\theta)$, where $J(\theta)$is the Jacobian, given by (1)

$$J = \begin{bmatrix} \frac{\partial x}{\partial \theta_1} & \frac{\partial x}{\partial \theta_2} & \frac{\partial x}{\partial \theta_n} \\ \frac{\partial y}{\partial \theta_1} & \frac{\partial y}{\partial \theta_2} & \frac{\partial y}{\partial \theta_n} \\ \frac{\partial z}{\partial \theta_1} & \frac{\partial z}{\partial \theta_2} & \frac{\partial z}{\partial \theta_n} \end{bmatrix} \text{, a } 3Xn\text{matrix} \quad (2)$$

As Jacobian is not necessarily a square matrix, the inverse can't be found out directly. Thus pseudo inverse is one of the methods devised to find the inverse of Jacobian

$$dX = J\, d\theta \quad (3)$$

$$J^T dX = J^T J\, d\theta \quad (4)$$

$$(J^T J)^{-1} J^T\, dX = d\theta \quad (5)$$

$$J^+\, dX = d\theta \quad (6)$$

where,

$$J^+ = \text{pseudo inverse}$$

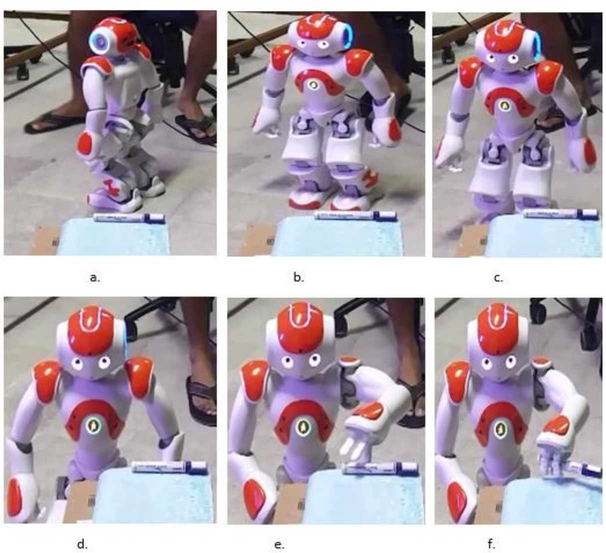


Figure 1: NAO performing pick and place operation

4 TEXT RECOGNITION FROM PRESCRIPTION FOR MEDICINE NAME

Optical Character Recognition (OCR) detects one glyph or character at a time. As Fig. 2 shows, input to an OCR is the image which undergoes optical scanning, segmentation, pre-processing, feature extraction and post processing (i.e., grouping the characters into meaningful words) to produce the editable digital file as output. Tesseract developed by Google is one of the finest OCR software because it incorporates an additional adaptive Recognition algorithm to learn the symbols passed in the first go to improve performance in the second pass, while using the conventional pattern matching/ image correlation, like K-nearest neighbor algorithm in the first pass. While it works best with typewritten text, even Tesseract has limitations while working with handwritten text, which needs to be addressed and the recognition needs to be made intelligent.

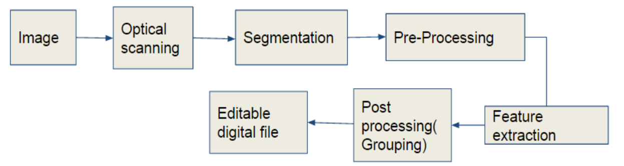


Figure 2: Layout of an OCR Process

To approach the problem of handwritten text identification in a simple manner, algorithms were developed in MATLAB, wherein MSER is employed to segment the non text regions which are removed using Stroke Width Transforms (SWT) and geometric features. SWT is a very stable algorithm and works on the concept that majority textural features follow a particular pattern of stroke, which is nearly constant, while non text regions have great variation in stroke. However, SWT is not very accurate as MSER, as can be seen from Fig 3. Fig 3A represents the regions detected by MSER, while Fig 3B represents the regions detected by SWT. It can be seen that MSER is more accurate. However, to make the detection very precise, a combination of both has been used in this paper.

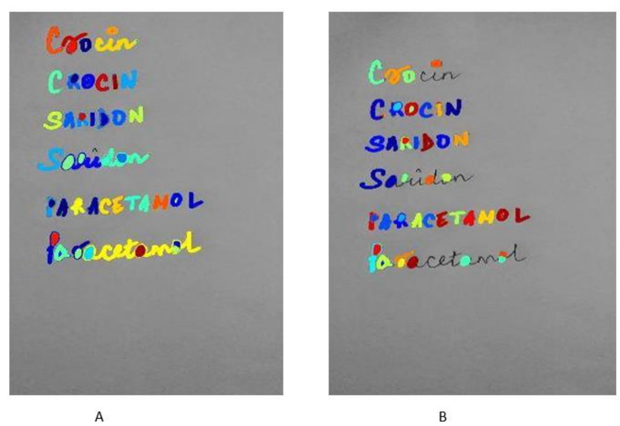
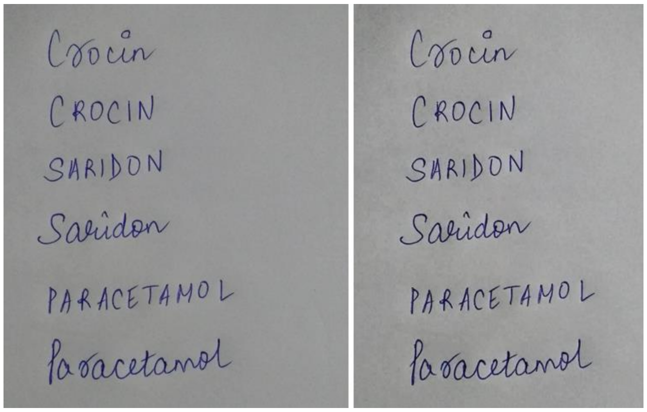


Figure 3. Text detection using A] MSER and B] SWT on a prescription template

The text regions are then merged for final detection result and then the detected text is post-processed (grouped) using OCR. What is novel here is that the image contrast is initially increased (Fig 4) before feeding it to MSER, to reduce its sensitivity to blur. Increasing the contrast results in histogram equalization and enables accurate detection in blurred regions whose edges are difficult to determine by gauging intensity alone (using canny edges). Fig 5 shows the corresponding change due to increase in contrast, which enhances text detection.



A B

Figure 4: **Raising Image entropy by enhancing contrast (B)**

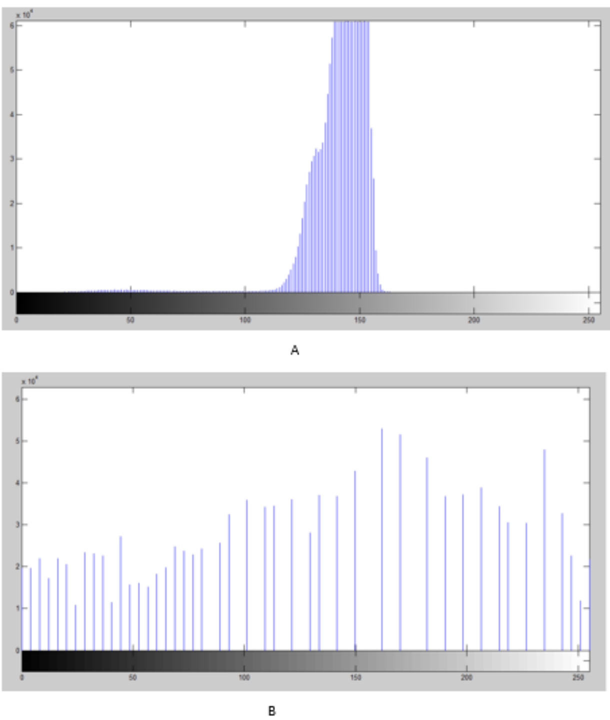


Figure 5: **Histogram Equalization by contrast enhancement. A is the initial intensity histogram, and B is histogram of the Fig. 4B, which has greater contrast.**

After the medicine name has been identified from the prescription, NAO can use automated smart detection technology like RFID (Radio Frequency Identity Detection), Barcode detection, or MICR (Magnetic Ink Character Recognition), to detect the medicine inside the pharmacy, at its respective place. Both RFID and Barcode are highly exhaustive and very secure, hence are preferred over MICR.

Barcode scanning is chosen, since it is portable, small, light and has cheap implementation cost. It is universal and has accurate detection, notwithstanding the fact that it has limitations to be in line of sight (LOS) with the scanner and scanner should be close enough in the range of ~15 feet. RFID on the other hand is expensive due to computerized chip and transponder installation. Also, tag and reader collision can occur due to presence of multiple transponders and signal interference.

So now, each medicine flap will be having an associated barcode and NAO will detect using smart detection technology.

5 BARCODE SCANNING OF MEDICINES WITH NAO CAMERA

5.1 Using ALBarcodeReader API

ALBarcodeReader converts the input image into greyscale and processes it to decode barcodes. As soon as the code is read, an event is raised, which contains *CodeData* and corner *Position* as depicted in Fig 6. The first corner point (labelled "*0*") is opposite to the corner with no black square, and remaining corners are labelled anti-clockwise.

$$CodeData = [Data, Position] \tag{7}$$

$$Position = [\,[x_0\,, y_0\,], [x_1, y_1], [x_2, y_2], [x_3\,, y_3]\,] \tag{8}$$

Position is an array containing the position coordinates of all the four corner positions of the barcode. Limitations of using this API is that it works only with QR codes, and the image resolution should be very high for detection. A technique needs to be developed to address this issue, and it should work for all types of barcodes. This is accomplished in real time using vertical descent Sobel Scharr operator in OpenCV.

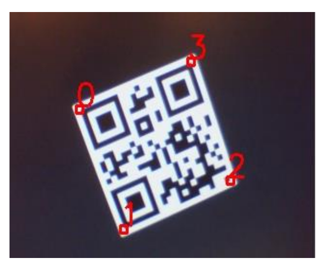


Figure 6: **Example of QR code and display of coordinates. Credits: Aldebaran Robotics.**

5.2 Proposed Algorithm

Fig 7 shows the Vertical descent Sobel Scharr operator algorithm used to perform Barcode scanning, which works for most types of barcodes in all development environments. Any image in the field of view of NAO is converted into greyscale and Scharr gradient is then calculated in x and y directions, which depends on Mean Square Angular Error (MSAE). Noise removal and binarisation is then performed on the gradient representation of the barcode image. Image blob analysis is then used and kernel is incorporated whenever the width exceeds the height. After successive erosions and dilations, contour detection is performed and then sorting is done to select the maximum area.

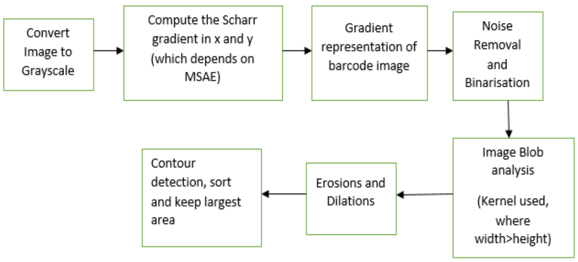
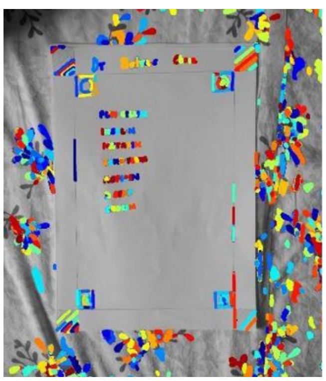
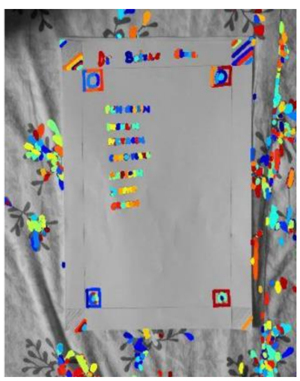


Figure 7: **Flowchart showing proposed Barcode scanning algorithm.**

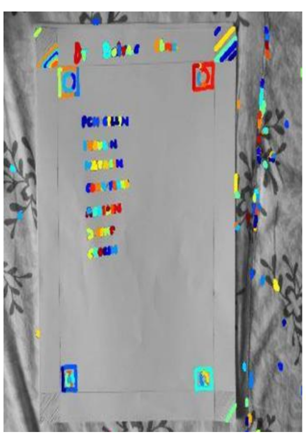
6 RESULTS AND DISCUSSION

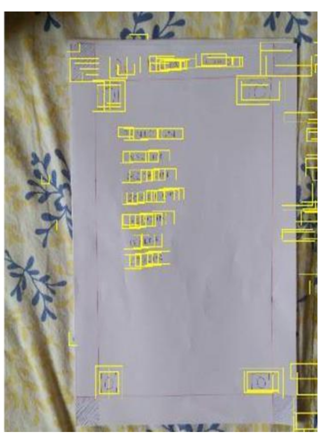
6.1 Text Identification in cluttered environment



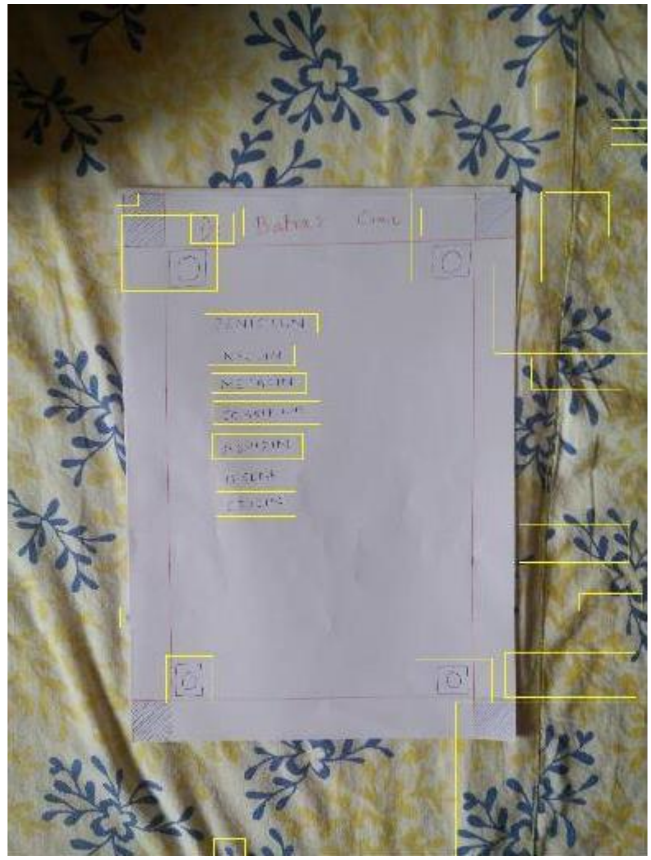


A **B**

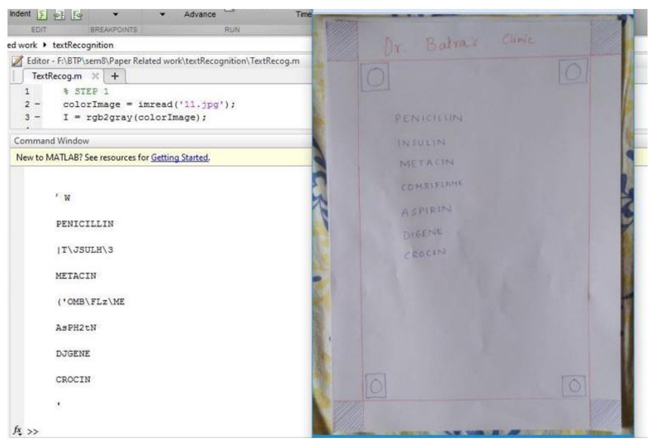




C **D**



E



F

Figure 8: **A] Detection of possible text regions using MSER. B] Removal of Non Text regions based on geometric properties. C] Removal of non-text regions based on SWT. D] Bounding Boxes around Text Regions E] Detected Text F] Output after OCR. [HIGH SCREEN RESOLUTION RECOMMENDED]**

Fig 8. shows text detection in cluttered environment. In A, the candidate text regions are detected in the higher contrast image using MSER In B, the non-text regions are removed based on basic geometric properties namely eccentricity and aspect ratio. In

C, the non-text regions are further removed using SWT and bounding boxes are used to merge text regions in D, followed by chain filtering in E, which is then fed to Tesseract to output the text regions in F.

6.2 Barcode Identification

Figure 9: **Proposed Barcode Recognition works fine for most barcodes.**

Proposed barcode recognition algorithm detects most barcodes, even when the image quality is not that high (Fig 9). However, it fails for embossed and raised barcodes as shown in Fig 10.

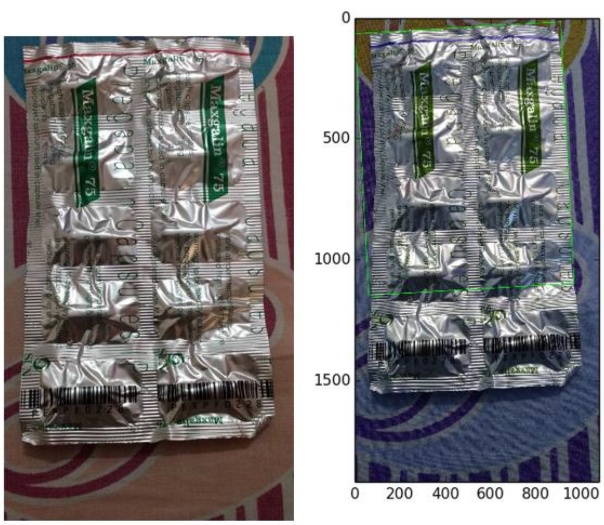


Figure 10: **Proposed Barcode Recognition fails for raised and embossed barcodes.**

7 CONCLUSIONS

In summary, socially assistive human robot interaction was explored using NAO robot, which was trained to do gripping, pick and place, text identification, and barcode scanning. Pharmacy Automation can be greatly enhanced using NAO for prescription identification. Robustness of text identification was improved using the high contrast image with histogram equalization and computational complexity was reduced using MSER algorithms in a combination with SWT. Barcode scanning was incorporated to cover the limitation of NAO detecting linear barcodes, which is the most frequent barcode used in pharmaceutical industries.

ACKNOWLEDGMENTS

This work was carried out at Mechatronics and Robotics V-Lab of the Department of Mechanical Engineering at IIT Guwahati, under the supervision of Prof. S.K Dwivedy.

REFERENCES

[1] N. Tejima. 2001. Rehabilitation robotics: a review. *Advanced Robotics*, 14, 7, 551-564.
[2] D. Floreano, A.J. Ijspeert, and S. Schaal. 2014. Robotics and neuroscience. *Current Biology*, 24, 18, R910-R920
[3] D. Feil-Seifer, and M.J. Matarić. 2011. Socially assistive robotics *IEEE Robotics & Automation Magazine,* 18, 1, 24-31
[4] JJ Cabibihan, H. Javed, M. Ang, and SM Aljunied. 2013. Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *International journal of social robotics*, 5, 4, 593-618
[5] A. Bruce, I. Nourbakhsh, and R. Simmons. 2002. The role of expressiveness and attention in human-robot interaction *IEEE International Conference on Robotics and Automation*, (ICRA'02) Vol. 4.
[6] D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maissonier. 2009. Mechatronic design of NAO humanoid *IEEE International Conference on Robotics and Automation,* (ICRA'09)
[7] S. Shamsuddin, H. Yussof, L. Ismail, FA Hanapieh, S. Mohammad, HA Piah, and NI Zahari. 2012. Initial response of autistic children in human-robot interaction therapy with humanoid robot NAO *IEEE 8th International Colloquium on Signal Processing and its Applications* (CSPA'12)
[8] KL Kim, K. Jung, and JH Kim. 2003. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 12, 1631-1639
[9] Y. Zheng, H. Li, and D. Doermann. 2004. Machine printed text and handwriting identification in noisy document images *IEEE transactions on pattern analysis and machine intelligence*, 26, 3, 337-353
[10] C.P Sumathi, T. Santhanam, and G. G Devi. 2012. A survey on various approaches of text extraction in images. *International Journal of Computer Science and Engineering Survey* 3.4 (2012): 27.
[11] H. Chen, SS Tsai, G. Schroth, DM Chen, R. Grzeszczuk, and B. Girod. 2011. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. *18th IEEE International Conference on Image Processing* (ICIP'11), 2609-2612
[12] J. Matas, O. Chum, M. Urban, and T. Pajdla. 2004. Robust wide-baseline stereo from maximally stable extremal regions *Image and vision computing*, 22, 10, 761-767.
[13] Y. Li, and H. Lu. 2012. Scene text detection via stroke width. *21st International Conference on Pattern Recognition* (ICPR'12)
[14] L. George, and A. Mazel. 2013. Humanoid robot indoor navigation based on 2D bar codes: Application to the NAO robot *13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*,329-335
[15] J. Figat, and W. Kasprzak. 2015 .NAO-mark vs QR-code Recognition by NAO Robot Vision *Progress in Automation, Robotics and Measuring Techniques. Springer International Publishing*, 55-64
[16] S.M Youssef, and RM. Salem. 2007 .Automated barcode recognition for smart identification and inspection automation *Expert Systems with Applications*, 33,4 , 968-977
[17] A. Abdiansah, and R.Wardoyo. 2015. Time Complexity Analysis of Support Vector Machines (SVM) in LibSVM *International Journal of Computer Application*, Vol 128
[18] S.R. Buss. 2004. Introduction to inverse kinematics with jacobian transpose, pseudoinverse and damped least squares methods *IEEE Journal of Robotics and Automation* , 1-19