A Novel Approach for Pick and Place Robot Control using Image Processing

Tivisha Goel
Electronics and Communication Engineering (B.Tech.)
Indira Gandhi Delhi Technical University for Women

Abstract - In this paper we portray a straight forward method for following a human hand based on pictures obtained by a dynamic stereo camera framework. We show the execution of this technique for pick and place robot as a major aspect of a multi-modular man-machine cooperation framework: recognizing the hand-position, the robot can translate a human pointing signal as the detail of an objective object to get a handle on.

Keywords - Image processing, man-machine interaction, robot control

I. INTRODUCTION

In the field of robotics, the interaction between man and machine typically consists of programming and maintaining the machine by the human operator. For safety reasons, a direct contact between the working robot and the human has to be prevented. As long as the robots act out preprogramed behaviours only, a direct interaction between man and machine is not necessary anyway. However, if the robot is to assist a human e.g. in a complex assembly task, it is necessary to have means of exchanging information about the current scenario between man and machine in real time. For this purpose, the classical computer devices like keyboard, mouse and monitor are not the best choice as they require an encoding and decoding of information: if, for instance, the human operator wants the robot to grasp an object, he would have to move the mouse pointer to an image of the object on a computer screen to specify it.

Along these lines of transmitting data to the machine isn't just unnatural yet in addition mistake inclined. In the event that the robot is furnished with a camera framework, it would be significantly more natural to simply point to the object to handle and let the robot recognize its position outwardly. Watching two humans in a similar circumstance uncovers another Interesting impact: by distinguishing the accomplice's look bearing the individual who focuses to an object can quickly control whether his expectation has been deciphered accurately. In the event that the accomplice takes a gander at the wrong object, this ends up clear promptly In a robot framework, this capacity can be actualized by giving the robot a stereo camera head that effectively locate the human hand position in work territory. To ensure a smooth collaboration amongst man and machine an assignment like this requires the visual handling. In the accompanying, we will portray a framework which satisfies these necessities (contrast and [1]). Before we go into insights about the vision handling techniques, we in the blink of an eye depict our pick and place robot on which we have actualized our strategy.

II. PICK AND PLACE ROBOT

Pick and place robot service robot system to assist a human partner in industrial assembly tasks. The research on this platform ranges from investigating the behavioral aspects of man-machine interaction and the representation and interpretation of scenes to the generation and organization of behavior. In our opinion successful man-machine interaction requires that both, the robot and the human partner, possess a similar sensor- and effectors equipment. Therefore, we designed Pick and Place robo: the six DOF manipulator is connected to a static trunk which is fixed on the edge of a table. Above the trunk as on head part we assembled a two stereo color camera system.

The objective of our examination is a robot framework that can play out the accompanying undertakings: a) visual recognizable proof of objects exhibited by a human educator, b) acknowledgment of these objects in a jumbled scene, c) getting a handle on and handing over objects, d) carrying on basic get together assignments. All practices ought to be acknowledged under visual, acoustical and haptic control by the human accomplice. This client association comprises of pointing signals, discourse commands and redresses of the controllers design by means of touch. In numerous applications particularly streamlined vision preparing techniques for the recognition of the human hand from camera images have been created.

One probability is to recognize the hand by methods for its development in the image. Under the supposition that whatever is left of the scene is static, this technique empowers an optimizing. Be that as it may, in the man-machine collaboration errand, the supposition of a static scene does not hold: got a handle on objects, the arm of the administrator and even the robot controller moves inside the scene. What's more, we need to have the capacity to identify a static pointing hand as well. A moment plausibility is to contrast the real image and components from a model database. By putting away models of hands in various postures, even hand motion acknowledgment can be accomplished by this technique. Be that as it may, to be adaptable and hearty, this approach requires either that an extensive number of various hand models are put away or that a complex visual pre handling decreases the fluctuation of the image to pick up a prototypic portrayal of the hand. For our Purpose, this model based technique isn't proper for the accompanying reasons: First, we don't have to perceive hand postures yet simply the general pointing course. Second, we can't bear the cost of a tedious pre-handling as the collaboration with the human occurs progressively. Third, we might want to perceive any human hand without the need a model of this hand in the database previously.

III. HAND SEGMENTATION AND DETECTION

Keeping all this in mind, we have decided to implement another very simple recognition method: we detect the hand on the basis of its typical skin color. In most situations, this parameter turned out to be sufficiently robust for our purpose. However, we have implemented a dynamical initialization phase in which the robot can "learn" the skin color of the specific operator (see Fig. 1). In the following, we will describe the recognition method in detail.

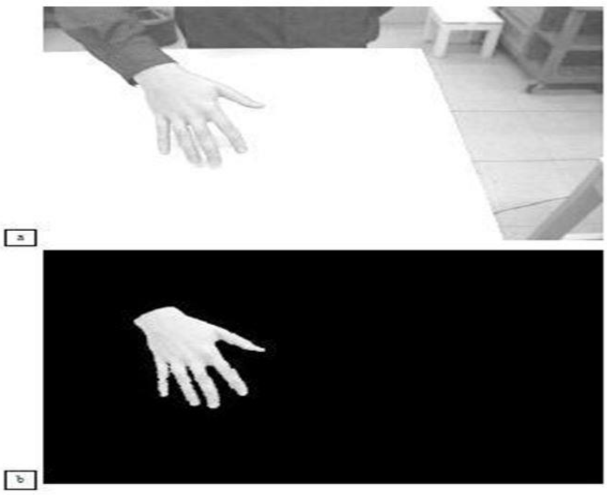


Figure 1: Figure (a) shows the human hand in a typical calibration phase. In figure (b) only that region got left, which represents the human skin color. These color samples are used for further operations, the calibration is complete.

A. Skin Segmentation – Our motivation is to section the quantized color image as indicated by skin color qualities utilizing the HSI (Hue, Saturation, Intensity) color demonstrate. Skin color patches have been utilized as a part of request to inexact the skin color subspace. Assessing a few skin color tests, covering an expansive scope of skin color appearances (changed skin colors, changed lighting conditions) one can watch that skin color tests frame a very minimized group in the HSI color space. Keeping in mind the end goal to find the client's hand, skin color regions must be sectioned and arranged. Their situation in space are computed by methods for a stereo-calculation from their divergence between the two camera images. Utilizing our insight about the limited workspace of CORA, we make the suspicion that one of the greatest skin colored bunches closest to the cameras must have a place with the client's hand. In this manner, the difference of a couple of purposes of this bunch is adequate to estimate the hand's mean separation.

B. Stereo based localization - Our test situation goes for a get together circumstance in which man and machine ought to understand a given assignment having a similar workspace is situated between the robot and the human. After the restriction of the skin color bunch having a place with the administrators hand by choosing the closest group to the cameras with an adequate size (see Fig. 3), we ascertain the difference of the pixel region at the base of this preselected hand group. This zone regularly speaks to the fingertip of the client for an ordinary pointing motion in this condition. The dispartity gives the 3-D position by a basic change. This confinement on such a little pixel territory empowers a quick assurance of the hand position.

C. Tracking - By restricting the time consuming calculation of the disparity on this single point, the described method for determining the hand localization is very fast. Therefore we can obtain a new hand position so frequently that a fast tracking of the hand is possible. As stated in the introduction, this behavior allows for a very natural interactive control of the detected hand position. During the tracking of the hand, human and robot become a coupled system, object can be rotated in the two dimensional plane, the correlation must be accomplished for a number of possible rotation angles.

IV. OBJECT RECOGNITION

The purpose of hand-position-recognition is the specification of an object in the workspace by pointing to it. The detection of the hand position serves as a pre-selection of the region of interest (ROI) in the image. In our case, the ROI is defined as a triangular area in the bird's eye view of the table which covers a sector with 30_ opening angle in front of the fingertip. Within this sector the object which is nearest to the fingertip is considered to be specified by the user's pointing gesture. The specification might be supported by a verbal command to avoid ambiguities. At this point the robot has simply to detect or, if there are further ambiguities, to recognize the specified object.

A. Learning phase - First, the robot must be enabled to recognize a specified object. Therefore, the robot must learn some of the object's special features. In our case these features are the color distribution and the calculated bird's eye view of the object. To obtain these features, all components of the image which are not part of the table are masked. This is done by a simple geometric calculation on the basis of the known spatial relation between the camera head and the table. In a second step, the color of the table is determined by assuming the table to represent the biggest cluster in the image. Image areas with similar colors are also masked as we assume them to belong to the table's surface. Thereby the scene gets restricted to its substantial part. On the basis of these limited data a color analysis and a determination of the disparity of each pixel follows. The color analysis provides the characteristic color values of the object within the HSI space. The disparity values provide enough data to calculate the three dimensional position of each pixel. These positions are transformed into a view of the scene from above. This bird's eye view is used as an object representation which is independent of the direction of view of the cameras and does not change its size, according to different distances.

B. Searching phase - In the searching phase the learned object characteristics enable a successful search for this object. Initially the scene gets reduced to its objects as done in the learning phase. At this time a search for the disparities could already take place, but due to the fact that this search is very time consuming, we implemented a data-reduction in a preprocessing step. This gets accomplished by using the color values, already known from the learning phase, to mask the extracted objects. Only areas with object clusters that are sufficiently well filled with the learned color values are used in the further image processing. The calculation of the bird's eye view on the basis of the disparity information is again the same as in the learning phase. At this time the bird's eye view of the learned object and the bird's eye view of the scene are present. In a process, similar to the Hausdorff Tracker described in [5], the bird's eye view gets mutually correlated with a blurred copy of the bird's eye views of the scene. The maximum of this correlation, verified by an inverse correlation, specifies the most probable position of the object in the scene.

V. GRASPING

As soon as the position of the object is known, its 3-d coordinates and orientation, calculated by a Hough transformation, are transmitted to the manipulator control to initiate the grasping process. The generation of the grasping trajectory is a complex problem the human arm. As only six degrees of freedom are needed to specify the position and orientation of the end-effectors, the additional joint angle introduces a redundancy into the kinematic problem.

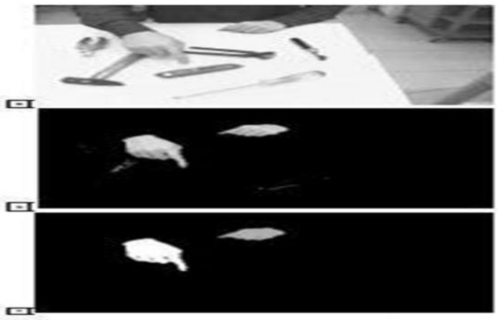
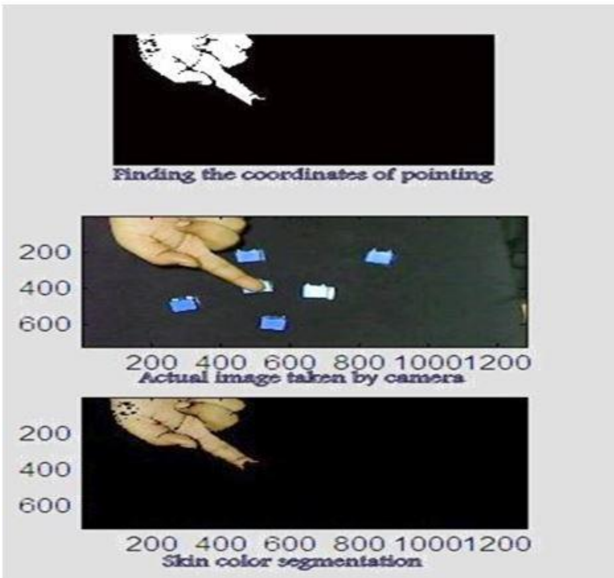


Figure 3: Figure (a) shows a typical situation during a man machine interaction. By using the color samples, acquired during the calibration phase, we get all regions, which possibly are skin areas (figure (b)). Finally we check the median disparity of adequate big clusters, to get the pointing hand, which should be the nearest cluster to the camera (figure (c)).

VI. INTEGRATION OF OTHER SOURCES OF INFORMATION

The major goal of our research is the design of a robot system which does not only have an anthropomorphic body structure but which is able to communicate with the operator on the basis of natural communication channels. Recognizing the human's hand position to identify a target object for grasping is only one of these channels. Within the current project we design several other means of man-machine interaction. We have, for instance, implemented a module for the recognition of the human's gaze direction.

This module can recognize the situation of the human's eyes with the goal that a harsh gauge of the focal point of consideration is conceivable. Thusly, the administrator can see the look course of the robot head with the goal that a characteristic method for recognizing the present district of enthusiasm on the table is conceivable. Moreover we have manufactured a discourse acknowledgment framework which can recognize talked catchphrase commands [7]. By methods for a discourse synthesizer, the robot can express its behavioral state. Based on regular commands it is conceivable to direct the robot's controller to a specific target or to end an erroneous conduct, (for example, the determination of a wrong object) by a talked command. Another correspondence channel depends on touch. We utilized a touch touchy fake skin by methods for which the administrator can adjust the stance of the robot arm. Moreover, unintended contact with deterrents can be evaded. The objective of our examination is the combination of all these distinctive correspondence directs in a total man-machine collaboration plot. Inside this plan, the excess of the diverse channels is changed into exactness. For instance, the administrator can determine an object by at the same time indicating and watching it and getting out the object's name. In actuality, the mix of these diverse wellsprings of data will empower the robot to defeat blunders in single sensor channels.



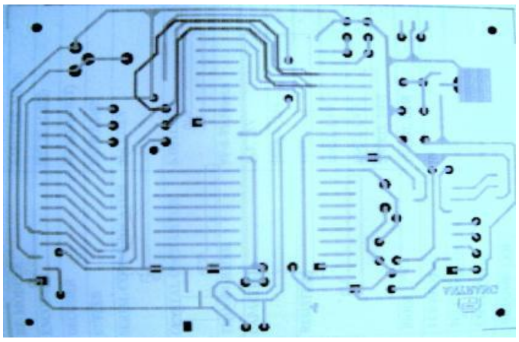
VII. RESULTS

The image processing method described so far has been used to track the operator's hand and to extract and interpret the pointing gesture. In this section we present our results by means of figure 4. In figure 4a the initial scene with three objects is shown and the operator's hand has been detected.

In b the moving hand is tracked and the pointing direction is extracted.

This pointing direction of the operator's hand serves as a pre-selection of the region of interest in the image. In c the highlighted object is detected as the only possible selection. For the case that two or more objects are possible candidates, the robot can react by asking for more information or simply grasp that object which is the only one known to him. The used robot arm for this implementation which is shown in fig 5.


a)


b)


c)


d)

Figure 5: the robot arm and in a) Robot arm, in b) interfacing circuit PCB layout, c) Web-cam as robot vision camera, d) Command window scenario

VIII. CONCLUSION

We have presented a basic concept for a vision based interactive robot control as a part of multi-modal man-machine interaction system. Implementing a straight forward technique for tracking a human hand and extracting a poiting gesture we could demonstrate how man and machine become a coupled system in a very natural way. The object recognition system can use those pointing gestures to minimize the searching area or, in the common case, can search the whole scene for a special object. Thereby the object's position and orientation is extracted very robustly. By using multiple channels of information, the whole system is able to overcome ambiguities and the robot can react in an adequate matter. For the future we plan to extend the multi-modal manmachine interaction system by integrating recognition of the human's direction of view.

IX. REFERENCES

[1]. T. Bergener and P. Dahm. A framework for dynamic man-machine interaction implemented on an autonomous mobile robot. In Proceedings of the IEEE International Symposium on Industrial Electronics, ISIE'97, 1997.

[2]. C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen. Walking Pedestrian Recognition. In ITSC, pages 292–297, Tokyo, Japan, 1999. IEEE.

[3]. Percy Dahm. Beiträge zu Anthropomorphen Robotorarmen (Ph.D. thesis). Ruhr Universität Bochum, Institut für Neuroinformatik, Lehrstuhl für Theoretische Biologie, 1999.

[4]. U. Handmann, T. Kalinke, C. Tzomakas, M. Werner, and W. von Seelen. An Image Processing System for Driver Assistance. Image and Vision Computing (Elsevier), 18(5):367 – 376, 2000.

[5]. D.P. Huttenlocher, G.A. Klanderman, and W.J.Rucklidge. Comparing Images Using the Hausdorff Distance. IEEE Trans. On PAMI, PAMI 15-9:850–863, 1993.