Robust Machine Vision Framework for Localization of Unknown Objects

S.M. Grigorescu and A. Gräser
Institute of Automation, University of Bremen, Bremen, Germany

***Abstract*— In this paper an approach to the detection of unknown objects is presented. The proposed algorithm is applied to the rehabilitation robot FRIEND II for the localization of objects situated in complex scenes. Also, the method was designed to cope with changes in the illumination conditions. The approach used in this work is the inclusion of feedback control in the image processing chain used by the *Machine Vision Framework* of the robot. A closed-loop control system was designed at image segmentation level for improving the robustness and reliability of the feature extraction module. The design of the closed-loop is based on an *Extremum Searching Algorithm* which searches for the optimal parameters of the image segmentation method. The performance of the proposed framework is investigated in comparison with a traditional open-loop method.**

I. Introduction

One key component of a robotic system that works in complex scenes is the robust perception of its environment. The problem in recognizing such scenes is that in order to detect the objects of interest a large amount of *a priori* information is required. Also, the vision algorithms behind this process are used in the traditional open-loop image processing chain (e.g. image acquisition, preprocessing, segmentation, feature extraction and object recognition) without any feedback information between them, thus making the 3D object localization unreliable with respect to external influences like more objects in the scene or changes in the illumination conditions. In such an open-loop image processing system the tuning parameters of the vision algorithms are constant. The reason behind this approach is that the input image is considered to be of good quality. However, in real world applications, where the system has to deal with numerous external influences, the overall result of the image processing chain is of poor quality.

In classical image processing applications *a priori* information about the objects of interest in the scene is used. In [1] a framework for tracking a hand using a predefined model is proposed. Also, in [2] color information was used for object recognition. An improvement of color based object recognition is discussed in [3] where an adaptive image segmentation method for coping with changing illumination conditions is presented. In comparison to the mentioned developments, only a few publications deal with the localization of objects without *a priori* data. An approach to detecting unknown objects in domestic settings is presented in [4].

In this paper a novel image segmentation algorithm for object detection without any *a priori* information about their characteristics is proposed and applied to the rehabilitation robot FRIEND II (*Functional Robot arm with frIENdly interface for Disabled people*) [5]. FRIEND II is a semi-autonomous robotic system designed to support disabled people in their daily life activities. The facilities of the designed robotic system are materialized in three working scenarios: an *All Day Living* (ADL) scenario where the user is helped in common tasks like preparing a meal, a *Library* scenario where the user is supposed to work at a books handling desk and finally a *Workshop* scenario. Because of the complex scenes in the scenarios, namely a large amount of objects placed in an unstructured background, a robust approach for developing the vision algorithms had to be taken. From the image processing point of view, the available objects were classified in two categories: container objects (e.g. fridge, microwave oven, etc.) and objects to be manipulated (e.g. bottles, glasses, meal-trays, books, etc.). In this paper only the detection of objects to be manipulated will be discussed, the container objects being detected with the help of a different method. The objects to be manipulated are considered to be unknown since they can appear in different forms and colors.

In [6], feedback control in image processing is used for the improvement of robustness of object recognition for objects to be manipulated in the system FRIEND II. The motivation behind the inclusion of closed-loop control in image processing is given in [7]. Within a feedback control system for image processing the definition of a *controlled variable* is that of a measure of image quality that can be used to control the values of the *actuator*

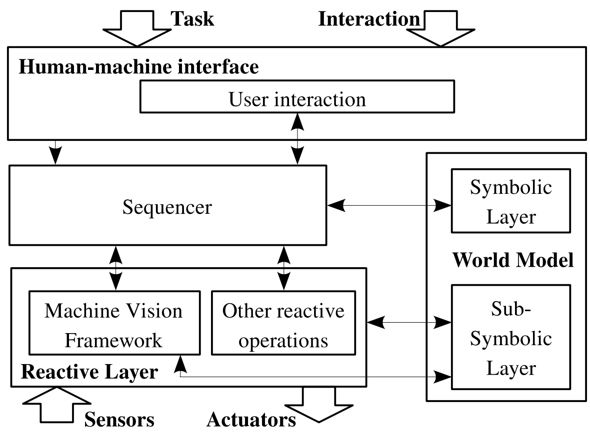

Fig. 1. The place of the Machine Vision Framework in the MASSiVE architecture.

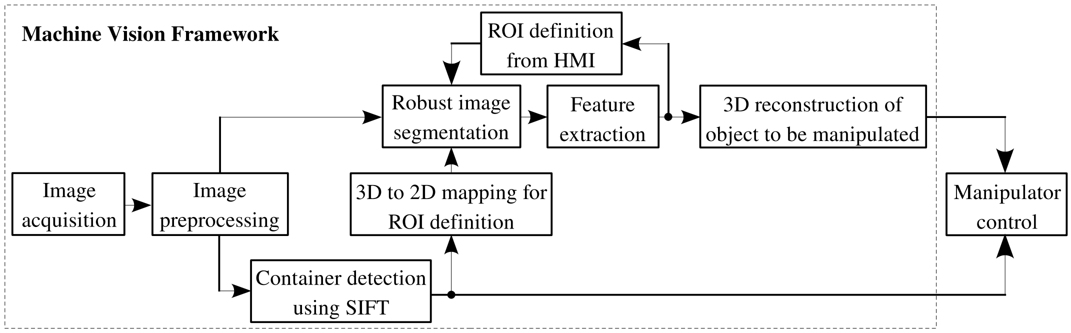

Fig. 2. Image processing flow of operations for recognizing a complex scene.

variable, namely the parameters of the image processing algorithms.

The paper is organized as follows. The architecture of the FRIEND II system along with the image processing flow is described in Section II. In Section III the choice of actuator and controlled variables and the design of the closed-loop system for the detection of unknown objects is presented. The performance evaluation of the proposed approach is given in Section IV and finally, in Section V, conclusions and outlook.

II. The FRIEND II System Architecture

A. Overall Control Architecture

The control of a complex robotic system like FRIEND II can only be achieved using an appropriate control framework. The used architecture, entitled MASSiVE (*MultiLayer Architecture for SemiAutonomous Service Robots with Verified Task Execution*), represents a distributed robotic control system which combines reactive behavior with classical artificial intelligence based task planning capabilities [8]. The MASSiVE architecture is represented in Fig. 1. The structure is divided in four specific modules. The *Human-Machine Interface* (HMI) operates at user interaction level. The user commands are acquired with the help of different input methods like Brain-Computer Interface (BCI), speech recognition or kin control and translated further in machine language for interpretation [9]. The processing algorithms that converts a user request into robot actions resides in the *Reactive Layer*. Here the data collected from different *Sensors* (e.g. stereo cameras, tactile tray, etc.) is processed for environment understanding and converted further into action by the available *Actuators* (e.g. 7 degrees-of-freedom manipulator). The sequence of operations needed to perform a specific task is computed by the *Sequencer* module. The *Sequencer* plays the role of a *Discrete Event Controller* (DEC) that plans sequences of operations by means of predefined task knowledge [8]. Through the functioning of the system the computed data is shared between the modules with the help of the *World Model*. The *World Model* defines the information produced and consumed by the operations in the *Reactive Layer*. The software interconnection between the processing layers is implemented using CORBA (*Common Object Request Broker Architecture*).

B. The FRIEND II Machine Vision Framework

To reliably implement the algorithms of the vision system an appropriate *Machine Vision Framework*, that encapsulates all the necessary image processing operations and their interconnections, had to be developed.

The vision hardware consists of a Bumblebee® Stereo Camera system connected to a computer dedicated to running the image processing algorithms. The stereo camera system is mounted on a frame-rack behind the user, above its head, and views the scene in front of the robotic system including the manipulator and the tray which is mounted on the wheelchair in front of the user. The acquired Bumblebee Camera images are represented in the RGB (*Red*, *Green*, *Blue*) color model.

C. Image Processing Flow for Complex Scene Recognition

In order to recognize a complex scene, like the environment of the FRIEND II robotic system, the image processing operations were interconnected between them and also to external modules (e.g. HMI). In Fig. 2 the image processing flow for the case of the detection of a bottle in the fridge is represented. The overall purpose of the vision system is to provide an augmented reality environment to the manipulator module for path planning and object manipulation. As discussed in introduction, after image preprocessing, the detection of objects has two directions: the recognition of container objects and the recognition of the objects to be manipulated.

Having in mind that the container objects will not change during the scenarios, a marker based approach detection can be used. The object recognition algorithm implemented for this task is a feature based detection method named SIFT (*Scale Invariant Feature Transform*) [10]. The method uses a model image to off-line train a classifier. During on-line operation the SIFT algorithm searches for the model image in the scene. When the model image is detected its pose (position and orientation in 3D space) can be reconstructed. Knowing the position of the model image on the fridge, the fridge container can be further reconstructed in augmented reality.

After SIFT recognition, the objects to be manipulated have to be detected. In contrast to the method used to locate container objects, the detection of objects to be manipulated has to be done without any *a priori* knowledge about their characteristics (e.g. shape, color etc.). For this type of object detection a robust image segmentation method has been developed. The method will be applied to a *Region of Interest* (ROI) in the input image, as will be further described. In this context, the ROI is a rectangular are which bounds either the container object, either the object to be manipulated.

The flow of image processing operations is applied in the following order. First, the ROI is defined by 3D to 2D mapping from the already detected container object (e.g. fridge) and the robust image segmentation algorithm is applied. If the feature extraction module detects more objects in the fridge than the user of the robotic system will be involved in the scenario and asked to chose one of the detected objects. Using one of the available HMI input methods, the user will select the desired object and the scenario can go further with the pose estimation of the object to be manipulated.

III. Closed-loop Control for Image Segmentation

The inclusion of feedback control in image processing differs significantly from the typical industrial control applications. One of the biggest challenges faced here is the absence of a mathematical model needed for the design of the closed-loop controller. In the next subsections the choice of the actuator variable - controlled variable and the design of the closed-loop image segmentation system will be presented.

A. Choice of Actuator-Controlled Variables

For segmenting the objects to be manipulated a thresholding operation on a gray level image was implemented. Because the objects are represented in different colors, the color plane of the HSI (*Hue*, *Saturation*, *Intensity*) color model is used. The acquired RGB images are converted by the image preprocessing module into HSI images.

The thresholding operation on an input gray level image $f(x, y)$ outputs a binary image $t(x, y)$ (x and y representing pixels coordinates). In $t(x, y)$ the pixels from $f(x, y)$ that reside in the segmentation interval $[T_{min}, T_{max}]$ are thresholded to foreground pixels (black pixels 1) and the pixels that are outside the threshold boundaries $[T_{min}, T_{max}]$ are set to background pixels (white pixels 0). The definition of the thresholding operation is:

$$t(x, y) = \begin{cases} 1, & \text{if } f(x, y) \in [T_{min}, T_{max}], \\ 0, & \text{if } f(x, y) \notin [T_{min}, T_{max}]. \end{cases} \quad (1)$$

As it can be seen from Fig. 3, only one segmentation interval can produce a good binary output image. Because of this reason a proper choice for the actuator variable are the boundaries of the segmentation interval.

For automatically adjusting the actuator variable to the optimal value in a closed-loop manner a controlled signal has to be defined. Having in mind that the controlled signal will be calculated in the ROI of the binary image, a proper definition for it would be that it has to give a measure of connectivity between the foreground pixels [7]. For this purpose the following ROI quality measure is proposed:

$$I = -log_2 p_8, \; I(0) = 0, \quad (2)$$

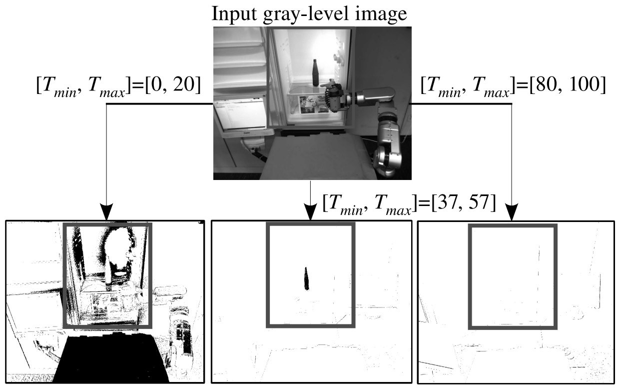

Fig. 3. Segmentation results obtained for different threshold intervals.

where p_8 represents the probability of a foreground pixel to be surrounded with 8 foreground pixels in its 8-pixel neighborhood:

$$p_8 = \frac{\text{nr. of foreground px. surrounded with 8 foreground px.}}{\text{total nr. of px. in the ROI}}. \quad (3)$$

The higher the uncertainty measure I is, the larger the disorder in the segmented ROI is.

In order to investigate the input-output controllability when considering variable thresholding interval as the input and the measure I as the output, the thresholding of the ROI on an input sample image was done. Having in mind that a SISO (*Single Input Single Output*) feedback control system will be developed, the interval between the low T_{min} and the high T_{max} threshold boundaries will be kept constant, thus obtaining only one variable to control. The initial value for $[T_{min}, T_{max}]$ was set to $[0, 20]$. To those initial boundaries an increment $u = i$ was added as $[T_{min} + i, T_{max} + i]$. For each segmented image corresponding to the increment $i \in [0, 159]$ the uncertainty measure I of the ROI was calculated. Because of the definition of the hue plane image in our implementation, namely containing pixels with values ranging from 0 to 179, the maximum threshold increment i can have a maximum value of 159. The resulting characteristic for the case of an object to be manipulated (e.g. bottle) placed inside a container (e.g. fridge) is presented in Fig. 4. From the analysis of more input-output characteristics similar to the one presented in Fig. 4 we have discovered that the optimal threshold increment u_{opt} for a specific object is found on one of the local minimums of the uncertainty measure I. The presentation of the automatic procedure for choosing the apropriate local minimum is out of the scope of this paper.

B. Closed-loop Control Design

The purpose of the closed-loop system is to automatically drive the threshold increment i to the optimal value for segmentation. Due to the nature of the reference value in the proposed feedback system, namely a local minimum on the input-output characteristic, a control action based on *Extremum Searching Algorithm* (e.g. hill-climbing) is suggested. In this type of control system the control action u_{opt} is realized through an optimization process [7]. The closed-loop control diagram is presented in Fig. 5.

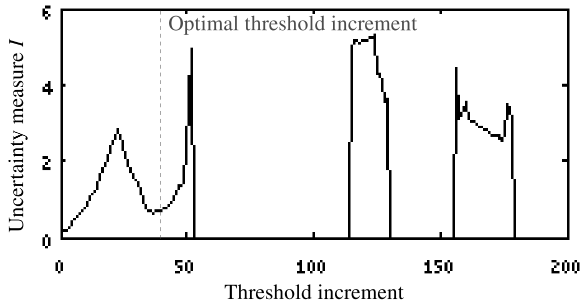

Fig. 4. The uncertainty measure I of segmented pixels vs. threshold increment.

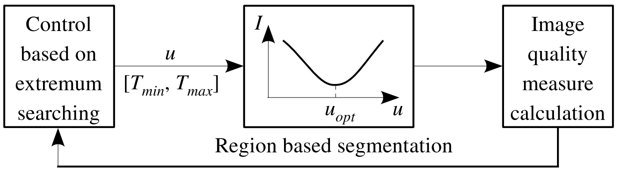

Fig. 5. Block diagram of the proposed closed-loop system.

IV. Performance Evaluation

For the evaluation of the proposed algorithm a set of images containing objects of different shape, color and size were used. In Fig. 6(a,d,g) three examples of such objects are given, along with the calculated ROI. The robust color segmentation method was applied, with the corresponding results shown in Fig. 6(b,e,h).

In order to evaluate the performance of the proposed algorithm we have choose to compare it with a traditional open-loop segmentation method which uses the mean gray level value of the image ROI to calculate the thresholding interval $[T_{min}, T_{max}]$. The output of the open-loop method is represented in Fig. 6(c,f,i). From investigating the output images it can be observed that, in comparison to the open-loop method, the closed-loop segmentation provides a result that satisfies our criterion, which is a full, well connected, shape representing the object to be manipulated.

V. Conclusions and Outlook

This paper considers the benefit of the use of control techniques in image processing for detecting unknown objects in complex scenes. The usefulness of the closed-loop image segmentation is demonstrated through the results on object recognition in the robotic system FRIEND II. The results were compared with a traditional open-loop segmentation algorithm. As future work, the designed framework can be modified for detecting unknown objects using a boundary segmentation method instead of a region based one. Also, the inclusion of more feedback loops in the system will be investigated.

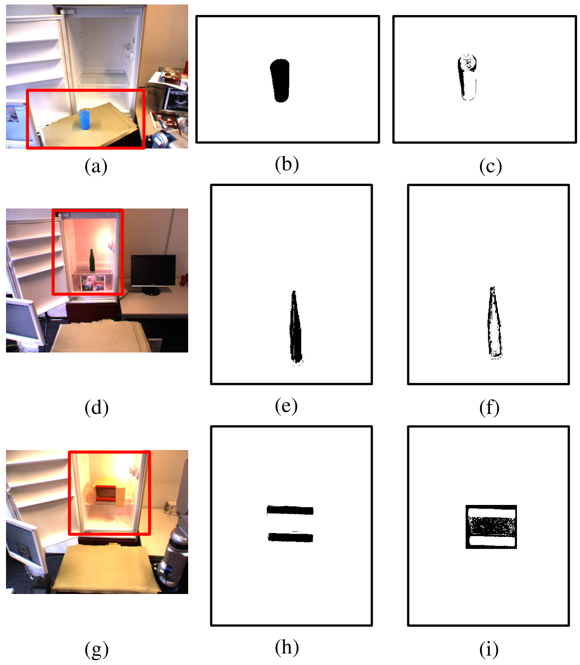

Fig. 6. Test images (a,d,g) along with the closed-loop (b,e,h) and open-loop (c,f,i) segmentation results.

References

[1] J.B. de la Rivière and P. Guitton, "Image-based analysis for model-based tracking," in *Proc. Mirage*, March 2007.

[2] P. Chang and J. Krumm, "Object recognition with color concurence histogram," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Fort Collins, pp. 498–504, 1999.

[3] L. Iocchi, "Robust color segmentation through adaptive color distribution transformation," in *Proc. RoboCup Symposion*, Bremen, 2006.

[4] D. Kragic, M. Bjorkman, H.I. Christensen, J.O. Eklundh, "Vision for robust object manipulation in domestic settings," *Robotics and Autonomous Systems*, vol. 52, pp. 85–100, July 2005.

[5] I. Volosyak, O. Ivlev and A. Gräser, "Rehabilitation robot FRIEND II - the general concept and current implementation," in *Proc. 9th ICORR*, pp. 540–544, Chicago, 2005.

[6] S.K. Vuppala, S.M. Grigorescu, D. Ristić and A. Gräser, "Robust Color Object Recognition for a Service Robotic Task in the System FRIEND II," in *Proc. 10th ICORR*, pp. 704–713, Netherlands, 2007.

[7] D. Ristić, Feedback structures in image processing. *PhD. thesis*, 2007.

[8] C. Martens, O. Prenzel, J. Feuser and A. Gräser, "MASSiVE: MultiLayer Architecture for SemiAutonomous Service Robots with Verified Task Execution," in *Proc. 10th Int. Conf. Optimization of Electrical and Electronic Equipments*, pp. 107–112, Brasov, 2006.

[9] O. Prenzel, C. Martens, M. Cyriacks, C. Wang and A. Gräser, "System controlled user interaction within the service robotic control arhitecture MASSiVE," *Robotica, Special Issue*, 2007.

[10] D.G. Lowe, "Object Recognition from Local Scale-Invariant Features", in *Proc. of the Int. Conf. on Computer Vision, Corfu*, September 1999.