# On Facial Features Tracking using an Active Stereo Camera Control Approach

**Gigel Macesanu, Sorin Grigorescu and Florin Moldoveanu**

**Department of Automation, Electronics and Computers, *Transilvania* University of Brasov, Romania,**
**Email: {gigel.macesanu, s.grigorescu, moldof}@unitbv.ro**
**Web address: http://rovis.unitbv.ro**

Abstract: In order to keep track the 3D *positions* and *orientations* (pose) of objects of interest over a large *Field of View* (FOV), the orientations and focal lengths of the stereo camera system must be adjusted, that is, the *pan*, *tilt* and *zoom* (PTZ) of both vision sensors. In this paper the practical implementation of human-robotic system interaction is proposed. Using human facial features, 3D position and a specific control low, the camera orientation and zoom is adapted. The overall system controller is taking into account the dead time inserted by the image processing algorithms.

*Keywords: Facial features detection, active vision, stereo control*

## 1. Introduction

Conventional robot vision systems use stereo cameras with a fixed orientation and zoom, this fact leading to physical limitations in the scene understanding process. In situation in which non stationary objects should be observed and tracked, the fixed cameras are replaced with systems that are able to adapt their viewpoint and zoom along with the poses of the objects. Thus, the visual perceptual capabilities of an autonomous robot are adapted according to the imaged environment, evaluating the camera's extrinsic parameters (i.e. pose and zoom) modifications. Using such an active vision system, a robotic platform can track objects of interest and reconstruct their poses in a virtual 3D space.

In this work an active vision system used for robotic perception is presented. Similar approaches can be found in older papers such as [7] and [8]. Both present first uses of active camera systems as modules able to modify camera parameters, such as: *position*, *orientation*, *focus*, *zoom*, *aperture* and *vergence*. Modern active vision systems use complex structures for camera's zoom adapting (focal length control) such that to maintain a desired distance between camera and an object of interest [2, 3, 11]. In the same context, systems used for active tracking represent an state of the art approach. Such a system should be able to detect multiple objects and to be designed as a humanoid configuration [1, 10].

Facial feature detection is a complex process and all methods used for this task are related to a head detection stage. To complete such a task, methods such as Haar classification or color based segmentation can be used. The first methods use the principle developed by Viola and Jones [9] and can be used for frontal and profile face detection. The second category use particular characteristics of the head, such as the color of the skin [6].

The rest of the paper is organized as follows. In Section II, the features detection algorithms are presented. The active camera control approach is presented in Section III followed by an performance evaluation stage in Section IV. Finally, conclusions are stated in Section V.

## 2. Facial Features Extraction

In this paper the facial futures extraction process supposed to use two classes of algorithms: Haar classifier and color based segmentation. For head detection, two *boosting* classifiers trained for recognizing human faces [9] have been used. The face detection algorithm is applied on each image delivered by the stereo vision system. In order to achieve a confident head detection, a series of face recognition parameters have been tuned, such as: the *scale factor*, used to determine the face scale difference between each search, the *search area size*, used to bound the minimal head search region, or the *head selection confidence*, used to select the best recognized face from an image where multiple faces are present.

The nose region extraction starts from the fact that it has a specific prominence and a specific color. Therefore, a color based segmentation approach, followed by a nose contour identification process has been constructed. The block diagram of the nose segmentation algorithm can be seen in Fig. 1.

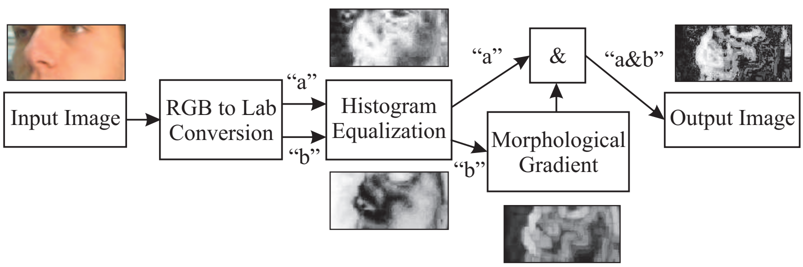



Figure 1: Block diagram of the proposed human nose segmentation algorithm.

The images are acquired RGB format. In order to have a proper segmentation which is, as much as possible, invariant to variable lighting conditions, the RGB image is converted to the L*a*b color space representation. The Lab color space format is composed of a "Luminance" value represented by $L$, and two color channels ($a$ and $b$). For segmenting the nose, a logical **AND** operation between the $a$ color channel and the resulted image is applied. A morphological gradient filter, applied on the $b$ channel image, is used to isolate perimeters of existing binary blobs. On the nose segmented image, all the segmented pixels are grouped into clusters based on their connectivity using a *K-means clustering* approach.

The stereo 2D positions of the nose tip are used with the internal camera parameters (i.e. focal length and baseline) to obtain the nose's tip 3D position [5]. This position is used as input by an active vision system.

## 3. Active camera control approach

The basic block diagram of the proposed architecture can be seen in Fig. 2. The feedback variable is represented by the nose detection system described in the previous section.

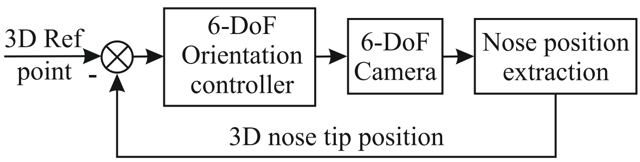



Figure 2: Basic block diagram of the proposed stereo active vision system.

In order to control a time-delay system, such as the one considered in this paper, a different control design technique is required as for the case of traditional linear approaches. This is mainly needed because a time-delay component introduces an infinity of poles in the overall transfer function of the system.

In order to stabilize the plant, a *prediction control* structure can be implemented, such as the *Smith predictor* approach. The core concept of this approach is to move the process's dead time outside the feedback loop of the control system and to determine a controller of a time-delay free system.

In order to design the system overall controller that is capable of stabilizing a system that have its dead time outside the control loop a Smith predictor compensator is obtained, having the following control structure:

$$C(s) = \frac{C_r^*(s)}{1 + C_r^*(s)G_p(s)[1 - e^{-s\tau}]}. \quad (1)$$

where, $C(s)$ represents the system time-delay controller, $C_r^*(s)$ is the controller of the free delay system (from Smith predictor), $G_p$ is the system transfer function and $\tau$ represents the dead time inserted into the system by the image processing algorithms.

Knowing the mathematical model of the open-loop system the design of the compensator is made according to the *poles placement method* [4].

## 4. Performance evaluation

During this stage the active tracking system was tested at different speeds. The human subject follow a predefined path, while the 3D position of its nose was computed. Active tracking results are presented in Fig. 3, where the pan-tilt real values and the position actuator variable are illustrated for a camera velocity of $50°$/sec. The zoom values was A tested for the focal length's translational velocities of 2mm/sec, 5mm/sec and 7mm/sec.

The acquisition system was placed at $1.7$ m above the ground and images was acquired at a resolution of 640x480 pixels.

## 5. Conclusions

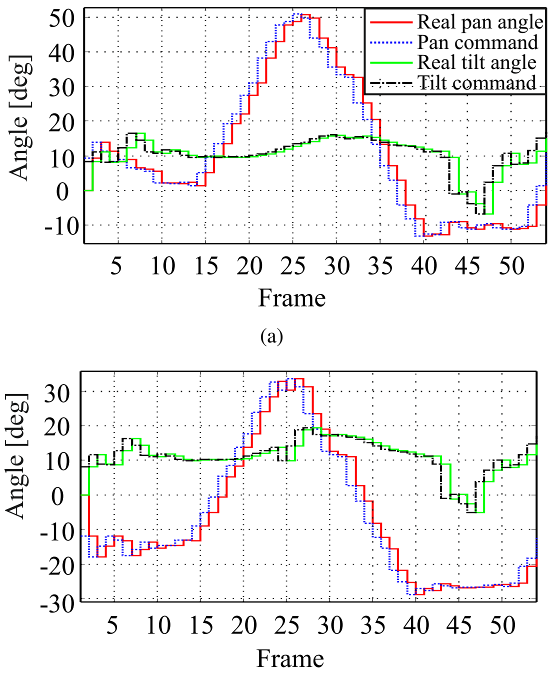
In this paper an active vision system that can be used in an human robotic interaction was presented. The main idea was to determine facial features in both stereo images and to determine the corresponding 3D positions. Using these data and the fact that a proper visualization suppose to have the subject into a central region, the camera position and zoom has to be adapted. In this sense, a time-delay system controller is designed.

## Acknowledgment

This paper is supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contracts number POSDRU/88/1.5/S/59321 and POSDRU/89/1.5/S/59323.

## References

[1] A. Bakhtari and B. Benhabib. An Active Vision System for Multitarget Surveillance in Dynamic Environments. *IEEE Trans. on Systems, Man, and Cybernetics, Part B:*




(a)

(b)

Figure 3: Active nose tip tracking for a velocity of $50^\circ$ /sec.

*Cybernetics*, 37(1):190–198, 2007.

[2] J.P. Barreto, L. Perdigoto, R. Caseiro, and H. Araujo. Active Stereo Tracking of Targets Using Line Scan Cameras. *IEEE Trans. on Robotics*, 26(3):442–457, 2010.

[3] Jeffrey A. Fayman, Oded Sudarsky, Ehud Rivlin, and Michael Rudzsky. Zoom Tracking and its Applications. *Machine Vision and Applications*, 13:25–37, 2001.

[4] G.C. Goodwin, S.F. Graebe, and M.E. Salgado. *Control System Design*. Prentice-Hall, Inc, New Jersey, USA, 2001.

[5] S.M. Grigorescu, G Macesanu, T.T.. Cocias, D. Puiu, and F. Moldoveanu. Robust Camera Pose and Scene Structure Analysis for Service Robotics. *Robotics and Autonomous Systems*, 59(11):899–909, 2011.

[6] G Macesanu, S Grigorescu, J. F. Ferreira, J. Dias, and F. Moldoveanu. Real Time Facial Features Tracking Using an Active Vision System. In *Proc. of the 13th Inter. Conf. on Optimization of Electrical and Electronic Equipment*, pages 1493–1498, Brasov, Romania, 2012.

[7] Rajeev Sharma. Role of Active Vision in Optimizing Visual Feedback for Robot Control. In *The Confluence of Vision and Control*, volume 237 of *Lecture Notes in Control and Information Sciences*, pages 24–40. Springer, Berlin/Heidelberg, 1998.

[8] Michael J. Swain and Markus A. Stricker. Promising Directions in Active Vision. *Inter. Journal of Computer Vision*, 11:109–126, 1993.

[9] P. Viola and M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. In *Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, Kauai, USA, 2001.

[10] Kai Welke, Tamim Asfour, and Rüdiger Dillmann. Active Multi-view Object Search on a Humanoid Head. In *Proc. of the 2009 IEEE Inter. Conf. on Robotics and Automation*, pages 2041–2047, Piscataway, NJ, USA, 2009.

[11] Yi Yao, Besma Abidi, and Mongi Abidi. 3D Target Scale Estimation and Target Feature Separation for Size Preserving Tracking in PTZ Video. *Inter. Journal of Computer Vision*, 82(3):244–263, May 2009.