A COLOR CONSTANCY APPROACH FOR ILLUMINATION INVARIANT COLOR TARGET TRACKING

Geert De Cubber **Hichem Sahli** **Hong Ping** **Eric Colon**

Vrije Universiteit Brussel
Pleinlaan 2
B-1050 Brussels-Belgium
{gdcubber, hsahli}@etro.vub.ac.be

Royal Military Academy
Avenue de la Renaissance 30
B-1000 Brussels-Belgium
eric.colon@rma.ac.be

Abstract

Many robotic agents use color vision to retrieve quality information about the environment. In this work, we present a visual servoing technique, where vision is the primary sensing modality and sensing is based upon the analysis of the perceived visual information. We describe how colored targets can be identified and how their position and motion can be estimated quickly and reliably. The visual servoing procedure is essentially a four-stage process, with color target identification, motion parameter estimation, target tracking and target position estimation. These individual parts add up to a global vision system enabling precise positioning for a demining robot.

Keywords : visual servoing, color constancy, robot positioning

1. Introduction

The research work presented in this article fits in a global research effort to develop intelligent humanitarian demining robots. One of the tasks set up for these robots is to build precise maps of the inspected terrain with an indication of all the suspicious points where mines could be located. To do this, the robot must be equipped with a very performant localization system. However, classical absolute positioning sensors like GPS do not deliver the demanded precision in some cases. In these cases, the robot positioning problem can be solved using a single external camera as we present here. The use of computer vision for solving detection, tracking and positioning problems is a major research area in robotics [16][17][18][19][20]. Two approaches have been considered: (1) a fixed camera configuration, that is, the camera is fixed at a certain point in a general frame (the world coordinate system) [18], and (2) the eye-in-hand configuration, where the camera is fixed on the end-effector or mounted on a mobile robot [19][20]. For both cases, monocular [16][19][21] or stereo [18][22][23][24] vision systems have been investigated.

In this work, we present a color target tracking algorithm aimed at robot localization in varying illumination conditions. In our approach a colored target is put on the top of the robot and a fixed camera is used to detect and track the target. The 3D robot position can be estimated knowing the camera parameters after an analysis of the camera image. The general setup of this approach is sketched on Figure 1:

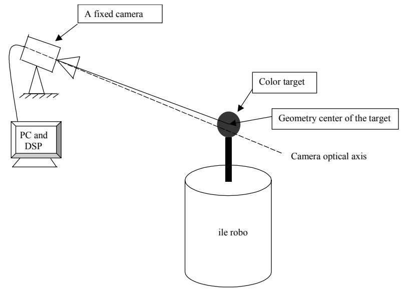


Figure 1: Setup for 3D robot localization

The first stage in the target tracking algorithm is the color target identification process, where a colored object that moves independently of the observer has to be found. Common image segmentation and object recognition were developed in the past to deal with this task. Yet, all of these algorithms have big problems as soon as the protected lab environment is left and tests are carried out in an outdoor environment where harsh and ever-changing illumination conditions cause great difficulties for the image-processing task, as a change in illumination will also change the perceived colors – or more generally the perceived image – of the environment. Numerous attempts have been made to solve this so-called “color constancy” problem and promising results have been shown before, as summarized in [8]. E. H. Land and J. J. McCaan were the first to tackle the problem with their retinex theory [7]. Others relied on finite-dimensional linear models [9][10][11][12][13], while also neural nets have been proposed as a solving technique [14]. However, these techniques commonly require hours of calculation time to process one non-synthetic image, which makes them totally unfit for real-time and real-world vision tasks, as is the case in the field of robotics. Here, we propose a color constancy technique used

for real-time target identification under varying illumination conditions.
Once the target object is identified, it can be tracked. Target-tracking refers to a method that enables a visual system to locate the target in its field of view using consecutive images. In our system the target is mounted on a mobile robot. The target tracking problem is regarded here as a camera control problem. We use the parameters estimated during the target detection step as well as the camera calibration parameters to control the camera's motion (pan and tilt) and try to keep the target center coincident to the image center. Moreover, the distance to the target is estimated using a proportional scaling, so that a precise positioning of this target object, the robot system as a whole or just the end-effector can be performed.
The rest of this paper is organized as follows: first, we will explain the used color constancy approach to achieve an illumination invariant color classification. Next, we show the working principles of the actual target tracking process. In paragraph 4, we will discuss the target localization procedure, which is most important for robotic applications. Finally, we present some results and conclusions.

2. Illumination invariant color classification

Our approach is directly based upon the physical characteristics of color reflection. The main problem for the correct interpretation of a camera image is that the measured intensities are function of a huge number of parameters and most of them cannot be retrieved in any possible way due to their strong interconnectivity. The color of an object in the image is therefore more an appearance than a real material property. Nevertheless, color can be used to identify objects as long as the parameters which influence the formation of the perceived color are taken into account. To do this, we make use of the dichromatic reflection model, which was first introduced by Shafer in [1]:

$$\rho_c = k_b(\bar{n},\bar{i},\bar{v}).\int_\lambda e(\lambda).f_c(\lambda).r_b(\lambda).d\lambda + k_s(\bar{n},\bar{i},\bar{v}).\int_\lambda e(\lambda).f_c(\lambda).r_s(\lambda).d\lambda \quad (1)$$

With:

- ρ_c: the measured intensity of channel c
- $e(\lambda)$: the normalized light spectrum
- $f_c(\lambda)$: the c^{th} channel sensor response function
- $r(\lambda)$: the surface reflectance function
- k_b: attenuation factor for the body reflectance
- k_s: surface reflectance attenuation factor
- $\bar{n}$: the normal to the surface patch
- $\bar{i}$: the direction of the illumination
- $\bar{v}$: the viewing direction

Among the different color spaces, our choice went out to the l1-l2-l3-space, a color space which was originally introduced by Gevers and Smeulders in [6] as a space that uniquely determines the direction of the triangular color in the RGB space. It poses an attractive alternative to the HSI space due to its computational simplicity. The space can be formulated as follows:

$$\begin{aligned} l_1 &= \frac{|R-G|}{|R-G|+|R-B|+|G-B|} \\ l_2 &= \frac{|R-B|}{|R-G|+|R-B|+|G-B|} \\ l_3 &= \frac{|G-B|}{|R-G|+|R-B|+|G-B|} \end{aligned} \quad (2)$$

In [15], Gevers and Smeulders prove that according to the dichromatic reflection theory, this space is invariant to highlights, viewing direction, surface orientation and illumination direction. This means that we can work with a simplified form of equation 1:

$$H_{l1-l2-l3}(x,t) = \int_\lambda e(\lambda,t).f_c(\lambda).r_b(\lambda,x).d\lambda \quad (3)$$

Equation 3 can be discretized by sampling over a number of wavelength bands. We chose to use a finite dimensional linear model with a limited amount of parameters and using 10 basis functions:

$$\begin{aligned} e(\lambda,t) &= B_e.q_e \\ r_b(\lambda,x) &= B_r.q_r \end{aligned} \quad (4)$$

The columns of the N x N_e B_e matrix and those of the N x N_r B_r matrix represent the basis functions for the light spectrum and the reflectance spectrum respectively. The N_e element q_e vector and the N_r element q_r describe respectively the illuminant and the body reflectance spectrum.
The problem with this representation is that the basis and sensor sensitivity functions are not well known. To avoid this difficulty, we use an approach similar to the one described in [4], which introduced a lighting and reflectance matrix, parameterized using 4 x N_e variables in a manner that is independent of basis functions and sensitivity functions. This leads to a general equation:

$$h^T = q_e^T.\sigma \quad (5)$$

With:

- $h^T = (h_1\ h_2\ h_3)$
- $\sigma = (\sigma_1\ \sigma_2\ \sigma_3)$ a N_e x 3 matrix holding all the reflection characteristics for a specific image point

In a learning phase, the algorithm learns the reflection characteristics of the object to be tracked. Small patches of images are accumulated over time while the material in question is subjected to a varying illumination. All intensity measurements are combined in an f x 3.p color measurement matrix H, while p is the number of pixels in the scene patch and f the number of frames sampled. If we sample for long enough, then eventually f will grow larger than p and the light spectrum matrix Q and the reflection characteristics matrix S can be recovered by applying singular value decomposition on H. Thus, all factors in equation 6 can be calculated:

$$H = Q.S \quad (6)$$

At this moment, the light spectrum distribution if the illuminant l is known $p(q_e|l)$ can be calculated. This can be done because Q is independent of the material. We use an Expectation Maximization (EM) clustering method [3] to derive the reflection distributions. This algorithm applies multivariate Gaussian mixture modeling with an unknown number of mixture components, so the number of clusters isn't fixed on beforehand, which makes the classification very flexible. The result of this calculation is an N_{LS} *x* N_e light spectrum matrix L and an N_e *x 3* reflectance spectrum matrix R, with N_{LS} the number of illuminant spectra distinguished by the EM algorithm.

Now that we have estimates of the reflectance spectrum of the target object and now that we've obtained illuminant spectra corresponding to different lighting conditions, we want to correctly classify newly presented pixels as belonging to the target object or not, while keeping track of newly arising lighting conditions. We present a Bayesian solution to solve these problems. New scene properties are brought into the model based upon the Maximum A Posteriori (MAP) estimate of these parameters given the color measurements. When applying this classification, we search for the conditions that maximize $p(o=o_{Target}, l, q_e, \sigma|h)$ for any values of the lighting condition l, the illuminant spectrum q_e and the reflectance spectrum of the target object σ, given the color measurement triplet h. The equation we want to solve is:

$$[\hat{o},\hat{l},\widehat{q_e}] = \underset{[l,q_e]}{argmax}\, p(o,l,q_e,\sigma\,|\,\hat{h}) \quad (7)$$

Using Bayes' rule, it can be shown that:

$$p(o,l,q_e,\sigma\,|\,\hat{h}) \propto p(\hat{h}\,|\,q_e,\sigma).p(q_e\,|\,l).p(l).p(o) \quad (8)$$

The pixel classification procedure calculates the probability $p(o,l,q_e,\sigma\,|\,\hat{h})$ for each pixel and labels the pixel as belonging to the target object or not based upon the result. Using this theorem, the pixel classification is no longer performed directly based upon the pixels color value, as is classically done, but based upon the derived reflection characteristics. This method makes the detection process very robust and recovers the target shape, which enables the estimation of the target image size.

During the actual tracking phase, the illumination model is continually updated using Bayesian reasoning. In this model updating stage, estimates for new lighting conditions and their corresponding illuminant spectra are calculated. It is this procedure that ensures the adaptive nature of the pixel classification process within the general target-tracking program. The philosophy of this procedure is that we take a small patch from the target object, try to recover the spectrum of the illuminant shining on this part of the target object and update our model if necessary. This algorithm doesn't need to run completely at every iteration, since there won't be a new illumination condition with every new frame and only noteworthy changes in illumination will result in the model being updated, so there are a lot of exit conditions built into the process. The calculation of the new illumination condition itself can happen very rapidly, since we already know the reflectance spectrum matrix. After acquiring a nominal color triplet measurement h_N, we can write:

$$q_e(N_{new}) = h_N.R^{-1} \quad (9)$$

With N_{new} the index of the rarest illumination condition within the L matrix, which will thus be replaced by the new lighting condition.

After the pixel classification process, the target object will never be completely recognized, there will always be outlier pixels. This is shown on figure 6 in the results paragraph. To solve this problem, we use morphology filtering. During the color detection we create a corresponding binary image. For this binary image we use morphology filtering to do image segmentation. A square mask of 5 by 5 pixels is used as the structuring element in our application. Figures 7 and 8 show processed results. The white color shows the detected region after image segmentation and all the pixels in this region are considered as the detected pixel. The region is connected and represents the target shape very well.

3. **Target tracking**

In fact, the target-tracking problem can be regarded as a visual servoing problem. In our system the target is mounted on a mobile robot. A calibrated camera fixed at the origin of the world frame is controlled through its pan (α) and tilt (β) angles to bring the target image center onto the image plane center. The camera zoom is also controlled to maintain a high signal-to-noise level. Figure 2 shows the camera control parameters which were defined.

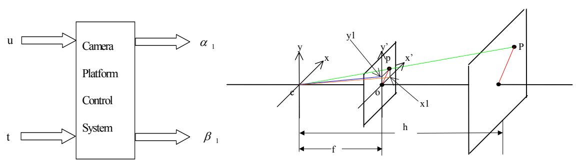


Figure 2: Camera control parameters

The proposed camera control method consists of two parallel processes: one process controls the pan/tilt camera platform in order to track the target; the other process uses a predictor to track the target in the image plane. Due to the fact that the robot moves with an unknown model, the servomotor-camera-target system is a time variant system. The target motion model has to be identified in real time. Motion parameter estimation uses the visual input to estimate the dynamic properties of the target object. Initially, we set up a computational model based upon the theoretical aspects of the different components of the visual servoing system. This step comes basically down to perform a parameter identification for this theoretical model. In order to meet the system dynamic characteristic requirements we developed a two-phase control strategy. The first one is an initialization phase, in which the motion dynamics are estimated and during which the target is tracked with a PI regulator. The more interesting second phase control consists of a feedback control strategy shown in Figure 3. The plant is modeled as a dynamic system as shown in Figure 4.

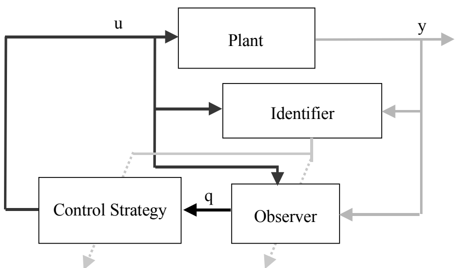


Figure 3: Feedback control strategy

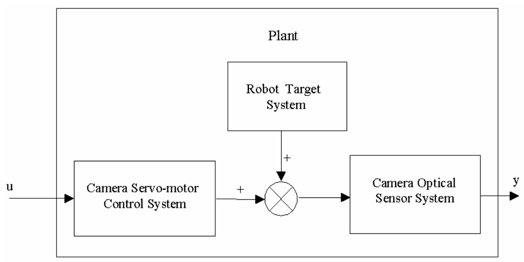


Figure 4: Model of the plant

In our implementation a second order difference model is considered. The system functions are

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_0(k) & -a_1(k) \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} b_0(k) & b_1(k) \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \qquad (10)$$

Where:

- (x_1, x_2) is the state vector corresponding to the camera angles and angular velocity.
- (a_0, a_1, b_0, b_1) are the system parameters to be estimated.

These parameters are estimated using LSM method from a set of input image frames and camera control parameters.

The feedback control strategy is implemented with system state vector estimation using Kalman filtering. The detail of the observer-based full-state-feedback control system configuration is shown in Figure 5.

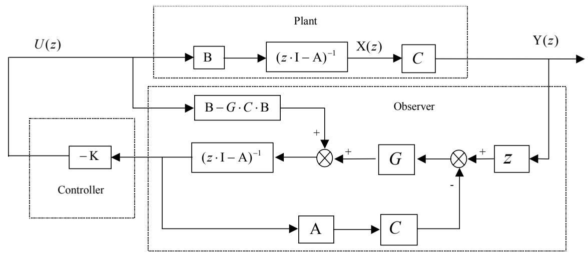


Figure 5: The Observer Based Full State Feedback Control System

With:

A: the plant system matrix given by $[a_0, a_1]$

B: the plant input matrix [0,1]

C: the plant output matrix given by $[b_0, b_1]$

G: the Kalman filter gain matrix

K: the control gain matrix defined as the difference between the required and identified system parameters of characteristic functions

Using this Kalman-filter-based camera control strategy, it is possible to achieve smooth and stable camera movement, even when the target object undergoes shaky movements. The ability to estimate the target motion and to perform very rapid processing makes window-tracking possible. The proposed window tracking method reduces the image processing time and increases the signal-to-noise ratio significantly.

4. Target size & distance estimation

The visual servoing system presented here involves a method for estimating the target position, i.e. the quantitative description of where the target is with respect to the observer's view. For our application, the similarity of the target shape and its projected image is used to estimate the camera/target distance.

The origin of world frame is set at the center of the camera. The camera platform is kept horizontal. Then, the position of the target can be described by 3 parameters: the horizontal angle, the vertical angle and the distance between camera and target. Angles are calculated using the pose of the camera and the orientation angles of the target image in the camera coordinate system. The distance between camera and target is estimated by simple similar triangle relationship of the real target size, the detected target image size and the effective camera focal length. The size of the target is estimated using circle and ellipse fitting procedures to more accurately measure the radius of the target object in the image plane. For the ellipse fitting, we used a very fast algorithm described in [5], whereas the circle fitting procedure is a much slower, but slightly more precise homemade algorithm.

5. Results

Showing the results of the presented color target tracking approach and its illumination invariant features is kind of hard if the use of color is not allowed. Figure 6 shows a result of the pixel classification procedure. Remember this is still before the morphology filtering.

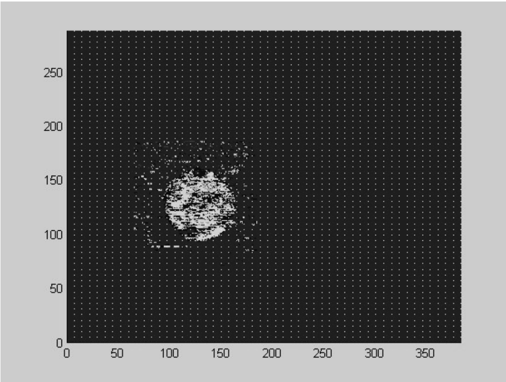


Figure 6: Pixel classification

Figure 7 shows one scene of an outdoors static trial. The target detection function works well (detected pixels are painted white).

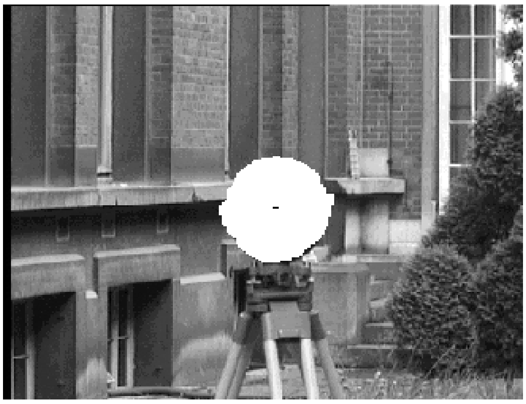


Figure 7: Outdoor trial

Figure 8 shows an indoor scene during a trial on varying illumination. You can see that the image is very dark, because the lights were turned off, yet the target object is still detected almost entirely.

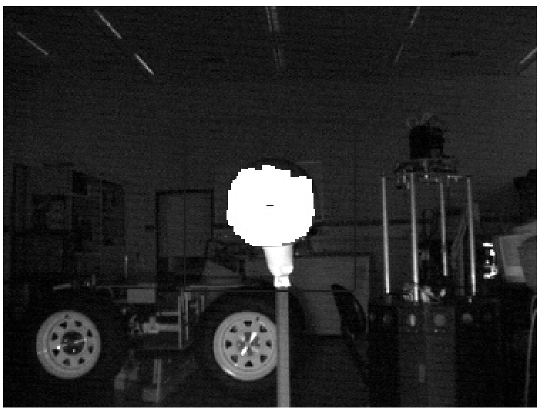
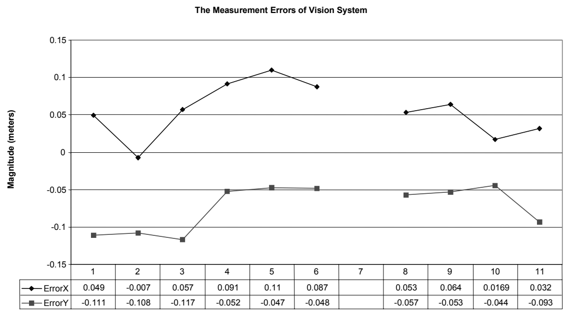


Figure 8: Extreme dark illumination conditions

Figure 9 shows the distance measurement errors during indoor tests. As you can see, the errors do not exceed 0,12 meters.



	1	2	3	4	5	6	7	8	9	10	11
ErrorX	0.049	-0.007	0.057	0.091	0.11	0.087		0.053	0.064	0.0169	0.032
ErrorY	-0.111	-0.108	-0.117	-0.052	-0.047	-0.048		-0.057	-0.053	-0.044	-0.093

Figure 9: Distance measurement error

Concerning the real-time capabilities, the classification algorithm takes about 60 ms to complete on a PC equipped with an Intel PIV 1.7GHz processor. When adding the 30 ms needed for morphology filtering and 10 ms for other tasks, we see the target-tracking program running at about 10 fps. This is adequate for every-day target tracking tasks, but not for high performance applications, so we still might want to improve the implementation a bit. The most processor-intensive process here is the management of the illumination maps.

6. Conclusions

We have shown a powerful set of algorithms, which were combined to form a universally useable system for automated target detection, tracking and position estimation, using a single and fairly simple pan/tilt camera. The Bayesian-based color constancy approach which was used ensures that this system can keep working, even in harsh illumination conditions. This research was specifically aimed at applicability in the field of robotics and due to its general structure it can also be used for a very wide range of applications. To conclude, we show a picture of the demining robot used for testing.

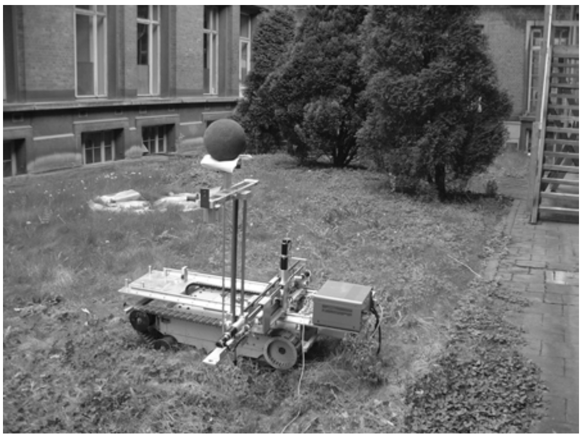


Figure 10: Demining robot

7. References

[1] S.A. Shafer, “Using color to separate reflection components” (1985). **COLOR research and application**, Vol. 10, No. 4, pp. 210-218.

[2] P. Hong, H. Sahli, E. Colon, Y. Baudoin (2001). “Visual Servoing for Robot Navigation”, **Third International Conference on Climbing and Walking Robots: Clawar 2001**, pp. 255 – 264, Karlsruhe, Germany.

[3] K. Zhang, L. Van Kempen, H. Sahli (2000). “Finding maximum likelihood mixture density parameters via Expectation Maximization and its application to mine detection”, Vrije Universiteit Brussel.

[4] Y. Tsin, R. Collins, V. Ramesh, and T. Kanade (2001). "Bayesian Color Constancy for Outdoor Object Recognition", **IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01)**.

[5] Markus Vincze, “Robust tracking of ellipses at frame rate” (2001). **The Journal of the Pattern Recognition Society**, vol. 34, No. 2, pp. 487-498.

[6] T. Gevers, A.W.M. Smeulders (1999). “Color-based object recognition,” **Pattern Recognition**, Vol. 32, pp. 453-464.

[7] E. H. Land and J. J. McCaan (1971). “Lightness and the retinex theory”, **Journal of the Optical Society of America**, 61:1–11.

[8] K. Barnard, “Practical Colour Constancy” (1999). PhD Thesis, Simon Frasier University, School of Computing.

[9] M. H. Brill and G. West, “Chromatic adaptation and color constancy: A possible dichotomy” (1986). **COLOR Research and Applications**, 11:196–204.

[10] M. D’Zmura and P. Lennie, “Mechanisms of color constancy” (1986). **Journal of the Optical Society of America** A, 3(10):1662–1672.

[11] B. V. Funt, M. S. Drew, and J. Ho, “Color constancy from mutual reflection” (1991). **International Journal of Computer Vision**, 6(1):5–24.

[12] J. Ho, B. V. Funt, and M. S. Drew, “Separating a color signal into illumination and surface spectral components” (1990). **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 12(10):966–977.

[13] L. T. Maloney and B. Wandell, “Color constancy: A method for recovering surface spectral reflectance” (1986). **Journal of the Optical Society of America** A, 3(1):29–33.

[14] B. Funt, V. Cardei, and K. Barnard, “Learning Color Constancy” (1996). **Proc. IS&T/SID Fourth Color Imaging Conference: Color Science, Systems and Applications**, pp. 58-60.

[15] T. Gevers, H. Stokman “Reflectance Based Edge Classification” (1999). **Vision Interface**, Trois-Rivières, Canada.

[16] A. Maruyama, M. Fujita, “Visual Feedback Control of Rigid Body Motion Base on Dissipation Theoretical Approach” (1999). **Proceedings of the 38th Conference on Decision & Control**, pp. 4161-4166.

[17] E. Zergeroglu et al., “Robust Visual-Servo Control of Robot Manipulators in the Presence of Uncertainty” (1999). **Proceedings of the 38th Conference on Decision & Control,** pp. 4137-4142.

[18] Ming Lei and Bijoy K. Ghosh, “Visually-Guided Robotic Motion Tracking” (1992). **Proceedings of Thirtieth Annual Allerton Conference on Communication, Control and Computing**, pp. 712-721.

[19] Billibon H. Yoshimi and Peter K. Allen, “Active, Uncalibrated Visual Servoing” (1994). **IEEE International Conference on Robotics and Automation**, pp. 156-161.

[20] Hong Zhang and James P. Ostrowski, “Visual Servoing with Dynamics: Control of an Unmanned Blimp” (1999). **IEEE International Conference on Robotics and Automation**, pp. 618-623.

[21] Rafael Kelly, “Robust Asymptotically Stable Visual Servoing of Planar Robots” (1996). **IEEE Transactions on Robotics and Automation**, Vol. 12, No. 5, pp. 759-766.

[22] Nicholas Ayache and Francis Lustman, “Trinocular Stereo Vision for Robotics” (1991). **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 13, No. 1, pp. 73-85.

[23] Philip F. McLauchlan et al., “The Pipe-group Architecture for Real-Time Active Vision” (1997). **Real-Time Imaging**, Vol. 3, pp. 319-330.

[24] James L. Crowley, “Dynamic Calibration of an Active Stereo Head” (1993). **ICCV93**, Berlin.