# Robust Color Object Recognition for a Service Robotic Task in the System FRIEND II

Sai K. Vuppala, Sorin M. Grigorescu, Danijela Ristić, and Axel Gräser

*Abstract* — **One of the key requirements of service and rehabilitation robotic systems is the robust perception of the environment. This paper presents the Machine Vision Framework within the rehabilitation robotic system FRIEND II and a method for improving the visual perceptual capability of this system. For the reliable, autonomous performing of the "beverage serving" task robust visual information about the objects to be manipulated in the presence of variable lighting conditions is necessary. A novel aspect of the proposed color object recognition method is its use of feedback control at the image segmentation level, which makes it able to cope with the object color uncertainty caused by different illumination conditions. The idea behind the closed-loop image segmentation is to provide the object recognition step with input data on which it can rely. The proposed object recognition method uses Hu moments which are invariant to rotation, translation and scaling of the object in the image. The benefit of the closed-loop color object recognition in service robotics is demonstrated through the example of the recognition of the green bottle. The presented experimental results on the performance evaluation show that the proposed method has robustness with respect to illumination as well as with respect to the different localization of the object to be manipulated.**

## I. Introduction

THE system FRIEND II (*Functional Robot arm with frIENdly interface for Disabled people*) is a semi-autonomous service robotic system designed to support disabled people in their daily life activities [1][2]. The system consists of a 7 degrees of freedom manipulator mounted on an electrical wheelchair. The system is equipped with various sensors that provide intelligent perception of the environment needed for task execution support. One of those sensors is a stereo camera system which provides visual information regarding the system's environment.

The "beverage serving" scenario of the system FRIEND II is an example of a typical service robotic task. After the user asks for a beverage, the robot has to pour it into the glass from the bottle and to serve it to the user. Reliable autonomous execution of such task relies on the exact 3D localization of objects to be manipulated. During the process of 3D localization based on the visual perceptual capability of the robotic system, the objects of interest first have to be reliably recognized in the camera image. Feature points obtained from the recognized object region of interest, together with the known additional information about stereo camera geometry, are used for the 3D object localization. Hence, object recognition is one of the most important tasks to be performed by the robot's vision system. Since FRIEND II is intended to support the user in daily life activities, the recognition task must be robust enough to work effectively in different lighting conditions that arise during the day.

For the purpose of object localization in robotics, objects which have good texture, or pattern, are typically used [3]. A number of algorithms for object recognition based on pattern information have been published [4][5]. However, those object recognition methods can give unreliable results for objects which do not have any pattern. In the case of untextured objects, the object color can be used as a useful object characteristic. Because of the distinguishing nature of object's color, the study of object recognition based on color information has been increased in recent years [6][7]. The use of color information in service robotic vision systems brings the challenge of dealing with uncertainty in the apparent colors of the objects to be manipulated. Color uncertainty arises from changes in the illumination conditions during image acquisition. The different lighting conditions range from the pure sunlight of the day to the purely artificial light available during the evening hours and includes various mixtures of the two. Therefore, color based object segmentation and the subsequent recognition process can become unreliable. Numerous approaches have been published to solve the problem of color uncertainty. An adaptive color object segmentation method for changing lighting conditions which avoids tedious and unreliable manual color calibration by using the color table which is updated periodically during the system operating time is presented in [8]. A comprehensive color image normalization approach which brings stable color values for the objects with Lambertian surfaces is presented in [9].

Manuscript received February 11, 2007.

The authors are with the Institute of Automation, University of Bremen, Otto-Hahn-Allee NW1, 28359 Bremen, Germany (phone: +49-421-218-4846 / 7523; fax: +49-421-218-4596 / 4707; e-mail: {vuppala, grigorescu, ristic, ag} @iat.uni-bremen.de ).

The authors claim that normalized color values are invariant to changing positions of light source as well to the illuminant color. In contrast to these publications, in this paper robustness with respect to illumination is achieved by inclusion of feedback control loops at the segmentation level of the object recognition chain. The idea of the inclusion of feedback structures at different levels of image processing to improve its robustness is introduced in [10]. The main idea behind this is that the control loop drives the current processed image to the image of the desired quality. This idea is further developed in [11] and [12] where the benefit of the feedback control loops for improvement of gray level image processing was demonstrated through the presentation of the results achieved in application of character recognition on metallic surfaces. In those papers different control structures were considered. The common characteristic of the implemented closed-loops was that the control actions were realized through optimization processes. Due to the absence of available ground truth that is due to the absence of the reference image, the controlled variables were defined as the measures of the image quality whose extreme values, minimum or maximum, correspond to the images of good quality. Therefore, optimal values of controlled variables were determined by appropriate search algorithms. In contrast to that, in this paper the case of available ground truth reference image, which allows use of proven error based control techniques, is considered. Also, the novelty in this paper, compared to the papers [11] and [12] is that the inclusion of feedback control to improve robustness of the color rather than gray level image processing is considered. The goal is to cope with the color variations in the hue plane of the HSI (*hue*, *saturation*, *intensity*) [13] color space. The reason for the use of the HSI color space is to perform object segmentation in the hue plane. Namely, in contrast to, for example, RGB (*red*, *green*, *blue*) color space, the pure color information is stored exclusively in one channel, which is the hue plane.

Besides the illumination conditions during the image acquisition, as stated in [14], the appearance of an object in the image depends upon the position of the object with respect to camera as well. Therefore, besides the color uncertainty, object recognition in computer vision systems basically suffers from the problems caused by variations due to changes in object pose with respect to camera and /or camera view point with respect to the object. In this paper, camera view with respect to object is considered as constant. The problem due to the uncertainty caused by the object pose variation is overcome using invariant Hu moments. Image invariant moments have been widely used in computer vision for a long time. Hu in 1962 proposed a set of seven image invariant moments [15]. Those moments are invariant to the translation and scaling of image pattern. Out of these, first six are also rotational invariant and the seventh one is skew invariant. In the same paper, the author provided the test results of recognition of alphanumeric characters. The general approach to object recognition using Invariant Moments is presented in [16]. The image analysis based on Moment Invariants for the purpose of object recognition is explained in [17]. In all these publications it is only the benefit of the Hu moments that is considered and the quality of the input image data is taken for granted. However, as shown in the following, the full use of the benefit of Hu moments is possible only when a good quality input image is available. As said above, the proposed inclusion of the closed-loop control at the segmentation level compensates for problems caused by variable illumination. Implicitly, it compensates the problem of poor binary input image to the Hu moments based recognition. In summary, the proposed image segmentation method provides reliable binary input image to the Hu moments based object recognition independently of the illumination condition and the pose of the imaged object.

The paper is organized as follows. The Machine Vision Framework that has been developed in system Friend II is presented in Section II. The choice of the operations at the image segmentation level, guided with the idea of achievement of the invariant Hu moments, is explained in Section III. Section IV presents the two closed-loops introduced at the segmentation level. The experimental results on the performance evaluation of the developed recognition method are given in Section V.

## II. THE MACHINE VISION FRAMEWORK IN THE FRIEND II SYSTEM

The robust control of a complex robotic platform like FRIEND II can only be achieved with an appropriate control architecture which separates the different levels of processing (e.g. image processing, robotic arm control, task planning etc.) into abstraction layers linked together by a system core module which acts as a system manager. The software architecture used for the control of the rehabilitation robotic platform is entitled MASSiVE (MultiLayer Architecture for SemiAutonomous Service-Robots with Verified Task Execution) and it represents a distributed control architecture which combines reactive behaviour with classical artificial intelligence based task planning capabilities [18]. Within MASSiVE, the CORBA (Common Object Request Broker Architecture) [19] framework is used for the communication between different layers of processing tasks.

The MASSiVE architecture interacts with the user through the Human-Machine interface interconnected with the Sequencer which is the core of the MASSiVE architecture. The user can specify a desired task to be performed by the robot through several interfaces like speech recognition or Brain-Computer-Interface [21]. During the system operation the task parameters can be viewed with the help of a Graphical User Interface available on a display system mounted on the wheelchair in front of the user.The Machine Vision Framework module acts on the commands sent by the Sequencer and performs appropriate tasks needed for the reliable object recognition and subsequent 3D localization of the object to be manipulated. The Sequencer plays the role of a Discrete Event Controller (DEC) that plans sequences of operations by means of predefined tasks knowledge [18].

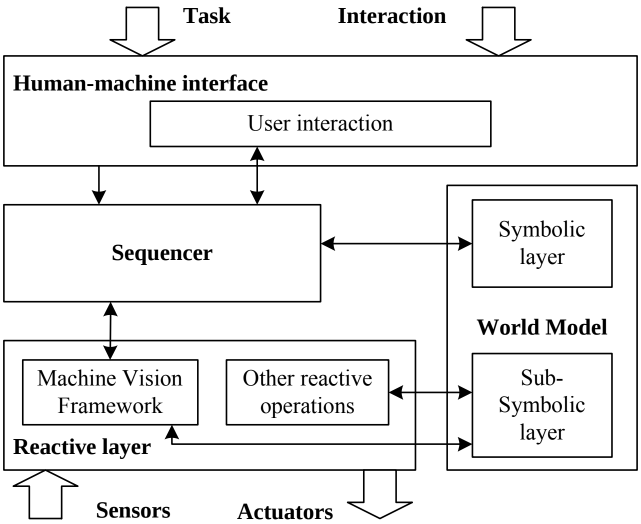


Fig. 1. The Machine Vision Framework place inside the MASSiVE architecture.

As it can be seen in the simplified diagram of the MASSiVE structure in Fig. 1, the vision framework is placed inside the Reactive Layer for providing vision information to the Sequencer component, which will act further with appropriate commands to the gripper element. The vision framework is communicating with the Sub-Symbolic layer of the World Model. Within this context the World Model defines the information produced and consumed by the operations in the Reactive Layer.

Vision is organized around a kernel element which controls a dataset of image processing algorithms. The kernel uses the available image processing methods through the algorithms FIFO (First In First Out) processing stack. Namely, one algorithm will take as input the output of the previous algorithm in the stack. In order to interconnect the inputs and outputs of the vision algorithms a standard base class was created for deriving all the implemented image processing classes from it.

For the robust color object recognition method the algorithms processing stack is composed of two image processing entities: the robust color segmentation algorithm, which provides the improved binary image, and the Hu moments based object recognition which classifies the objects in the improved binary segmented image. As it can be seen in Fig. 2, the sequential steps considered for object recognition are the conversion of the input RGB image to the HSI color image, robust image segmentation and finally Hu moments based object recognition. The extracted object features are further feed into the 3D localization part of the framework.

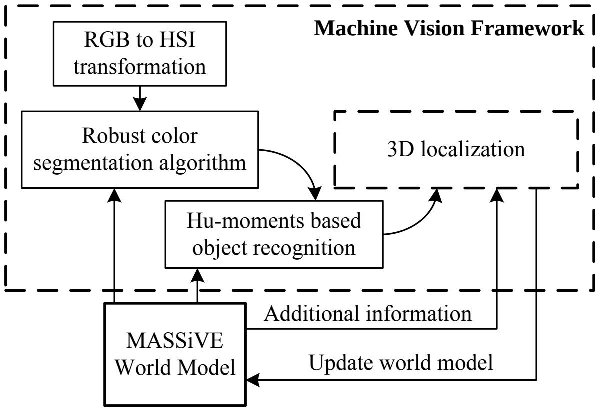


Fig. 2. Machine vision processing chain.

The reference values of parameters of the object segmentation and recognition algorithms are loaded from the MASSiVE World Model according to the current object of interest. After the objects in the image are detected and classified, the information about them must be wrapped into a predefined data container which will be passed further to consequent higher level machine vision operations. The data container holds vision information like the coordinates of feature points, color, image invariant moments, etc. In the application considered in this paper, the vision information, used for subsequent 3D object localization, consists of the height and the middle point of the bounding box of the recognized object.

The vision framework is implemented on a Pentium IV computer with 512MB of RAM, running a Debian Linux operating system, which is interconnected with the other processing modules via the CORBA interface. The used stereo camera system consists of two CCD Sony EVI-D70 color video cameras [2]. The camera images are transferred via the RS-232 serial port using the Sony VISCA (VIdeo System Control Architecture) protocol. The stereo camera system is mounted on the frame-rack behind the user above its head. The stereo cameras view 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. All camera parameters are used in default settings. The idea is to develop image processing software able to cope with the eventual problems, regarding the image quality and consequently image processing, that arose from the certain camera characteristics like default gain and offset settings or auto white balance.

## III. Choice of the Image Segmentation Steps in the Hu Moments Based Object Recognition

The input image to the Hu moments based object recognition is a binary segmented image. In the presented system, binary segmented image is obtained by thresholding the hue plane image. Since the pixel values of this image represent the color values, pixels are segmented either to black background or white foreground based on color information.

Let us denote the representation of colors from the hue plane as a color space $C$. Color classification can be seen as partition of the entire color space $C$ into different color classes $C_l$ such that $\bigcup_{l=1}^{n} C_l = C$. In our application a color class $C_l$ is nothing but the interval of pixel values in the hue image histogram. Therefore, segmentation of an object of the color belonging to the class $C_l$ is done using a thresholding function $t(i)$ which can be expressed as:

$$t(i) = \begin{cases} 1, \text{if} \quad f(i) \in C_l \\ 0, \text{if} \quad f(i) \notin C_l \end{cases} \tag{1}$$

where $f(i)$ is the pixel value of the pixel $i$ in the hue image. For the sake of clarity an object color class $C_l$ in the following is referred to as an *object thresholding interval*.

In the presented application, appropriate reference thresholding intervals for the objects of interest in the FRIEND II environment are determined off-line from the so-called reference hue image. The reference hue image corresponds to the reference RGB image of the scene from the "beverage serving" scenario taken in artificial light. The thresholding interval of an object of interest is determined by manually thresholding the hue reference image. The thresholding interval is said to be appropriate for the object of interest if the resulting binary image contains as much object pixels as possible. Fig. 3 shows the object segmentation results for the green bottle imaged in different illumination conditions. Segmentation of the hue images was performed using the reference thresholding interval for the green bottle which is determined by segmenting the reference hue image as explained above. As it can be seen, the segmented binary image corresponding to the image of green bottle taken in artificial light conditions is of good quality since it contains the majority of object pixels. In contrast, the result of segmentation of the image captured in daylight conditions is quite poor. This is an expected result since different illumination conditions caused significant variance in the image colors. Because of that the used threshold interval, determined for the case of reference artificial light, can yield good segmentation result only for the images taken in similar illumination conditions. Therefore, in order to achieve reliable object segmentation it is necessary to adjust the object thresholding interval according to the changes in illumination. Another solution to the problem is to use the constant predefined object thresholding interval but to adjust the pixel values in hue image so that as much object pixels as possible are segmented. However, manual adjustment of either object thresholding interval or hue pixel values is quite time-consuming and meaningless during autonomous functioning of the robotic system. To overcome this problem the inclusion of closed-loop control in the image segmentation chain is suggested as explained in Section IV.

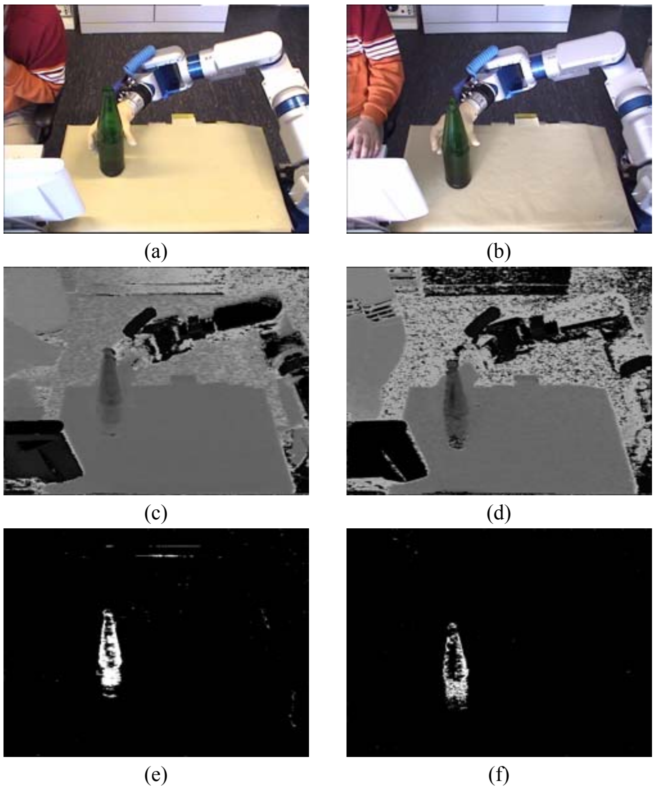



Fig. 3. Binary segmentation of the green bottle in Friend II environment corresponding to different lighting conditions. RGB image captured in artificial (a) and daylight illumination condition (b). (c) and (d) corresponding hue images. (e) and (f) segmented images obtained using reference thresholding interval.

A binary segmented image of good quality, which is appropriate for subsequent processing, is the image that contains full and well shaped segmented object regions. Bearing this in mind, from the above shown results it is obvious that the thresholding operation can not provide good segmentation result on its own. Namely, even in the case of good segmentation shown in Fig. 3(e), the segmented region of the bottle contains holes that have to be "filled". This result is expected since, as explained, the reference thresholding interval can yield as much segmented object pixels as possible but not all pixels. The reason is that due to the light reflection during the image acquisition object pixels can belong to different color classes.

In order to improve the quality of binary segmented image, the application of the morphological dilation [13] is suggested in this paper. The basic effect of the dilation operator on a binary image is to gradually enlarge areas of foreground pixels while suppressing holes within those regions. The dilation operator takes two inputs. One is the binary image to be dilated and the other is the so-called structuring element. The structuring element is nothing but a matrix consisting of 0's and 1's. The distribution of 1's determines the shape of the structuring element and the size of the pixel neighborhood that is considered during the image dilation. The structuring element is shifted over the image and at each pixel of the image its elements are compared with the set of the underlying pixels according to some predefined operator. As a result, basically, a black background pixel turns to white (object) foreground pixel if there are white pixels in its neighborhood that are covered by the 1's of the structuring element. The effect of "filling of segmented regions" by dilation strongly depends on shape and size of the structuring element as well as on the number of performed dilations. Manual tuning of the dilation parameters is time-consuming and rather meaningless for autonomous functioning of the robotic system. As described in Section IV, the second included closed-loop in the proposed object recognition system provides automatic performing of the image dilation.

In the considered approach, the choice of the above described steps of the image segmentation, thresholding followed by binary image dilation, is justified by the achievement of the invariant Hu moments as explained in the following.

Hu [15] presented a set of seven invariant moments which can be derived from the second and third order moments. In case of a digital image $I(x, y)$, the moment of order $(p+q)$ is denoted with $m_{pq}$ and is determined by:

$$m_{pq} = \sum_x \sum_y x^p y^q I(x, y) . \tag{2}$$

The central moments $\mu_{pq}$ are defined as:

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q I(x, y), \quad p, q = 0,1,2,... , \tag{3}$$

where $\bar{x} = m_{10} / m_{00}$ and $\bar{y} = m_{01} / m_{00}$ and $I(x, y)$ is the intensity level of an image point with coordinates $x$ and $y$. In case of binary image, $I(x, y)$ is 1 for each object (white) pixel and 0 for each background (black) pixel. In our approach, in order to evaluate the robustness of proposed object recognition system, first three Hu moments $I_i$, $i = 1,2,3$ of binary image region of desired object were considered. They are defined by the following formulas:

$$I_1 = \eta_{20} + \eta_{02} , \tag{4}$$

$$I_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}{}^2 , \tag{5}$$

$$I_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 , \tag{6}$$

where $\eta_{pq}$ is the normalized central moment of order $(p+q)$:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\frac{p+q}{2}+1}} . \tag{7}$$

In Table I, the first three Hu moments of segmented binary image regions of green bottle are shown for three different object positions. These binary images were obtained by segmenting the hue images corresponding to RGB images captured in artificial lighting conditions. The used thresholding interval was the reference one obtained by the manual segmentation of reference hue image as explained above.

TABLE I
HU MOMENTS FOR SEGMENTED BOTTLE REGIONS

| ObjectType | BOTTLE | | |
|---|---|---|---|
| Binary images |  |  |  |
| $I_1$ | 9.270e-002 | 7.964e-002 | 8.229e-002 |
| $I_2$ | 6.667e-003 | 4.812e-003 | 5.140e-003 |
| $I_3$ | 3.387e-005 | 5.217e-005 | 6.265e-005 |

Hu moments are considered to be invariant with respect to rotation, translation and scaling of the object in the image [15]. Because of that they are used in the Computer Vision community for the process of object classification [16]. However, as it can be seen from Table I, the calculated Hu moments vary significantly indicating that the segmented object regions are not of enough good quality for subsequent classification. In order to improve the quality of segmented regions, the above shown binary images are dilated using suitable dilation operator. The improved binary images together with the calculated Hu moments of the object regions are shown in Table II. As evident from the Table II, the first two Hu moments for the full and well shaped object regions are almost invariant. Though there is a slight variation in the third moment, it is considered in order to have larger number of object features for the better classification process. The effect of the third moment variations is handled through the predefined tolerance of the Euclidean distance:

$$d_r = \sqrt{(I_{r1} - I_1)^2 + (I_{r2} - I_2)^2 + (I_{r3} - I_3)^2} , \tag{8}$$

which is used as decision function for classification of an object into particular class. In (8) $I_{ri}$, $i = 1,2,3$ refer to the reference Hu moments of an object of interest which are manually defined from the reference binary image. If the distance $d_r$ for an unknown object is less than the predefined minimum tolerable distance $d_{\min}$ for particular object class, then the object is classified into that class.

TABLE II
HU MOMENTS FOR IMPROVED SEGMENTED BOTTLE REGIONS

| Object Type | BOTTLE | | |
|---|---|---|---|
| Binary images |  | | |
| $I_1$ | 9.468e-002 | 9.644e-002 | 9.638e-002 |
| $I_2$ | 7.221e-003 | 7.865e-003 | 7.467e-003 |
| $I_3$ | 1.384e-005 | 1.651e-005 | 2.751e-005 |

Hu moments of different objects differ significantly. This can be observed by comparing the considered Hu moments for the well segmented bottle regions with the Hu moments for the well segmented glass regions given in Table III. The comparison results justify the use of Hu moments as a classification criterion. However, as demonstrated above, the benefit of the Hu moments based object recognition can be fully used only if the objects of interest are correctly segmented from the background. Therefore, image segmentation is a crucial step in the object recognition chain and the achievement of reliable and robust object segmentation is of essential importance.

TABLE III
HU MOMENTS FOR IMPROVED SEGMENTED GLASS REGIONS

| Object Type | GLASS | | |
|---|---|---|---|
| Binary images |  | | |
| $I_1$ | 6.991e-002 | 6.745e-002 | 6.568e-002 |
| $I_2$ | 1.556e-003 | 1.482e-003 | 1.753e-003 |
| $I_3$ | 5.281e-006 | 5.502e-006 | 6.473e-006 |

The proposed inclusion of closed-loop control, as explained in the following section, aims to improve the robustness of recognition process with respect to external influences by providing the binary input image that the classification process can rely on.

## IV. CLOSED-LOOP COLOR IMAGE SEGMENTATION

The block-diagram of the proposed color object recognition based on closed-loop image segmentation is shown in Fig. 4. The first included control loop is the thresholding closed-loop which is realized through the feedback between the quality of the binary segmented image and a parameter of the thresholding operation. The second closed-loop is introduced at the second level of image segmentation. It is realized through the feedback between the quality of the improved binary segmented image and a parameter of the applied binary image dilation operation.

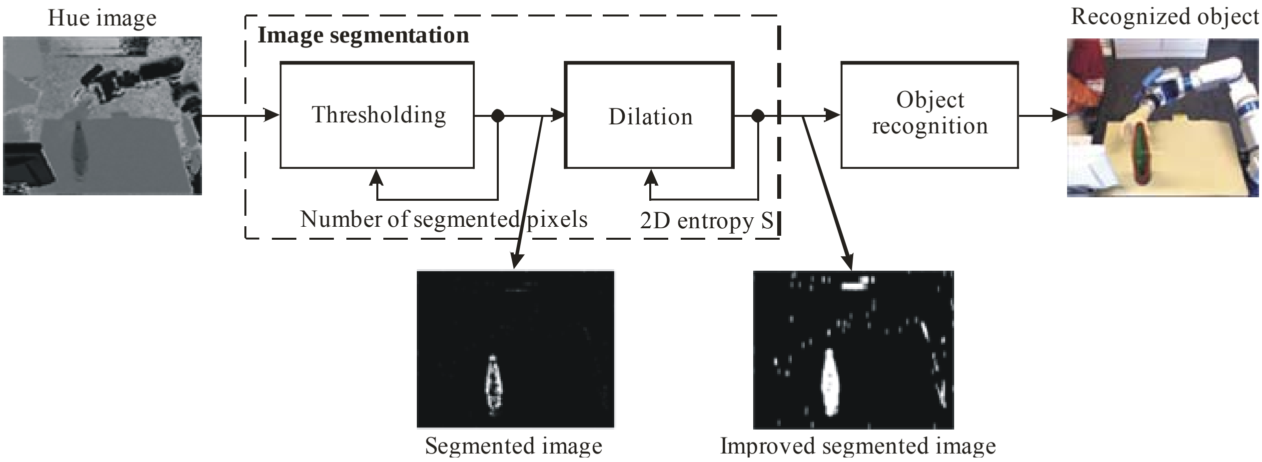
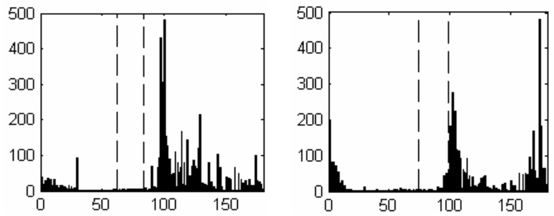



Fig. 4. Robust color object recognition with included feedback control at the segmentation level.

The control goal of the thresholding closed-loop is to provide a binary image which is as free of noise as possible and that contains as many segmented object pixels as possible. The control goal of the dilation closed-loop is to determine the optimal parameter of the dilation operation which provides further improvement of the binary segmented image by "filling" the segmented object region. The overall control goal of both proposed control loops is to provide a good binary input image to the Hu moments based object recognition process. As discussed in the previous section, a good binary input image here means that the segmented object region is "full" and well shaped so that its Hu moments are close to the predefined reference values.

The use of closed-loop control in image processing differs significantly from its use in conventional industrial control, especially concerning the choice of the actuator and the controlled variables. Generally, the actuator variables are those that directly influence the result of image processing. The image resulting from the first closed-loop is a binary segmented image whose quality is basically determined by the distribution of pixel values in the hue image. As discussed in the previous section and illustrated with Fig. 3, the object thresholding interval depends on the lighting conditions. This is further demonstrated in the histograms shown in Fig. 5.




(a) (b)

Fig. 5. Histogram of the reference (a) and alternative (b) hue image overlaid (dashed lines) with the thresholding interval of the object of interest.

Fig. 5(a) shows the histogram of the reference hue image which corresponds to the RGB image acquired in reference artificial light. Fig. 5(b) shows the histogram of the hue image which corresponds to the daylight illumination condition during the image acquisition. Both histograms are overlaid with the thresholding interval of the object of interest (imaged green bottle). Object thresholding intervals were determined by the manual thresholding of both hue images. The result of thresholding was considered as good, once the segmented object region containing as much object pixels as possible was achieved. It is obvious that the thresholding intervals for the same imaged object significantly differ in the images corresponding to different illumination condition. The object thresholding interval in the hue image different from the reference one is shifted with respect to the object thresholding interval in the reference hue image.

Evidently, in order to use the reference object thresholding interval for successful segmentation of particular object of interest, it is necessary to adjust the hue image pixel values. If the lighting condition differs from the reference one, the hue pixel values $f(i)$ in the corresponding hue image differ from those of the reference one. The difference in these pixel values, $\Delta f$, is therefore considered as the actuator variable in the thresholding closed loop. The hue pixel value increment $\Delta f$ can take any integer value from the full pixel value range in the hue gray level image. The number of object white pixels in the current binary image is considered as the measure of the binary image quality and so it is used as the controlled variable. This is justified by the fact that the output binary image is of good quality if it is free of noise and if majority of object pixels are segmented. The reference number of object white pixels *r*, that is the reference quality of binary image, was predetermined off-line by manually segmenting the reference hue image.

Once a binary thresholded image of good quality is achieved, the second feedback loop in the proposed sequential control structure is initialized. In contrast to the first control loop, the second closed-loop is applied only to the region of interest rather than to the entire image. In order to define the region of interest, the contour extraction operation, which searches for the contours of the regions of connected white pixels, is applied to the image resulting from the first closed-loop. Since its quality is equal or close to the reference one within some tolerance interval, the largest segmented region of connected pixels corresponds to the object of interest and so the largest extracted contour defines the region of interest. Bearing in mind that the segmented region of interest is of good quality if it contains full and well shaped object, the measure of the connectivity of white object pixels is used as the controlled variable in the dilation closed-loop. The used connectivity measure, introduced in [11], is the so-called two-dimensional entropy of segmented object pixels in a binary image defined as:

$$S = -\sum_{i=0}^{8} p_{(0,i)} \log_2 p_{(0,i)} , \qquad (9)$$

where $p_{(0,i)}$ is the relative frequency, that is the estimate of the probability of occurrence of a pair $(0,i)$ representing the white segmented pixel 0 surrounded with $i$ white pixels in its 8-pixel neighborhood:

$$p_{(0,i)} = \frac{\text{number of white pixels surrounded with } i \text{ white pixels}}{\text{number of white pixels in the image}} . \qquad (10)$$

The two-dimensional entropy *S* can be also considered as a measure of disorder in a binary image since, as demonstrated in [12], *the higher the two-dimensional entropy S, the larger the disorder in a binary image is.* In the case of object recognition it means that higher values of *S* correspond to broken object regions and lower values of *S* correspond to well connect object pixels forming a well shaped object region. The reference value *r* of *S* was calculated off-line from the manually dilated reference segmented hue image. The process of "filling" the segmented region is strongly dependent on the choice of the dilation structuring element [13]. Since the height of the structuring element is one of the adjustable parameters, it is used as the actuator variable in the second closed-loop.

The discrete PI controller structure used in both closed-loops is implemented in the following velocity form:

$$\Delta u(k) = K_P\left(\Delta e(k) + \frac{1}{T_I}e(k)\right), \tag{11}$$

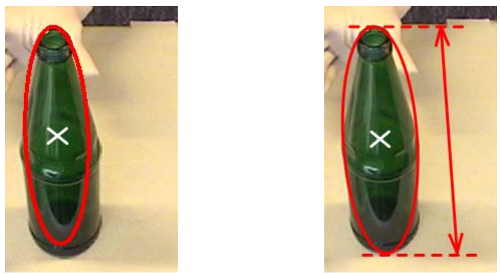
where $k$ is the discrete time index, $e(k)$ is the control error ($e(k) = r - y(k)$), $y(k)$ is controlled variable (number of segmented object pixels in the first closed-loop and the two-dimensional entropy of segmented object pixels in the second closed-loop) and $u(k)$ is the actuator variable (the hue pixel value increment and the height of the dilation structuring element in the first and the second closed-loop respectively). The benefit of using closed-loop control is that the image processing parameters are adjusted automatically. In the case of the first closed-loop the current hue pixel values are incremented so that majority of pixels corresponding to the object of particular color belong to the reference object thresholding interval, as explained before. In the second closed-loop the holes in the segmented object region are filled in the desired way by adapting the considered parameter of the dilation operation. In other words, properly tuned proportional $K_P$ and integral $K_I = \frac{K_P}{T_I}$ gains in (11) provide closed-loops that drive the current binary segmented images to the reference ones independent of the illumination condition during the image acquisition. Controllers were tuned using the so-called "tuning maps" method [20]. Different pairs ($K_P$, $K_I$) were tested in images taken in different illumination conditions including artificial lighting, which is closed to the reference one, as well as daylight condition which significantly differs from the reference lighting condition. Control objective was as small as possible settling time under stability specification for all tested cases. Since the chosen pair ($K_P$, $K_I$) provides stable closed-loop system in all tested cases including bad illumination conditions it is to expect that the system will be stable in conditions different from tested. Off-line predefined reference values for the particular object to be manipulated are stored in the MASSiVE World Model of the system FRIEND II, as explained in Section II. These reference values are delivered on-line to the Machine Vision Framework once the user has defined the task through the Human-Machine interface.

Once the improved binary segmented image is achieved, Hu moments are used to confirm the segmented object region, that is, to distinguish the segmented object of interest from other segmented regions of white pixels. These "extra" segmented regions appeared as noise in the binary segmented image and can be enlarged due to the dilation operation as shown in Fig. 4. In this way, the invariant Hu moments are used for performance evaluation of the proposed closed-loop color object segmentation as it will be demonstrated in the following section. After the segmented regions classification, the recognized object is bounded with the bounding ellipse. The parameters of the bounding ellipse, the major axis and the center point, are the outputs of the object recognition process. These outputs are further used for the 3D localization of the object of interest, as explained in Section II.

## V. Performance Evaluation

In order to evaluate the performance of the proposed closed-loop control based color object recognition, its robustness has been assessed using two measures: human-judged robustness measure and Hu moments of the achieved segmented object region.

A set of images of the FRIEND II environment were taken in different instances of time during the day. Since "beverage serving" scenario is considered in this paper, the imaged object of interest is a green bottle without any texture on it. In the first experiment, the bottle was kept in the same position on the tray of the system FRIEND II, while in the second experiment the position of the bottle was changing. When changing the position, the bottle can be in any location on the tray which is mounted on the front side of the wheelchair. The tray is of the size 54cm x 34cm. The assumption was that the bottle is not occluded by other objects. In both experiments the lighting conditions were varying so that the scene was imaged under the artificial as well as under the daylight conditions. Object of interest was automatically segmented and consequently recognized in each of the acquired images using the proposed closed-loop object recognition method. Automatically obtained object recognition result, containing the parameters of the object bounding ellipse, was compared with the human-judged object recognition result as shown in Fig. 6.



(a) (b)

Fig. 6. Object bounding ellipse resulted from the closed-loop object recognition (a). Human-judged object bounding ellipse (b).

In the human-judged performance evaluation, for every captured image, the position of the vertices of the object bounding ellipse was manually determined by a person. In this way, human-judged values of the major axis and center point of the bounding ellipse were determined. Those values correspond to the ideal (reference) recognition of the object of interest. The difference between the reference object recognition result and the actual result obtained by the proposed recognition algorithm was estimated by calculating the Euclidean distance:

$$e = \sqrt{(x_{1d} - x_1)^2 + (x_{2d} - x_2)^2 + (x_{3d} - x_3)^2}\ , \qquad (12)$$

where $x_1$, $x_2$ and $x_3$ are respectively the actual length of major axis (in pixels), $x$ image coordinate and $y$ image coordinate of the middle point of the actual object bounding ellipse and $x_{1d}$, $x_{2d}$ and $x_{3d}$ are their corresponding human-judged reference values.

In order to obtain a normalized accuracy measure based on the calculated Euclidean distance $e$ the algorithm accuracy measure $\alpha$ was introduced as:

$$\alpha = \frac{\sqrt{x_{1r}^2 + x_{2r}^2 + x_{3r}^2}}{\sqrt{x_{1r}^2 + x_{2r}^2 + x_{3r}^2} + e}\ . \qquad (13)$$

In (13), $x_{1r}$, $x_{2r}$ and $x_{3r}$ are respectively the length of major axis (in pixels), $x$ image coordinate and $y$ image coordinate of the middle point of the object bounding ellipse determined off-line by human-judgment in the reference image. As explained in Section II, the reference image is the image of the green bottle in the FRIEND II environment captured in the reference artificial illumination condition. The accuracy $\alpha$ can have a value from the interval $[0,1]$. As evident from (13) the accuracy value $\alpha$ decreases with the increasing of the Euclidian distance $e$. The value $\alpha = 1$, which corresponds to $e = 0$, is the optimal accuracy value.

To evaluate the invariance of the achieved quality of segmented object region, the accuracy measure $\alpha$ (12) was also calculated using Hu moments of the segmented object regions in binary images resulted from the proposed closed-loop object recognition method. In this case $x_1$, $x_2$ and $x_3$ in Euclidian distance (12) are Hu moments of the actual segmented object region calculated using (4)-(7) and $x_{1d}$, $x_{2d}$ and $x_{3d}$ are their reference values calculated off-line from the reference binary segmented image. For the case of Hu moments based robustness measure, the values $x_{ir}$ in (13) are equal to the described $x_{id}$ values, where $i = 1, 2, 3$.

Fig. 7 shows the closed-loop object recognition algorithm accuracy variation, for human-judged and Hu moments based evaluation, for images captured in different illumination conditions ranging from dark (daily light) to bright (artificial light). The position of the imaged object of interest (bottle) in all considered test images is the same. It is evident that the proposed method is robust with respect to illumination since the human-judged accuracy measure $\alpha$ for all tested different illumination conditions apart from very dark is close to 1. Also, the invariance of Hu moments is evident. This indicates that the achieved segmented object regions are of same quality. The bad object recognition result in images captured in dark illumination condition can be considered irrelevant for the robustness evaluation of the proposed method. Namely, applications of the rehabilitation robotic system FRIEND II are considered to be indoor. For that reason the condition of dark illumination can be avoided since the system FRIEND II operates always either in the very bright daily light conditions or in bright artificially light conditions.

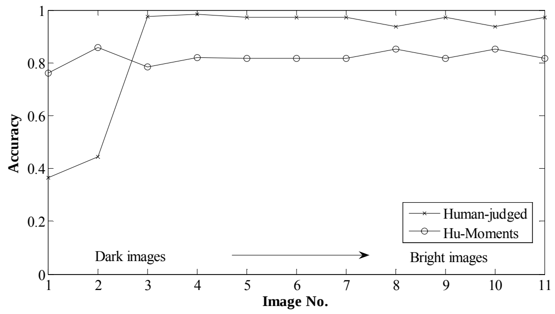



Fig. 7. Accuracy measures $\alpha$ of the object recognition in images captured in different illumination conditions ranging from dark to bright; the case of the constant location of the object of interest.

The algorithm accuracy variation for the second experiment, in which the position of the imaged bottle was changing, is shown in Fig. 8. The tested illumination conditions were ranging from bright to dark.

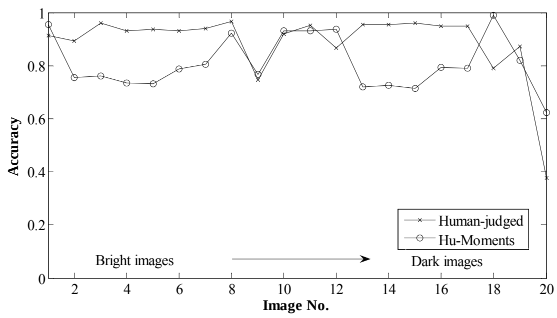



Fig. 8. Accuracy measures $\alpha$ of the object recognition in images captured in different illumination conditions ranging from bright to dark; the case of the variable location of the object of interest.

From Fig. 8 it is evident that the values of both considered performance measures are a bit decreased when considering the object recognition in images of the object placed in different positions. This is an expected result since the reference values for both closed-loops in the proposed object recognition system are determined off-line from the reference image of the object placed in a certain position. Nonetheless, the values 0,86 and 0,81 of the average accuracy, for the human-judged and Hu moments based performance measures respectively, show that the proposed closed-loop color based recognition method has satisfactory robustness with respect to illumination as well as with respect to the different positions of the object of interest. The average accuracy was calculated as:

$$\frac{1}{n} \cdot \sum_{i=1}^{n} \alpha_i \ , \qquad (14)$$

where $n$ is the number of tested images.

## VI. Conclusions and Outlook

In this paper a novel color object recognition method is proposed. The steps of the object recognition chain, including RGB to HSI color space transformation, thresholding of the hue image and binary image dilation, were chosen so that invariant Hu moments can be used for the object recognition. The novel difference between a traditional object recognition chain, consisting of same steps, and the proposed configuration is the inclusion of two control loops at the image segmentation level. The main idea behind this is that the closed-loop drives the actual quality of binary segmented image to the given reference value. In this way, closed-loop image segmentation provides a "good" binary input image to the Hu moments based object recognition independently of the illumination condition. The effectiveness of the proposed method has been demonstrated through the results achieved in the recognition of an object of interest from the "beverage serving" scenario of the rehabilitation robotic system FRIEND II. Although shown experimental results on performance evaluation demonstrated the benefit of closed-loop object segmentation in the achievement of the robust visual perception of the environment, which is a necessary prerequisite for reliable autonomous action of a service robotic system, a few open questions and possible extensions of the presented work come up. These issues are summarized in the following.

The presented object recognition method assumes that the objects of interest are not occluded by other objects in the camera field of view. In order to improve robustness with respect to object occlusion, a measure of the segmented image quality that can be used as feedback variable in thresholding closed-loop, which differs from the one considered in this paper, should be defined.

The presented closed-loop thresholding method considers the entire input image. System parameters (like reference values and control parameters) should be adjusted to achieve reliable thresholding of the region of interest. In this way the presented benefit of closed-loop image processing in coping with variable illumination can be combined with the possibility of semi-autonomous service robotic systems that user selects the object of interest in the displayed image of the viewed scene clicking on it by the cursor [21].

## References


[1] O. Ivlev, C. Martens, A. Graeser, "Rehabilitation robots FRIEND-I and FRIEND-II with the dexterous lightweight manipulator," Restoration of Wheeled Mobility in SCI Rehabilitation, vol.17, July. 2005, pp.111–123.

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

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

[4] M. E. Munich, P. Pirjanian, E. Di Bernardo, L.Goncalves, N. Karlsson, and D. Lowe, "Break-through visual pattern recognition for robotics and automation," in *Proc. IEEE International Conference on Robotics and Automation,* Spain, 2005.

[5] D.G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int.Conf. Computer Vision*, 1999, pp. 1150–1157.

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

[7] S. Ekvall, D. Kragic and F. Hoffmann, "Object recognition and pose estimation using color concurrence histograms and geometric modeling," Image and Vision Comp., vol. 23, 2005, pp. 943-955.

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

[9] G.D.Finlayson, B. Schiele, and J. Crowley, "Comprehensive Colour Image Normalization," in *Proc. 5th European Conf. Computer Vision,* Freiburg, 1998 , pp. 475–490

[10] D. Ristic, I. Volosyak and A. Gräser, "Feedback control in image processing," atp int. automation technology in practice, 2005, pp. 61-70.

[11] D. Ristic and A. Gräser, "Performance measure as feedback variable in image processing," EURASIP Journal on Applied Signal Processing, vol. 2006, 2006, pages 12.

[12] D. Ristic, S. K. Vuppala and A. Gräser, "Feedback control for improvement of image processing: an application of recognition of characters on metallic surfaces," in *Proc. 4th IEEE Int.Conf. Computer Vision Systems,* New York, 2006, pp. 39-39.

[13] R.-C. Gonzalez, R. –E. Woods, *Digital Image Processing.* NJ, 2002.

[14] T. Okabe and Y. Sato, "Support vector machines for object recognition under varying illumination conditions", in *Proc. Asian Conf. Computer Vision,* Jeju Korea, 2004, pp.724-729.

[15] M.K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. on Information Theory*, 1962, pp. 179-187.

[16] M. Mercimek, K. Gulez, and T.V. Mumcu, "Real object recognition using moment invariants," *in Proc. Sadhana-Academy Engineering Sciences,* India, 2005, pp. 765-775.

[17] J. Flusser, "Moment invariants in image analysis," Tran. Engineering Computing and Technology, vol 11, 2006, pp.196-201.

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

[19] S. Vinoski, and M. Henning, *Advanced CORBA Programming with C++*, Addison Wesley, 2004.

[20] D. Ristic, "Feedback structures in image processing", Ph.D. dissertation, to be published.

[21] 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.