Automated Assistance Robot System for Transferring Model-Free Objects From/To Human Hand Using Vision/Force Control

Mohamad Bdiwi[1], Alexey Kolker[2], Jozef Suchý[1], and Alexander Winkler[1]

[1] Department of Robotic Systems, Chemnitz University of Technology, Chemnitz, Germany
[2] Automation Department, Novosibirsk State Technical University, Novosibirsk, Russia
mohamad.bdiwi@s2008.tu-chemnitz.de

Abstract. This paper will propose an assistance robot system which is able to transfer model-free objects from/to human hand with the help of visual servoing and force control. The proposed robot system is fully automated, i.e. the handing-over task is performed exclusively by the robot and the human will be considered as the weakest party, e.g. elderly, disabled, blind, etc. The proposed system is supported with different real time vision algorithms to detect, to recognize and to track: 1. Any object located on flat surface or conveyor. 2. Any object carried by human hand. 3. The loadfree human hand. Furthermore, the proposed robot system has integrated vision and force feedback in order to: 1. Perform the handing-over task successfully starting from the free space motion until the full physical human-robot integration. 2. Guarantee the safety of the human and react to the motion of the human hand during the handing-over task. The proposed system has shown a great efficiency during the experiments.

Keywords: Assistance robots, human-robot interaction, visual servoing, force control, handing-over tasks.

1 Introduction

Recently, it has been noticeable that the advanced robot applications have become focused on the tasks which require from the robot to interact with the human physically, e.g. assistance robot for blind, disabled or elderly people for helping in fetching, carrying or transporting objects. In other applications, robot could serve as rescue, service robots or even as member of human-robot teamwork for supporting humans in such applications as space exploration, construction, assembly etc. In such kind of applications, where human hand will interact with robot hand, the initial core is to perform the handing-over task from/to human hand successfully starting from hand tracking in free space motion until to the full physical interaction between the human hand and robot hand.

Numerous papers which have suggested different approaches of service robot systems have discussed the problems of objects transfer between human and robot. With robot HERMES [1] and in the work [2], the transfer tasks are performed exclusively by the human. This means that the robot will bring its hand into a specified pose and

G. Herrmann et al. (Eds.): ICSR 2013, LNAI 8239, pp. 40–53, 2013.

then it will wait until the human places the transported object between the fingers of the gripper. When the robot detects that an object has been placed in its hand, it attempts to grasp the object. In fact, this scenario will not be fit to assist blind, disabled or elderly people or even to support workers concentrating on their work. References [3] and [4] have presented different algorithms of grasp planning during handing-over objects between human and robot. These algorithms depend on human preferences, on the kinematic model of the human and on the availability of the 3D model of the object. Related to that, various papers have focused on analysing and detecting human body movements especially on hand gesture and facial features, e.g. [5], [6] and [7]. The common properties of them are that they detect the human hand in order to direct or to lead the robot for performing of some tasks without any physical interaction with the robot. In addition to that, they have implemented algorithms which are able to handle with only human hand free of load. Hence, what about if the human hand has carried an object and the task needs the robot to interact with the human physically? Moreover, what about if the user is disabled or elderly?

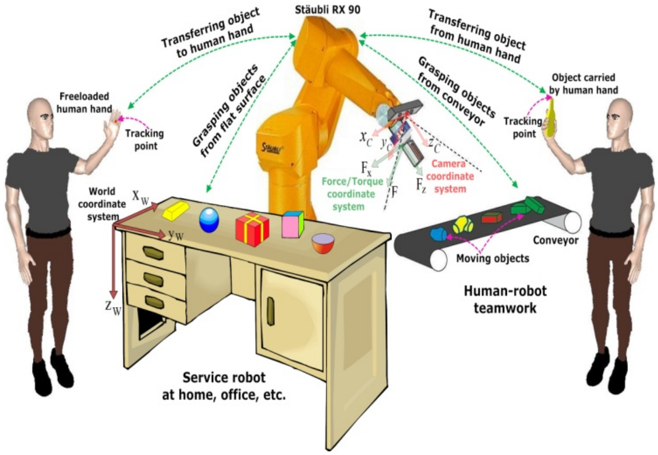

Fig. 1. Handing-over from/to human hand

On the whole, according to our knowledge none of the previous works has proposed a fully automated robot system which combines vision and force control in order to transfer model-free objects from/to human hand. Furthermore, in this work, the proposed robot system will consider the human as the weakest party (negative party) during the handing-over task, i.e. this party could move in a wrong direction (not toward the robot), e.g. elderly or blind or he/she is doing something else at the same time. In this case, the robot system should expect some random motions from human during the task and react accordingly to them. In addition to that, the proposed system has implemented real time vision algorithms for detecting and segmenting loadfree human hand, any objects carried by human hand and any object located on a flat surface. The fusion of vision and force sensors is an optimal solution to guarantee the safety of the user and for performing the handing-over task successfully.

In the next section, the hardware and software equipments will be presented. Section 3 contains general description of the whole algorithms of the proposed system. In section 4, real results will be illustrated for transferring objects from/to human hand. The last section contains conclusion, future work and the benefits of improving the physical interaction between human and robot.

2 System Equipment

As shown in Fig. 2, the overall experimental setup consists of Stäubli RX90 robot with a JR3 multi-axis force/torque sensor together with the eye-in-hand Kinect camera as shown in Fig. 2. The end-effector is installed on the collision protection device. In our application the end-effector is two-finger gripper. JR3 (120M50A) is a six component force/torque sensor and its effective measurment range is ± 100 N for forces F_x, F_y, F_z and ± 10 N.m for torques M_x, M_y, M_z. Kinect camera is RGBD camera which delivers depth and color images with VGA resolution (640x480 pixels). As shown previously in Fig. 1, the active human hand will interact with the robot and it could be free of load or carry an object which will be transferred from it.

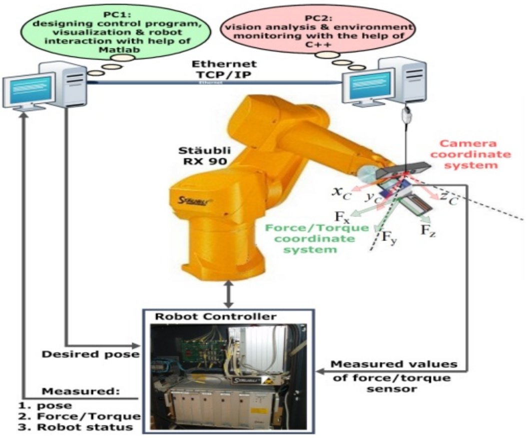

Fig. 2. Hardware/software equipments

The proposed system has used two PCs. PC1 will control the robot to perform the visual and force tracking tasks with the help of MATLAB program and V+ (programming language of Stäubli robot). A special program is designed to integrate Matlab program with V+ language, which means that all V+ instructions, e.g. read/write pose, read force/torque or robot status, could be written directly in Matlab program. PC2 is connected with the Kinect camera, and all the image processing algorithms are performed in PC2 using C++ language (OpenCV and Openkinect libraries). PC2 will send the position of the face and of the object as well as the status of the task to PC1 in every frame using Ethernet TCP/IP protocol as follows:

$$\left[Obj_{x_C}, Obj_{y_C}, Obj_{z_C}, time\_dif, vision\_status\right] \tag{1}$$

where $(Obj_{x_C}, Obj_{y_C}, Obj_{z_C})$ is the tracking point of the carried object or of the human hand in the case of loadfree hand, this point will be later the contact point between the object and the gripper. $time\_dif$ is time difference between two frames. $vision\_status$ represents the current status of the vision system if only loadfree hand is detected or a carried object is segmented and detected.

3 Description of Proposed Algorithms

This section will present the algorithms of the whole process of the proposed system. As shown in Fig. 3, firstly the user will define the type of the task, either transferring from or to human hand, based on keyboard commands. If the task is to transfer to

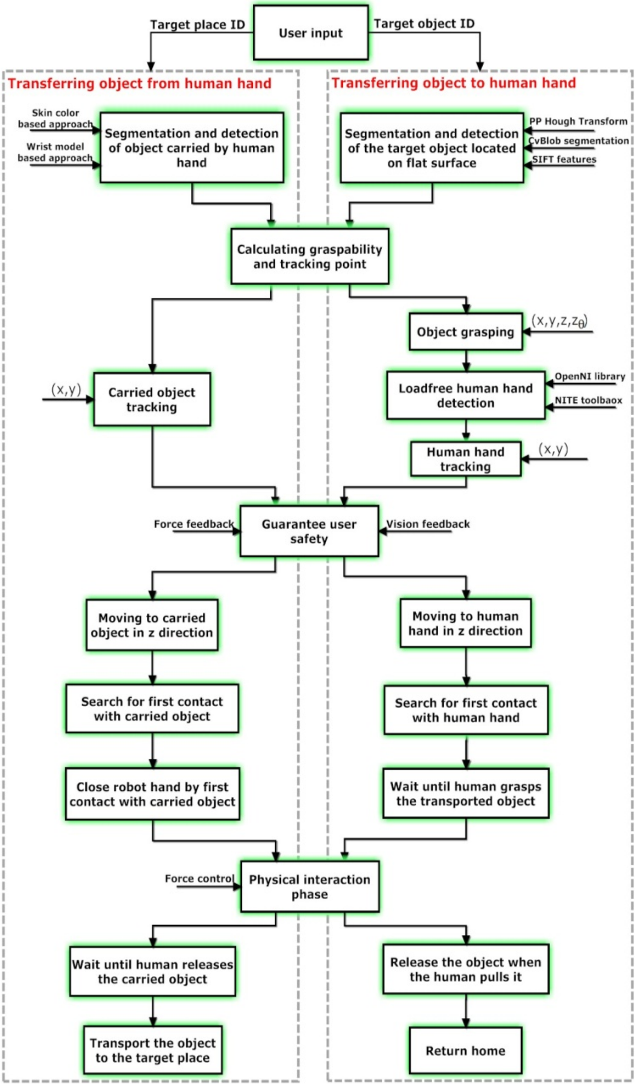

Fig. 3. Algorithm flowchart

human hand, the user should enter either target object ID or name under which its corresponding SIFT features are saved in the database. In other hand, if the task is transferring from human hand, the user needs only to enter the ID of the target place where the object should be transported, e.g. table, workspace, conveyor, etc.

User interface could be improved in the future to be based on voice commands. The currently proposed system is supported only by voice subsystem, i.e. it will announce the current phase (what it is going to do), the status of the operation or whether some errors have occurred. The proposed speaking subsystem will give the human the opportunity to learn and to understand what the robot is doing now and to be prepared if any error has occurred during the task. It will increase the safety factor between human and robot, especially if the user is disabled or blind.

A. Vision algorithms

The next step in the proposed robot system are two different real time vision algorithms. The first one will be used, if the task is to transfer the object from human hand. The proposed vision algorithm will detect and segment any object carried by human hand without any a priori information about its model or color. The proposed vision algorithms for detecting and segmenting the carried object consists of two approaches: 1. Skin color based approach and 2. Wrist model based approach.

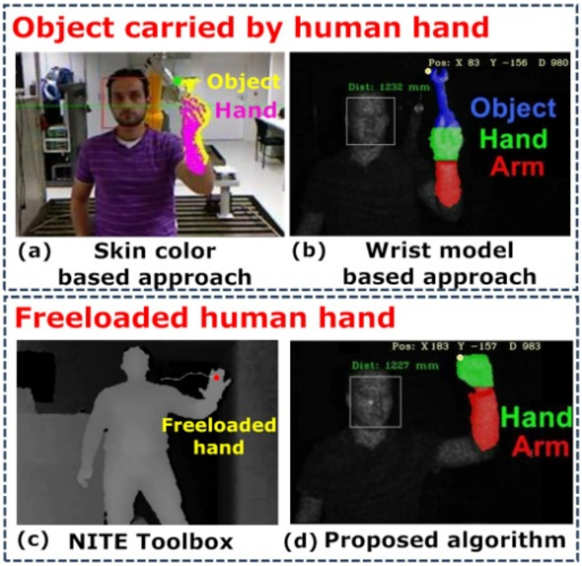

Fig. 4. Results of vision algorithm for carried object

Briefly, in the skin color based approach the vision system will segment the human body and carried object from the background, then it will define the area of interest which contains the human hand with the object. Finally, it will segment the human hand from the carried object using skin color mask. The proposed system will implement face detector to define the average skin color of the active user in HSV space by scanning the HSV values of the detected face. The previous work [8] has defined the general range for human skin color in HSV space as follows (by assuming the range of H, S and V components is 0,...,255):

$$\begin{aligned} G_{H_{\min}} &= 0 \quad and \; G_{H_{\max}} = 128 \\ G_{S_{\min}} &= 59 \; and \; G_{S_{\max}} = 175 \\ G_{V_{\min}} &= 0 \quad and \; G_{V_{\max}} = 255 \end{aligned} \tag{2}$$

It is clear that the general range of human skin color in HSV space is very wide, e.g. V values could be spread in whole space. H and S values could be spread in almost the half of the space. Hence, the general ranges could be sufficient to detect skin but they will not be enough to segment the hand from the object in a complicated scene or in different light conditions, especially if the object has almost the same color as the skin such as wood objects. Hence, by scanning the color of the detected face, the color of human skin will be updated every frame if light source, active person and skin reflection are changed. This will help to narrow the ranges of skin color in HSV space and to define exactly the color of the human hand. Fig. 4(a) presents experimental results of the skin color based approach. One limit of this approach is that it cannot be used when the human wears gloves or he/she has Vitiligo disease. Hence, another approach, which depends on wrist model, has been implemented using infrared frames.

Fig. 5 presents the basic principle of the proposed approach. The most important region of the hand is the wrist (the joint between the hand and the arm). By analyzing the cross-section of the human hand parallel with plane xz, it can be concluded that an algebraic equation of the second order can approximately represent all the points of cross section at any value of y in the wrist region as follows:

$$\bar{Z}_k(x_i) = a_2 \cdot {x_i}^2 + a_1 \cdot x_i + a_0; \quad i = 1, \dots, N \tag{3}$$

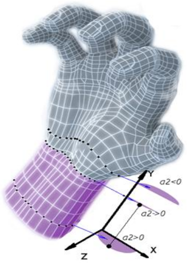

Fig. 5. Human hand profile

where $\bar{Z}_k(x_i)$ are the approximated depth values. Using least square polynomial approximation, the system can calculate the coefficients a_2, a_1 and a_0. In general, the coefficient a_2 in any second order equation describes the flexion of the curve. Hence, by analysing the human hand profile, it can be concluded the following: If the cross section locates below the wrist joint, the coefficient a_2 will be always positive (as shown in Fig. 5, the region of violet color). The value of coefficient a_2 will become unstable (sometimes negative and sometimes positive) exactly when the cross section crosses the wrist joint upwards. This approach will be useful even if the human wears gloves or has Vilitigo disease and it is able to segment the human hand from the object even if they have the same color starting from complete darkness and ending with different color temperature lamps, as shown in Fig. 4 (b). The frequency of first approach is 5 frames/sec, whereas the frequency of the second approach is 10 frames/sec, i.e. they are fit for fast real time visual servoing.

Furthermore, the proposed vision algorithms are able to detect and segment the loadfree hand as shown in Fig. 4 (d). In contrast, the hand control of NITE toolbox from Kinect camera could be implemented to track only the loadfree hand, as shown in Fig. 4. (c).

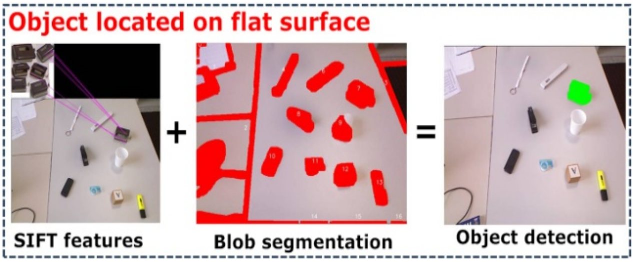

Fig. 6. Result of vision algorithm for object on flat surface

On the other hand, as shown in Fig. 6 the second algorithm for transferring objects to human hand is able to detect and segment any object located on a flat surface. This approach works with SIFT features and it is able to detect the object in real time even if the object has complicated shape, if it has bad contours and if the illumination has been permanently changed and unequally distributed with no need to a priori model. The proposed algorithm in this section has combined different algorithms of image processing such as Canny filter, progressive probabilistic Hough transform (PPHT [9]) and some morphological operations for improving objects contours. In general, the proposed vision algorithms in this work have many contributions which will be illustrated in details in other papers.

B. Graspability and tracking point

After detection and segmentation the target object (even if it is carried by human hand or located on a flat surface), the next phase will include algorithm common to both tasks as shown previously in Fig. 3. In this phase the robot system will calculate the graspability and the position of the tracking point depending on the results of vision processing.

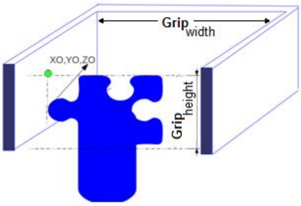

Fig. 7. Object *and robot hand*

As shown in Fig. 7, the graspability will be calculated by comparing the width and the height of the target object $(obj_{width}, obj_{height})$ with robot hand width $Grip_{width}$ (distance between fingers of the robot hand) and with (height of robot hand) $Grip_{height}$ as follows:

$$\begin{aligned} & if\ \left(obj_{height} > Grip_{height}\right)\ \&\ \left(obj_{width} < Grip_{width}\right)\ then \\ & \qquad \{Grasp\_ability\ = 1\} \\ & else\ \{Grasp\_ability\ = 0\} \end{aligned} \tag{4}$$

In brief, the object should have grasp area GA(i,j), where its width is less than the gripper width and its height is greater than gripper height in order to activate the graspability. If the robot is able to grasp the object then it will calculate the tracking point of the object which will be later the contact point. The vision system will scan the whole grasp area GA(i,j) and it will define three points of it: 1. Upper point (x_U, y_U, z_U), 2. Left point (x_L, y_L, z_L) and 3. Nearest point (x_N, y_N, z_N). The tracking point (x_O, y_O, z_O) will be calculated as follows:

$$\begin{aligned} x_O &= x_L \\ y_O &= y_U \\ z_O &= z_N \end{aligned} \tag{5}$$

In the task of transferring object to human hand, the next step will be grasping the object from the flat surface. Fig. 8 illustrates the coordinate system of the object. The vision system will control the directions (x, y, z) and the orientation (θz) in order to rive the robot hand to the best grasping position. When the robot arrives at the tracking point of the object, it will move 20 mm in z direction toward the object before establishing any contact with the object. After that the robot will start searching for the first physical contact with the object in x direction. In this way, the robot can guarantee that it will grasp the object successfully by closing the gripper fingers.

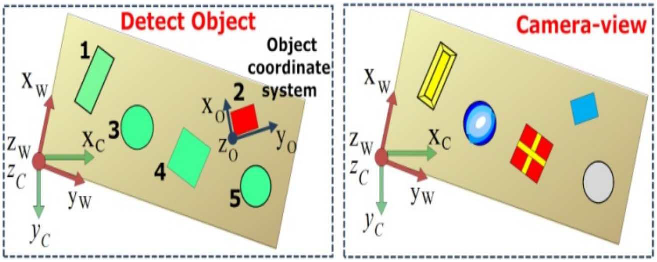

Fig. 8. Grasping object from flat surface

After grasping the object from the flat surface, the vision system will start searching for the loadfree human hand. In this case, the NITE toolbox which supports Kinect camera could be implemented for detecting the human hand and finding the middle point of it, because its frequency is 30 frames/sec.

C. Tracking phase

As shown in Fig. 3, the following phase will be the tracking phase either of the loadfree human hand or of the object carried by human hand depending on the task. The robot will start to track the target object or the hand in two directions, x and y. The tracking algorithm can be easily extended to three directions x, y and z. However, in this phase, a safety distance in z direction between the robot hand and the human hand should be always kept. The robot will not move toward the human hand in z direction to establish the contact until it ensures that the human hand doesn't move anymore. When the human hand is in a stable position and after the robot has already tracked it successfully, the robot system will start moving toward the human hand in z direction.

Hence, the main purpose of the visual tracking in this phase is to preserve the target point of the object or the hand at the middle of camera frame. $(\mathrm{Cam_x}, \mathrm{Cam_y})$ represent the position of the middle point of the camera. Hence, the required distance (mm) in order to locate the tracking point of the object in the middle of camera's view (position based visual servoing approach) is as follow:

$$\begin{aligned} Err_x &= Obj_{x_C} - Cam_x \\ Err_y &= Obj_{y_C} - Cam_y \end{aligned} \tag{6}$$

However, real time tracking of moving object is not an easy task, especially when this object is moving randomly e.g. when tracking the hand of blind user. In our approach, the robot is able track the loadfree hand or the object carried by a human hand smoothly and with sufficient speed in real time in spite of the following difficulties: 1. The position of the object continually changes and it can move in all directions. 2. The speed and acceleration of the object motion are not constant and the motion direction of the object can suddenly change. 3. The motion speed and desired position in the implemented robot (older commercial robot) cannot be changed when the robot is moving, i.e. the robot is not able to receive a new target position unless the previous motion is finished. Hence, the common position controller will not be sufficient, especially when the object direction and position can suddenly be changed by human hand. Therefore, in order to track the object smoothly and with sufficient speed in real time, the proposed approach will not control the speed of the robot directly but it will control motion steps of the robot. Whenever the target point is farther, the robot will move toward it with a greater step.

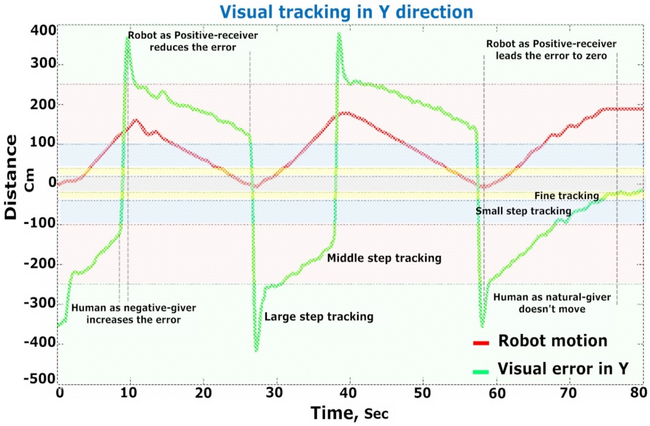

Fig. 9. Visual *tracking in Y direction*

Fig. 9 shows the experimental results of the proposed visual tracking approach in Y direction. The green diagram presents the visual error Err_x (the difference between the tracking point of the object or hand and the camera middle point) in the camera coordinate system, whereas the red diagram represents the motion of robot in the world coordinate system. The different colored zones refer to the different motion step of the robot. From the behaviour of robot motion (red diagram) it is clear that the robot always moves toward the target point as positive party therefore the visual error is diminished to zero. On the contrary, from the behaviour of the human it is clear that the human increases the visual error extremely (as negative party) by moving his/her hand, look e.g. at time t = 10s in Fig. 9).

At time t = 0s, the visual error was near to -350cm which made the robot move upwards to reduce the visual error. Starting in time t = 0s, until t = 2s, the visual error was inside the large step tracking zone and it has diminished. At time t = 2s, the visual error has come inside the middle step tracking zone. However, at time t = 8s the human hand has suddenly been moved and it has changed the visual error from -120cm to 380cm which has led the robot to react rapidly and to change the direction of the motion in order to track the target point. The rapid visual tracking and changing directions will be continued until the human hand acts at least as a natural giver (doesn't move anymore), e.g. at time t = 58s, so the robot will move toward the target point smoothly and the visual error will be diminished until it enters the zone0 (visual error is less than 2cm). It means that the robot has arrived at the target point, after that the visual tracking will be finished and next phase will be started.

D. Safety measures

After tracking phase, the robot will be ready to move toward the human if the safety procedures are taken into account. Numerous papers have been published which have proposed different solutions for improving the safety factor during the human-robot interaction, such as improved mechanical design by reducing the robot mass [10] and motion trajectories related to the human body constraints [11]. However, in our opinion, even if the system would use lightweight robot and predefined trajectories, it is indispensable to integrate vision and force information to guarantee the safety especially when unexpected problems or errors happen during the physical interaction. In this work, three safety measures are implemented: The first one (SF_body) is related to the safety of the whole human body depending on the depth map, whereas the second safety (SF_hand) is related only to the safety of the fingers during handing-over the object (if the robot is able to grasp the object without touching the human fingers). The third one is (SF_force) which monitors the force values, especially when the robot is moving toward the human in z direction. Values of these factors will be set to zero as long as the safety requirements are fulfilled. Otherwise, if any error or dangerous situation of human is recognized or unexpected obstacle is encountered, the safety variables will be immediately activated and the task will be cancelled.

E. Physical interaction phase

If the tracking phase is successfully performed and the robot or the human has grasped the transported object, the physical interaction phase will start. In the first phase, the first party has grasped the transported object but the other party has not released it yet, i.e. the transported object will be like a kind of connection bridge between the human and the robot hand. In this phase, we will assume that the human or the robot may not release the object immediately. Perhaps, the human would like to drive the robot toward another place, so the robot should be able to react to the motion of the human hand, so the robot will comply with the motion of the human hand based on user of the force sensor. Finally, if robot transfers the object to the human hand, it will release the object when the human pulls the object with desired force and after that it will return to home position. Otherwise, by transfer of object from human hand, robot will wait until the human releases the object (no more applied forces) then it will transport the object to the target place.

4 Experiment Results

Some results have been already presented in Fig. 4, Fig. 6 and Fig. 9, whereas, Fig. 10 and Fig. 11 present further experimental results from handing-over task.

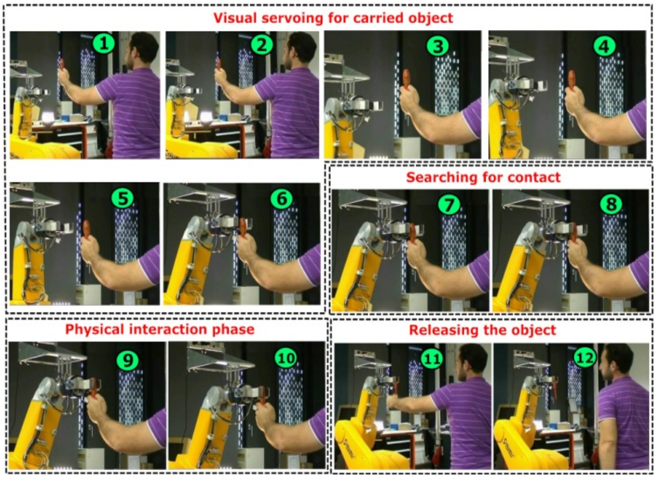

Fig. 10. Transfer object from human hand

In Fig. 10, as shown in the first six pictures, the robot is tracking and moving toward the object carried by human hand in order to grasp it from the human hand with the help of vision system. When robot reaches the tracking point of the object, the robot will start searching for the first contact point with the object using force reading, as shown in pictures 7 and 8. When the force sensor measures that the desired contact with the object has been established, then the robot will close its hand to grasp the object as shown in pictures 9 and 10. Hence, the physical interaction between the human and the robot will start and the robot will react to the motion of the human hand on the basis of force sensor measurements. When the interaction phase is finished, the human will release the object as shown in pictures 11 and 12.

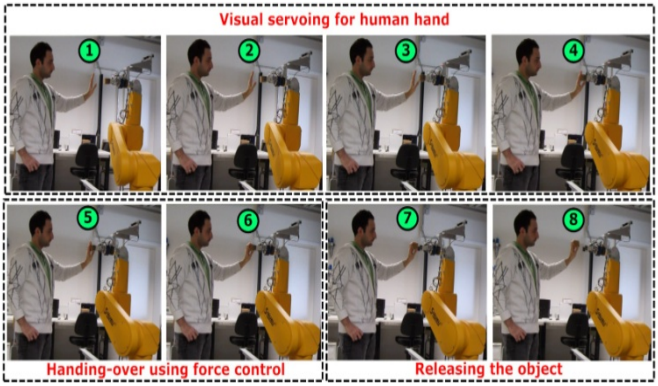

Fig. 11. Transfer *object to human hand*

In Fig. 11, as shown in the first four pictures, robot is moving toward the human hand in order to deliver the cube with the help of vision system. In pictures 5 and 6, the human will start grasping the object by closing his/her hand. In pictures 7 and 8, robot will perceive the contact force. It will release the object when the human starts pulling.

As shown previously in Fig. 10 and Fig. 11, the proposed system will assume that the human as giver/receiver will be the weakest part (blind, elderly etc.) of the task and the robot will play the main role as a positive party to perform the transfer task. In other words, human only needs to open/close his/her hand in order to release/grasp the object.

The proposed algorithms have been repeated for more than 15 different objects which were carried by different users.

Fig. 12. Dataset *of different transported objects*

Fig. 12 presents the dataset of the transported objects from/to human hand. In all experiments of handing-over objects from human hand, the robot system was able to segment, to track and to grasp the carried object from the human hand successfully. As is shown in the Fig. 12, the implemented objects are textureless, so in the experiments of handing-over objects to human hand, if the robot system was able to detect the target object which is placed on flat surface depending on SIFT features, surely the robot system will be able to deliver that object to the human hand.

5 Conclusion

This paper has just illustrated the general description of the proposed robot system, more details about the vision algorithms, vision/force control structure [12], tracking algorithms, safety procedures and physical interaction phase will be found in other papers. This work has proposed an automated robot system which is able to deliver and receive model-free objects to the human hand automatically. This work has proposed different real time image processing algorithms in order to detect and to segment loadfree hand, any object carried by human hand and any object located on a flat surface. The proposed robot system has implemented visual servoing algorithms for tracking the loadfree hand or any object carried by it in real time. In the proposed system, the transfer object from/to human hand is performed exclusively by robot and the human has been considered as the weakest part in this task (elderly, blind or disabled). Furthermore, for improving the handing-over task the proposed system has combined vision and force control. The fusion of vision and force will ensure the safety of the user during the physical human-robot interaction, it will ensure the fulfilment of the grasping/releasing task and it will make the robot able to react to the motion of human hand during the interaction phase. The proposed robot system could be easily modified in order to fit different scenarios and applications, e.g. in connection with the service, assistance, rescue and mobile robots or even to work simultaneously with the human in industrial human-robot teamwork.

In future work, tactile sensors could be integrated with the system to optimize the handing-over tasks. In addition to that, a voice command subsystem could be implemented.

References

[1] Bischoff, R., Graefe, V.: HERMES - a Versatile Personal Assistant Robot. Proc. IEEE - Special Issue on Human Interactive Robots for Psychological Enrichment 92, 1759–1779 (2004)

[2] Edsinger, A., Kemp, C.: Human-Robot Interaction for Cooperative Manipulation: Handing Objects to One Another. In: 16th IEEE International Conference on Robot & Human Interactive Communication, vol. 2, pp. 1167–1172 (2007)

[3] Cakmak, M., Srinivasa, S., Lee, M., Forlizzi, J., Kiesler, S.: Human Preferences for Robot-Human Hand-over Configurations. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, USA, pp. 1986–1993 (2011)

[4] Kim, J., Park, J., Hwang, Y., Lee, M.: Advanced Grasp Planning for Handover Operation Between Human and Robot: Three Handover Methods in Esteem Etiquettes Using Dual Arms and Hands of Home-Service Robot. In: 2nd International Conference on Autonomous Robots and Agents, New Zealand, pp. 34–39 (2004)

[5] Hussain, A.T., Said, Z., Ahamed, N., Sundaraj, K., Hazry, D.: Real-Time Robot-Human Interaction by Tracking Hand Movement & Orientation Based on Morphology. In: IEEE International Conference on Signal and Image Processing Applications, pp. 283–288 (2011)

[6] Chuang, Y., Chen, L., Zhao, G., Chen, G.: Hand Posture Recognition and Tracking Based on Bag-of-Words for Human Robot Interaction. In: IEEE International Conference on Robotics and Automation, pp. 538–543 (2011)

[7] Yuan, M., Farbiz, F., Mason, C., Yin, T.K.: Robust Hand Tracking Using a Simple Color Classification Technique. The International Journal of Virtual Reality 8(2), 7–12 (2009)

[8] Phung, S., Bouzerdoum, A., Chai, D.: Skin segmentation using color pixel classification: Analysis and comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(1), 148–154 (2005)

[9] Galambos, C., Matas, J., Kittler, J.: Progressive probabilistic Hough transform for line detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 554–560 (1999)

[10] Zinn, M., Khatib, O., Roth, B.: A New Actuation Approach for Human Friendly Robot Design. Int. J. of Robotics Research 23, 379–398 (2004)

[11] Mainprice, J., Sisbot, E., Simeon, T., Alami, R.: Planning Safe and Legible Hand-Over Motions for Human-Robot Interaction. In: IARP Workshop on Tech. Challenges for Dependable Robot in Human Environments (2010)

[12] Bdiwi, M., Suchý, J.: Library Automation Using Different Structures of Vision-Force Robot Control and Automatic Decision System. In: IEEE/RSJ International conference on Intelligent Robots and Systems, pp. 1677–1682 (2012)