# Stereovision System for Estimation of the Grasp Type for Electrotherapy*

## Matija Štrbac[1], Marko Marković[1]

**Abstract:** This paper presents hardware and software for scene analysis that are designed for the system used in treatment of post stroke hemiplegic patients using electrical stimulation. New hardware includes two cameras and a laser pointer, while new software is given as a Matlab program that performs real-time estimate of size and shape of targeted object. Based on heuristic contemplation the system makes a decision grasp type and necessary actions for the purpose of hand opening and closing. The system was tested on 13 objects and in 95% of cases it worked according to demands, i.e. corresponding to choices of healthy subjects when they wanted to grab that same object.

**Keywords:** Stereovision, Scene analysis, Distance estimation, Image correlation, Object identification, Ggrasp type decision, Functional electrical stimulation.

## 1 Introduction

Functional electrical stimulation (FES) is used for the treatment of individuals whose body parts are paralyzed or paretic as a result of neurological disorders (e.g. stroke, spinal cord injury). In recent years it has been shown that the application of FES in the treatment of upper extremities on the patients who suffered from stroke has a therapeutic effect [1], and it is of interest to improve the system to allow automatic selection of the grasp type. This addition to the electrical stimulator would eliminate the need for manual grasp type selection from a patient or a therapist, and would allow automatic selection of the required electrical impulses that stimulate muscles whose contraction would lead to suitable grasp for observed object. During this assessment three types of grasping, ones most commonly used for daily activities, are taken into consideration (Fig. 1).

To be precise, the results of this study are related to the Man-Machine Interface (MMI) [2] with direct reliance on the area of computer vision. General requirements that are placed in front of such a system are:

---

[1]University of Belgrade, School of Electrical Engineering, Kralja Aleksandra 73, 11000 Belgrade, Serbia; E-mails: sir.matija@gmail.com;  konvergentan@gmail.com

1. The ability to make decisions in real time;
2. Following defined system restrictions, a small error in decision making, which can be corrected by a user, is anticipated;
3. Clarity, comfort and user-friendliness.



**Fig. 1 –** *Pinch, lateral and palmar grasp are used
in more than* 90% *of all typical daily activities.*

Specific task is to extract and identify the object of interest in a particular scene using computer vision, taking into account some of the objects used in daily living activities. The recognition refers to the object contour detection and assessment of its size and orientation.

Most of the known solutions use a single camera, in static or in eye-in-hand [3] configuration, and different methods of determining the distance from the object to the camera (sound or light echo, stereovision) to estimate its actual size. The assessment is based on the initial knowledge of the image size within a single pixel when the distance of the analyzed scene is known.

One of the tasks of this study was to test the applicability of two web cameras for the distance estimation using methods that are characteristic for stereovision in the quasi-static circumstances.

## 2   Methods and Materials

Stereo Vision System (SVS) is consisted of:
1) Two identical CCD web cameras, the highest resolution of 1600×1200 pixels, both placed on the plastic stand, which guarantees the same horizontal orientation. The distance of focal points of cameras is 6.2 cm and their optical system is set so they can see sharp images at distances between 30 and 90 cm. Each camera is connected to PC with USB cable.
2) One red laser diode, placed in the center between two cameras and focused in the same direction. The power of the diode which is shared with one of the cameras is recurrently switched on and off, while serial communication with PC provides the information of desired state. Essential purpose of the laser is to enable feedback on the camera orientation, i.e. whether the object of interest is in the central part of the scene viewed by cameras. Also, the laser mark is used as the element to which the distance is estimated, all with the intention of later object size assessment.
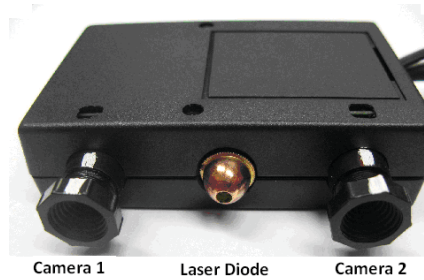
**Fig. 2** – *Image of the formed SVS that includes two web cameras, laser pointer and the required electronics. The system is connected to the PC and it communicates via serial and USB interface with the program that works in Windows and Matlab environment.*

SVS is designed to be integrated into a light hat or some similar article that a patient could put on the head so that the cameras and laser are oriented in the direction of the working space in which the objects are located (approximately 0.8 radians down the horizontal plane). This approach is different to the eye-in-hand method shown in the references [4], which was developed for an intelligent artificial hand. The proposed concept can be viewed as quasi-stationary, because the cameras move only when it comes to new object selection. The advantage of this setting in relation to the eye-in-hand is primarily a stationary scene which is analyzed, eliminating the effects of camera vibrations caused by hand tremor that is to some extent present in most patients, and also a reduced amount of overlapping objects in the camera perspective. The disadvantage that occurs in both configurations is that the system provides 2D image as a result of 3D space projection on camera plane, and that can lead to an error in object size assessment, thus resulting in a grasp which is not appropriate for that object.

The image acquisition process is performed in YUV color space, at the resolution of 320x240 pixels on the desktop PC (single core P4 3GHz, 1GB RAM) using the MATLAB 2007b software package.

Fundamentals of how the program operates are described in a given flow chart (Fig. 3).

Laser detection starts with the image conversion from the YUV to the YCbCr color space for the easier detection of the red component. All pixels that meet the condition that their Y component and 1.5Cr-0.4Cb component subtraction is greater than the threshold which is 90% of maximum value in both terms are potential candidates. In order to reduce the impact of false detection (primarily as a consequence of reflecting surfaces that have large Y component), on pixels that have passed the first threshold additional morphological operations with hit&miss [5] algorithm is performed. Only group of pixels whose morphology is similar to imperfect dot will pass.
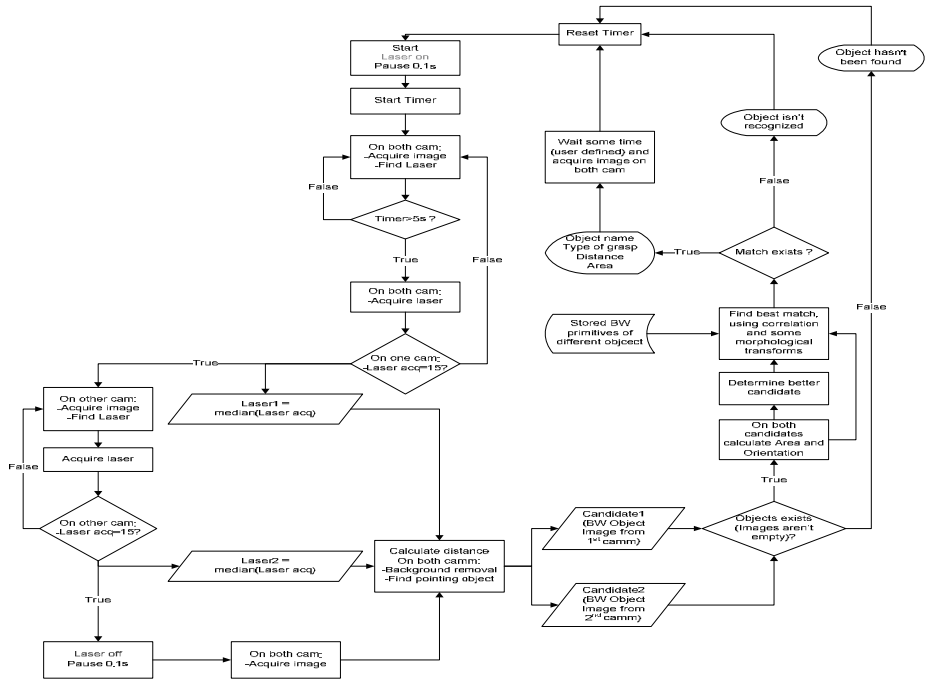
**Fig. 3 –** *Algorithm for object detection and grasp type identification.*

Laser detection starts with the image conversion from the YUV to the YCbCr color space for the easier detection of the red component. All pixels that meet the condition that their Y component and 1.5Cr-0.4Cb component subtraction is greater than the threshold which is 90% of maximum value in both terms are potential candidates. In order to reduce the impact of false detection (primarily as a consequence of reflecting surfaces that have large Y component), on pixels that have passed the first threshold additional morphological operations with hit&miss [5] algorithm is performed. Only group of pixels whose morphology is similar to imperfect dot will pass.

Object detection is done by removing the background with the usage of color segmentation [6]. The candidates for the background are those pixels whose value of RGB component slightly differs from the median of the same. In this process the initial primitive of the original image, where the background is marked by ones and the rest are zeros, is obtained. The object of relevance extraction is done by filling the primitive with ones, starting with coordinates of the image where the laser is located and ending with pixels declared for background, and then subtracting the initial from the newly created primitive.

Distance to chosen object marked by the laser beam is obtained by prior measurements of Euclidean distance, between pixels on the left and the right

camera, in relation to the real distance of that pixel. By analogy, the measurement of the real screen size at different distances leads to relation from which area of a pixel at a given distance can be easily calculated.
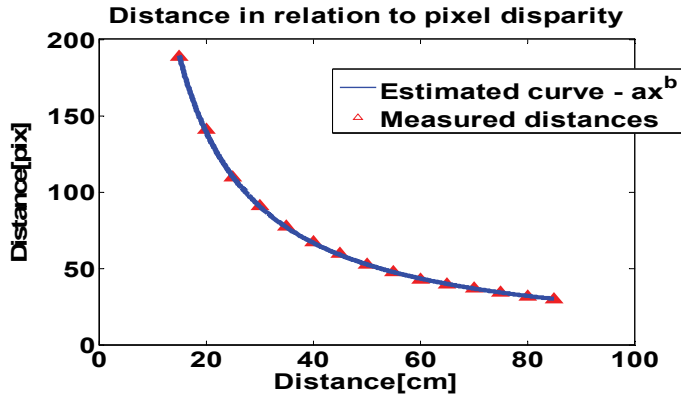


**Fig. 4** – *Distinguishing the distance between selected object and the camera based on changes in the pixel disparity.*

**Table 1**
*Distribution of different approaches to object grasping.*

| Object | Palmar | Lateral | Pinch | Spherical |
|---|---|---|---|---|
| Coffee cup | 0% | 100% | 0% | 0% |
| Mug | 20% | 80% | 0% | 0% |
| Coffee pot | 10% | 90% | 0% | 0% |
| Tetra pak | 100% | 0% | 0% | 0% |
| Bottle | 100% | 0% | 0% | 0% |
| Disc | 0% | 0% | 100% | 0% |
| Comb | 10% | 80% | 10% | 0% |
| Tooth brush | 0% | 80% | 20% | 0% |
| Pencil | 0% | 80% | 20% | 0% |
| Small ball | 0% | 0% | 0% | 100% |
| Mobile | 0% | 10% | 90% | 0% |
| Key | 0% | 30% | 70% | 0% |
| Spoon | 0% | 80% | 20% | 0% |

Object identification starts with the image rotation (so that the main axis of object matches with the horizontal axis) and expansion (so that the object occupies as much of the image, size of 320×240, as possible) of its primitive, and then using correlation, with additional morphological operations, its comparison with those objects from the database whose area does not differ too much from the one calculated for the given object (eg. spoon will not be compared with the ball) is performed. In the correlation process, every pixel to pixel match is „rewarded" with +1 and disparity „punished" with -1 in the sum. The object is declared to be the item from the database with which that correlation is the strongest.

To test this method and to form the database, the closed set of objects, that people use in common daily activities, was chosen. In the database beside the object image primitive is also its respective area, calculated by already described routine, whereas asymmetrical objects have two primitives (e.g. cup when the handle is in viewpoint and when it is concealed). The optimal grasp for each object was obtained by heuristic analysis, to be exact, by recording 10 healthy, 22-25 years old, right-handed subjects grasping the selected objects.

## 3    Results

The way SVS works is illustrated with a few characteristic examples.
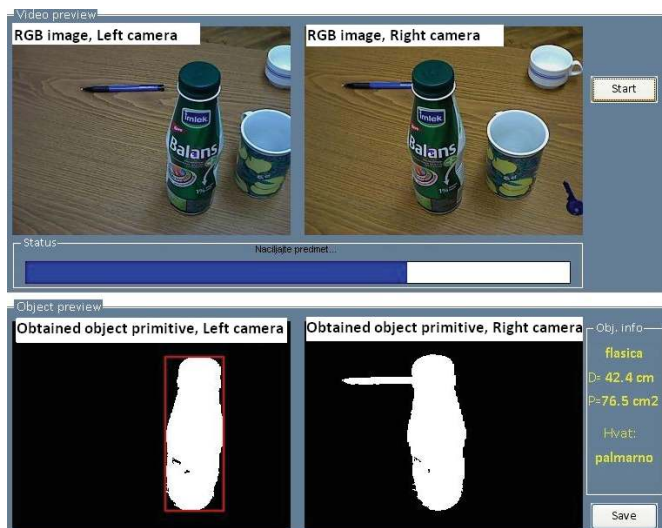


**Fig. 5** – *Objects overlap in the perspective of the right camera resulting with incorrect silhouette in a given primitive. Using discrimination by area (in pixels) that each primitive possesses the algorithm will select the primitive from the other camera (framed), which does not contain overlapping objects.*
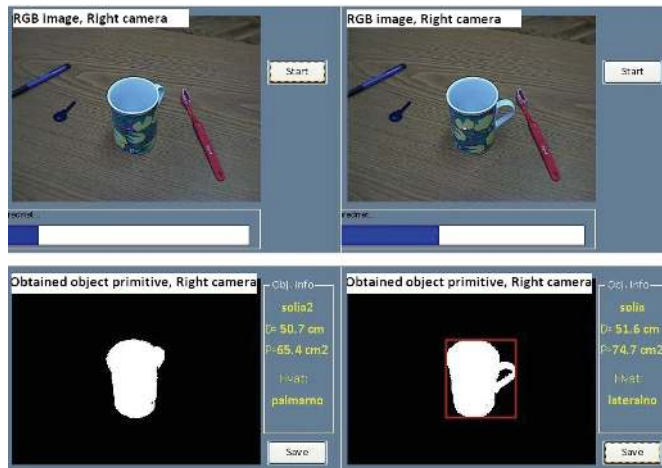
**Fig. 6 –** *By varying the orientation of the object, the algorithm will decide on a different grasp, the one more suitable for that orientation.*
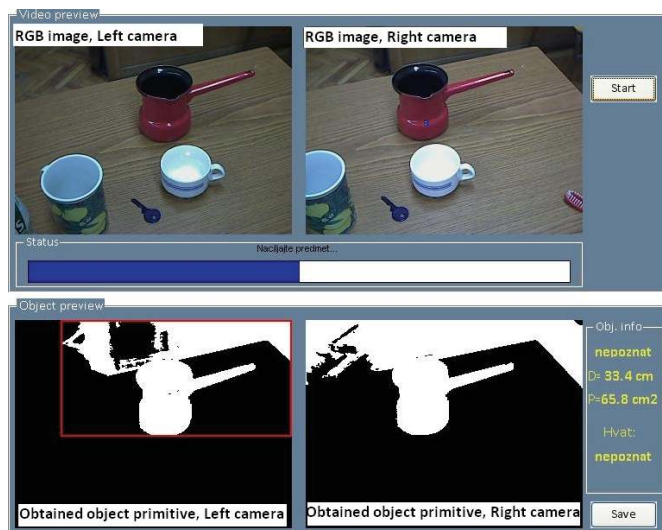


**Fig. 7 –** *Error in detecting laser mark and inadequate segmentation.*

Each camera in SVS observes the object of interest under a different angle. This feature is used in the algorithm when it comes to deciding which of the two extracted primitives represents better candidate for further analysis. SVS testing proved that the best candidate is more often primitive that occupies smaller area in pixels (Fig. 5). This kind of discrimination has also proved successful in other situations, i.e. when the object shadow is more noticeable on one camera

than the other, thus the primitive with minor area will represent the one with a smaller shadow influence.

Depending on the orientation, objects can be approached with different grasping strategy (Fig. 6). In order to enable the algorithm to adjust the decision in this scenario, it is required that the appropriate, different primitives of the same object exist in the database.

Although the laser beam marks the cup (Fig. 7), due to the inferior auto-tuning of „white-balance" in camera drivers, as well as the presence of red surface that in some areas has similar characteristics to the laser, an incorrect point is identified as laser mark. Another obvious error is also noticeable here (the object improperly isolated is attached to the background therefore acting as an entity) as a result of some hypothesis that were made when the algorithm for image segmentation was made.

## 4    Conclusion

Bearing in mind that the test machine is way below average for today's terms, the system has demonstrated exceptional performance above all thanks to good utilization of its resources, a simple but efficient algorithm and the absence of additional sensors. Nevertheless, due to some assumptions and restrictions of the system hardware, it is necessary to comply with following directions to ensure the adequate results:

1. Objects whose surfaces are extremely bright and reflective are not anticipated in the examined scene;
2. All objects should be opaque;
3. Shadows of the objects may exist, but their intensity should not be comparable to the object itself;
4. In the perspective of at least one camera should not be any overlapping objects with the one emphasized;
5. The surface on which the objects are placed should be relatively of consistent coloring (otherwise, some contours in it could be recognized as an object feature).

Testing of image processing algorithm, at different distances (45-75cm) and object orientations, has shown that the percentage of correct grasp choices is greater than 90%. Errors that can occur are usually a result of deformation of the object contour perceived in both primitives when it is moved away from the camera (due to the isometric perspective), which leads to inaccurate area calculation and ultimately unsuitable correlation.

The utilization of SVS potential is not outright, since stereovision along with the depth map calculation offers the possibility of 3D image reconstruction

of an object [6], which could be used as a much reliable source of information. Unfortunately, this approach would require far more expensive and much more accurate system, the one that would consequently be far more demanding in terms of computer resources. It should also be noted that this system was developed for testing new ideas and concepts, therefore, it still remains to examine its applicability in authentic situations. To make that possible, it is essential to fulfill the terms of portability and ease of use, which would be achieved with the embedded system configuration.

## 5    Acknowledgments

## 6    References

[1]    M.B. Popović, D.B. Popović, T. Sinkjær, A. Stefanović, L. Schwirtlich: Clinical Evaluation of Functional Electrical Therapy in Acute Hemiplegic Subjects, Journal of Rehabilitation Research and Development, Vol. 40, No. 5, Sept/Oct. 2003, pp. 443 – 454.

[2]    KF. Kraiss: Advanced Man-machine Interaction: Fundamentals and Implementation, Springer, London, 2006.

[3]    L. Fausett: Fundamentals of Neural Networks, Prentice-Hall, NY, 1994.

[4]    K. Ramnath: A Framework for Robotic Vision-based Grasping Task, Project report, The Robotics Institute.

[5]    Đ. Klisić, M. Kostić, S. Došen, D.B. Popović: Control of Prehension for the Transradial Prosthesis: Natural-like Image Recognition System, Journal of Automatic Control, Vol. 19, No. 1, 2009. pp. 27 – 31.

[6]    R.C. Gonzalez, R.E. Woods, S.L. Eddins: Digital Image Processing using Matlab, Prentice Hall, 2003.

[7]    B. Cyganek, J.P. Siebert: An Introduction to 3D Computer Vision Techniques and Algorithms, John Wiley and Sons, Chippenham, 2009.