

Global Visual Features based on Random Process

Application to Visual Servoing

Laroussi Hammouda¹, Khaled Kaaniche^{1,2}, Hassen Mekki^{1,2} and Mohamed Chtourou¹

¹*Intelligent Control Design and Optimization of Complex Systems, University of Sfax, Sfax, Tunisia*

²*National School of Engineering of Sousse, University of Sousse, Sousse, Tunisia*

Keywords: Visual Servoing, Global Visual Features, Mobile Robot.

Abstract: This paper presents new global visual features: random distribution of limited set of pixels luminance. Our approach aims to improve the real-time performance of visual servoing applications. In fact, using these new features, we reduce the computation time of the visual servoing scheme. Our method is based on a random process which ensures efficient and fast convergence of the robot. The use of our new features removes the matching and tracking process. Experimental results are presented to validate our approach.

1 INTRODUCTION

Computer vision is progressively playing more important role in service robotic applications. In fact, the movement of a robot equipped with a camera can be controlled from its visual perception using visual servoing technique. The aim of the visual servoing is to control a robotic system using visual features acquired by a visual sensor (Chaumette and Hutchinson, 2008). Indeed, the control law is designed to move a robot so that the current visual features s , acquired from the current pose r , will reach the desired features s^* acquired from the desired pose r^* , leading to a correct realization of the task.

The control principle is thus to minimize the error $e = s - s^*$ where s is a vector containing the current values of the chosen visual information, and s^* its desired values. The basic step in image-based visual servoing is to determine the adequate set of visual features to be extracted from the image and used in the control scheme in order to obtain an optimal behavior of the robot.

In the literature several works were concerned with simple objects and the features used as input of the control scheme were generally geometric: coordinates of points, edges or straight lines (Espiau and al., 1992), (Chaumette and Hutchinson, 2007).

These geometric features have always to be tracked and matched over frames. This process has proved to be a difficult step in any visual servoing scheme. Therefore, in the last decade, the

researchers are focused on the use of global visual features. In fact, in (Collewet and al., 2008) the visual features considered are the luminance of all image pixels and the control law is based on the minimization of the error which is the difference between the current and the desired image.

Others works are interested in the application of image moments in visual servoing, like in (Chaumette and Hutchinson, 2003) where the authors propose a new visual servoing scheme based on a set of moment invariants. The use of these moments ensures an exponential decoupled decrease for the visual features and for the components of the camera velocity. However this approach is restricted to binary images. It gives good results except when the object is contrasted with respect to its environment.

In (Dame and Marchand, 2009), the authors present a new criterion for visual servoing: the mutual information between the current and the desired image. The idea consists in maximizing the information shared by the two images. This approach has proved to be robust to occlusions and to very important light variations. Nevertheless, the computation time of this method is relatively high.

The work of (Marchand and Collewet, 2010) proposes the image gradient as visual feature for visual servoing tasks. This approach suffers from a small cone of convergence. Indeed, using this visual feature, the robotic system diverges in the case of large initial displacement. Another visual servoing approach which removes the necessity of features tracking and matching step has been proposed in

(Abdul Hafez and al., 2008). This method models the image features as a mixture of Gaussian in the current and in the desired image. But, using this approach, an image processing step is always required to extract the visual features.

The contribution of this paper consists in the definition of new global visual features: random distribution of limited set of pixels luminance. Our features improve the computation time of visual servoing scheme and avoid matching and tracking step. We illustrate in this work an experimental analysis of the robotic system behavior in the case of visual servoing task based on our new approach.

This paper is organized as follows: Section 2 illustrates our new visual features and the corresponding interaction matrix. Section 3 recalls the optimization method used in the building of the control law. Finally, experimental results are presented in section 4.

2 RANDOM DISTRIBUTION OF LIMITED SET OF PIXELS LUMINANCE AS VISUAL FEATURES

The use of the whole image luminance as global visual features for visual servoing tasks, as in (Collewet and Marchand, 2011), requires too high computation time. Indeed, the big size of the interaction matrix related to the luminance of all image pixels leads to a very slow convergence of the robotic system.

Therefore, we propose in this paper a new visual feature which is more efficient in terms of computation time and doesn't require any matching nor tracking step.

In fact, instead of using the luminance of all image points, we work just with the luminance of a random distribution of a limited set of image points (n pixels). Thus, the visual features, at a position r of the robot, are:

$$s_i(r) = E_1^i(r) \quad (1)$$

with $E_1^i(r)$ is the luminance of random set of image pixels taken at frame i .

$$E_1^i(r) = (I_1^i, I_2^i, I_3^i, \dots, I_n^i) \quad (2)$$

where I_k^i is the luminance of the pixel k taken randomly at the frame i .

For each new frame, we get a new random set of image pixels. Thus, the desired and the current

visual features will continuously change along the visual servoing scheme. In that case, the error e will be:

$$e_i = E_1^i(r) - E_1^i(r^*) \quad (3)$$

where $E_1^i(r)$ represent the current visual features and $E_1^i(r^*)$ the desired ones at the frame i .

Consequently, in our method, the error used in the building of the control law is variable, it changes at each frame. This change is like a kind of mutation. Convergence to global minimum is then guaranteed.

The choice of n is based on the image histogram. We take n equal to the maximum value of the current image histogram. We can then avoid the fact that the n pixels randomly chosen will have the same luminance. Hence, we guarantee the good luminance representation of the image. We note p_l the probability that the n pixels will have the same luminance. It is given by:

$$p_l = \frac{C_n^n}{C_N^n} = \frac{1}{\frac{N!}{n!(N-n)!}} \quad (4)$$

where n is the number of pixels deduced from the image histogram and chosen as visual features and N is the number of all image pixels. This probability is null (see Table 1).

Since the number n depends on the histogram of the current image, it slightly changes during the visual servoing scheme. Let us point that n is always very small compared to the total number of image pixels (in our case 320×240). We note that the more the image is textured, the smaller n is.

Figure 1 shows an example of image, the luminance of all its pixels form the ancient global visual features.

The histogram of this image is illustrated on Figure 2. In our approach, instead of using all image pixels, we take randomly n pixels as global visual features, with n is the maximum value of this histogram (in this example $n = 2452$ which is 3.1% of all image pixels).

After ensuring that the n pixels are good representatives of the image luminance, we can confirm that these n pixels randomly chosen will be well distributed in the image and not concentrated in one particular zone. For that, we compute the probability that the n pixels will be all in one zone z . This probability is given by:

$$p_s = \frac{C_z^n}{C_N^n} = \frac{\frac{z!}{n!(z-n)!}}{\frac{N!}{n!(N-n)!}} \quad (5)$$

with z is the number of pixels in a compact zone of the image.



Figure 1: Ancient visual features: The whole image luminance.

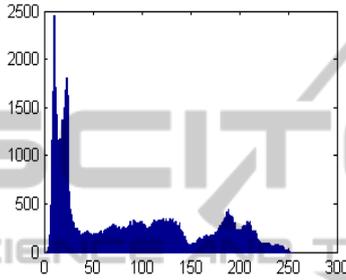


Figure 2: Image histogram (essential for the choice of n).

In our work, we take z as the half of all image pixels (Beyond this value of z we assume that good image representation is ensured).

The probability p_s is equal to zero (see Table 1). This proves that the n pixels chosen as visual features will always ensure good spatial representation of the scene.

We present in Table 1 the histograms and the probabilities (p_l and p_s) related to different images.

The visual servoing is based on the relationship between the robot motion and the consequent change on the visual features. This relationship is expressed by the well known equation (Chaumette and Hutchinson, 2006):

$$\dot{s} = L_s v \quad (6)$$

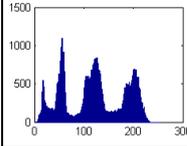
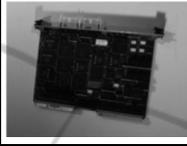
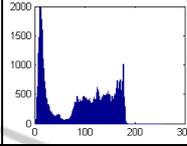
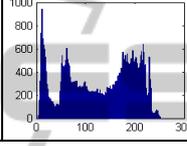
where L_s is the interaction matrix that links the time variation of s to the robot instantaneous velocity v (Chaumette and Hutchinson, 2008).

So, after identification of the visual features, the control law requires the determination of this matrix which is at the center of the development of any visual servoing scheme. In our case, we look for the interaction matrix related to the luminance of a pixel x in the image.

The computation of this matrix is based on the optical flow constraint equation (OFCE) which is a hypothesis that assumes the temporal constancy of the luminance for a physical point between two

successive images (Marchand, 2007).

Table 1: Examples of images with the corresponding histograms and probabilities.

Image	Histogram	n	p_l	p_s
		1089	0	0
		1979	0	0
		940	0	0

If a point x of the image realizes a displacement dx in the time interval dt , according to the previous hypothesis we have:

$$I(x + dx, t + dt) = I(x, t) \quad (7)$$

After development of this equation we get:

$$\nabla I^T \dot{x} + \dot{I} = 0 \quad (8)$$

where $\dot{I} = \frac{\partial I(x)}{\partial t}$ and ∇I is the spatial gradient of x .

We know that: $\dot{x} = L_x v_c$ (9)

where L_x is the interaction matrix that relates the temporal variation of x to the control law.

Using (8) and (9) we obtain:

$$\dot{I} = -\nabla I^T L_x v_c \quad (10)$$

So the interaction matrix that relates the temporal variation of the luminosity $I(x)$ to the control law v_c is:

$$L_{I(x)} = -\nabla I^T L_x \quad (11)$$

In this case, we can write the interaction matrix $L_{I(x)}$ in terms of the interaction matrices L_x and L_y related to the coordinates of $x = (x, y)$ and we obtain:

$$L_{I(x)} = -(\nabla I_x L_x + \nabla I_y L_y) \quad (12)$$

with ∇I_x et ∇I_y are the components along x and y of $\nabla I(x)$.

In the case of a mobile robotic system, we take into account just the components of L_x that correspond to three degrees of freedom: Translation

along x axis, translation along z axis and rotation around y axis. Therefore, we have:

$$L_x = \begin{pmatrix} -\frac{1}{z} & \frac{x}{z} & -(1+x^2) \end{pmatrix} \quad (13)$$

$$L_y = \begin{pmatrix} 0 & \frac{y}{z} & -xy \end{pmatrix} \quad (14)$$

where z is the depth of the point x relative to the camera frame.

We get the interaction matrix related to our new features ($L_{E_i^i}$) by combining the interaction matrices related to the n pixels randomly chosen.

Thus, the size of the interaction matrix related to our visual features ($L_{E_i^i}$) is very small compared to the size of the interaction matrix related to the whole image luminance.

3 THE CONTROL LAW GENERATION

In our work we use a global photometric visual features. In this case most of classical control laws fail. Therefore, we have interest in turning the visual servoing scheme into an optimization problem to get the convergence of the mobile robot to its desired pose (Abdul Hafez and Jawahar, 2006), (Abdul Hafez and Jawahar 2007). In fact, the aim of the control law will be the minimization of a cost function which is the following:

$$C(r) = (s(r) - s(r^*))^T (s(r) - s(r^*)) \quad (15)$$

where $s(r)$ are the current visual features ($E_i^i(r)$) and $s(r^*)$ are the desired ones ($E_i^i(r^*)$).

The cost function minimization is, essentially, based on the following step:

$$r_{i+1} = r_i \oplus d(r_i) \quad (16)$$

where " \oplus " denotes the operator that combines two consecutive frame transformations, r_i is the current pose of the mobile robot (at frame i), r_{i+1} is the next pose of the mobile robot and $d(r_i)$ is the direction of descent.

This direction of descent must ensure that $d(r_i) \nabla C(r_i) < 0$. In this way, the movement of the robot leads to the decrease of the cost function.

Optimization methods depend on the direction of descent used in the building of the control law. The control law usually used in visual servoing context is given by:

$$v = -\lambda L_s^+ (s(r) - s(r^*)) \quad (17)$$

where λ is a positive scalar and L_s^+ is the pseudo inverse of the interaction matrix.

This classical control law gives good results in the case of visual servoing task based on geometric visual features (Chaumette and Hutchinson, 2006).

Since we work with photometric visual features this classical control law fails and doesn't ensure the convergence of the robot (Collewet and al., 2008). Thus, in our work we use the control law based on the Levenberg-Marquardt approach. The control law generated to the robot, using our new features, is then given by:

$$v_c^i = -\lambda \left(H_{E_i^i} + \mu \text{diag} \left(H_{E_i^i} \right) \right)^{-1} L_{E_i^i}^T e_i \quad (18)$$

where e_i is the error corresponding to these new features:

$$e_i = E_i^i(r) - E_i^i(r^*) \quad (19)$$

and with

$$H_{E_i^i} = L_{E_i^i}^T L_{E_i^i} \quad (20)$$

4 EXPERIMENTAL RESULTS

4.1 Experimental Environment

We present the results of a set of experiments conducted with our visual features. All the experiments reported here have been obtained using a camera mounted on a mobile robot. In each case, the mobile robot is first moved to its desired pose r^* and the corresponding image I^* is acquired. From this desired image, we extract the desired visual features s^* . The robot is then moved to a random pose r and the initial visual features s are extracted. The velocities computed, at each frame, using the control law, are sent to the robot until its convergence. The interaction matrix is calculated at each frame of the visual servoing scheme. In a first step we conduct our experiments on a virtual platform of VRML, therefore we can recuperate, at each frame, the pose of the mobile robot in terms of position along two translational axes and around one rotational axis. In a second step we validate our results on a real mobile robot (Koala robot).

4.2 Interpretation

During the experiments conducted on the VRML environment we take as initial positioning error: $\Delta r_{\text{int}} = (18 \text{ cm}, 12 \text{ cm}, 9^\circ)$. We illustrate the results obtained using our new visual features on

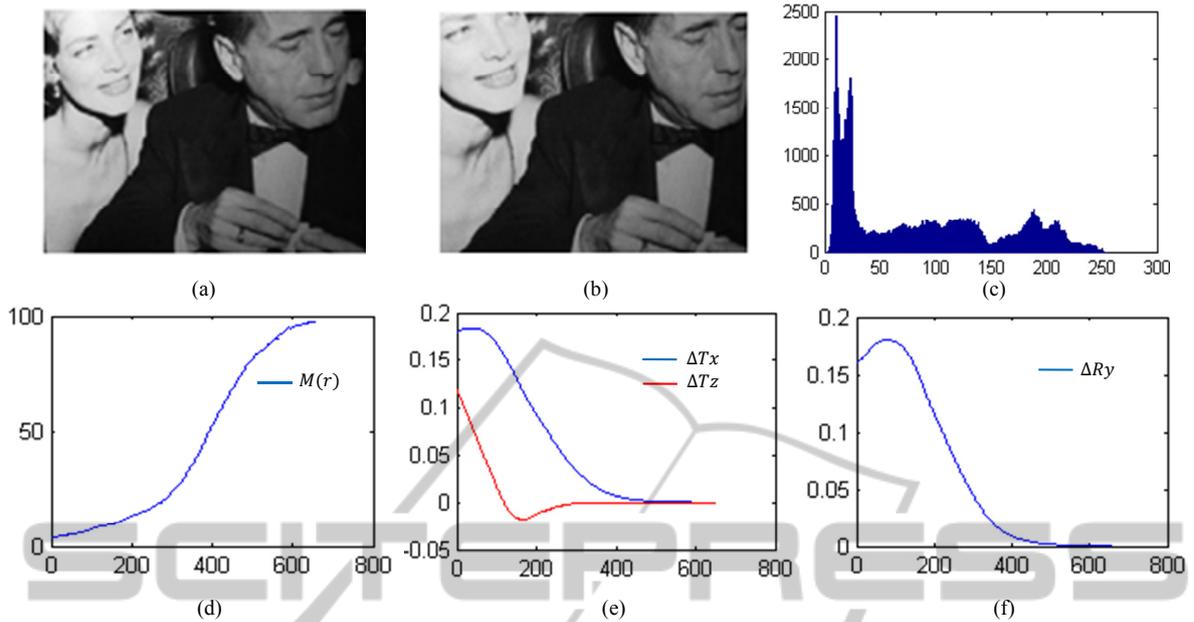


Figure 3: First experiment with our new global visual features (x axis in frame number for (d), (e) and (f)): (a) Initial image, (b) Desired image, (c) Initial image histogram, (d) Stopping criterion evolution: $M(r)$ in percentage (%), (e) Translational positioning errors: ΔTx and ΔTz in meter (m), (f) Rotational positioning error: ΔRy in radian (rad).

Figures 3 and 4 (first and second experiment).

Figures 3a and 4a present the initial scenes. Figures 3b and 4b depict the desired scenes. The histograms of the initial images are shown on Figures 3c and 4c.

We choose as stopping criterion of our program the following measure: $M(r)$ which is the proportion of the number of pixels, in the error image ($I - I^*$), whose luminance is below a certain threshold compared to the total number of image pixels.

$$M(r) = \frac{N_{thres}(r)}{N_{total}} \times 100 \quad (21)$$

where $N_{thres}(r)$ is the number of pixels in the error image whose luminance is below a predefined threshold at pose r of the robot and N_{total} is the total number of pixels (320×240).

In our experiments we choose the luminance value 3 as a threshold. We suppose that the convergence is achieved and the robotic system reaches its desired pose when $M(r)$ get at 98%.

Figures 3d and 4d depict the behavior of this stopping criterion. The translational positioning errors ($\Delta Tx, \Delta Tz$) between the current and the desired pose during the positioning task are shown on Figures 3e and 4e. The rotational positioning errors (ΔRy) are illustrated on Figures 3f and 4f.

We note that the robotic system converges with good behaviour using our global visual features

($s(r) = E_i^j(r)$) and it spend very less time compared to the method of (Collewet and al., 2008).

Indeed, our method reduces the size of the visual features vector s . Thus, the size of the interaction matrix related to our visual features (L_{E_i}) is very small compared to the size of the interaction matrix related to the whole image luminance. Therefore, our approach is more suitable to real-time applications. As an example, the experiment of Figure 3 has demonstrated that, using our approach, the computation time for each 320×240 frame does not exceed 40 ms while it is 270 ms when we work with the whole image luminance as visual features.

After using the virtual platform of VRML, we validated our new approach using the Koala mobile robot which is a differential wheeled robot. The results of the experiments conducted on the Koala are illustrated on Figure 5. We remark that this mobile robot correctly converges to its desired pose using our new global visual features. The initial and the desired scene are reported respectively on Figures 5a and 5b. The evolutions of the velocities of the two robot wheels are illustrated on Figure 5c where φ_r is the right wheel and φ_l is the left one. The stopping criterion evolution is shown on Figure 5d. So, we can confirm that our new visual features give good results in the case of real conditions of visual servoing task.

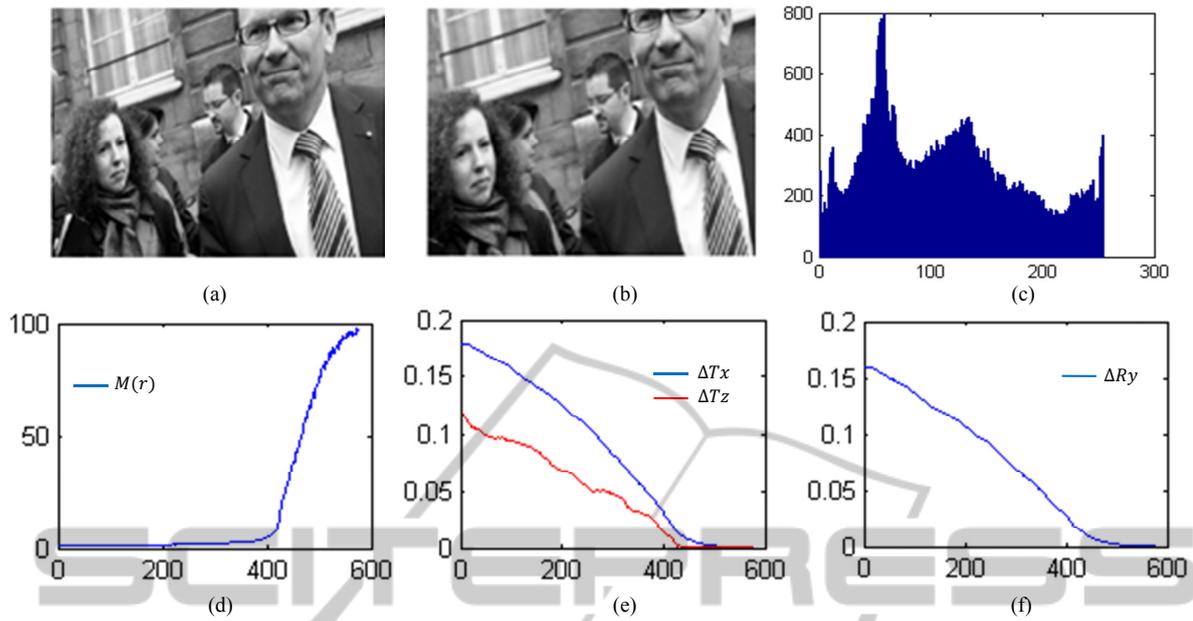


Figure 4: Second experiment with our new global visual features (x axis in frame number for (d), (e) and (f)): (a) Initial image, (b) Desired image, (c) Initial image histogram, (d) Stopping criterion evolution: $M(r)$ in percentage (%), (f) Translational positioning errors: ΔTx and ΔTz in meter (m), (f) Rotational positioning error: ΔRy in radian (rad).

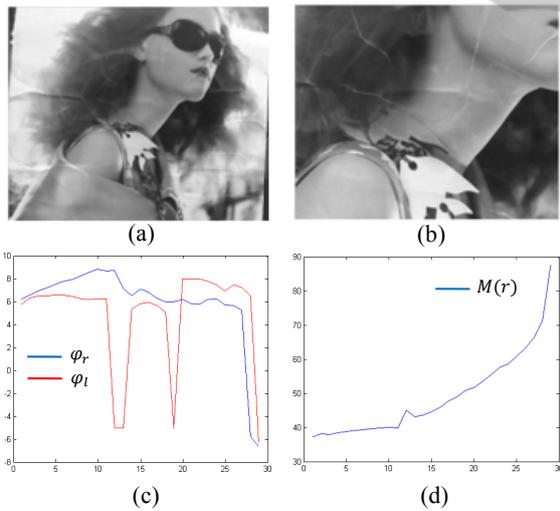


Figure 5: Our global visual features (x axis in frame number). (a) Initial scene, (b) Desired scene, (c) The velocities of the two robot wheels (mm/s), (d) Stopping criterion evolution: $M(r)$ (%).

4.3 Robustness with Respect to Image Content

Our approach does not depend on the image content. In fact, the experiments demonstrate that the control law converges even in the case of low textured scenes.

Figure 6 shows that using different types of scenes the control law converges in all the cases (we keep the same initial positioning errors). The images presented here are those used in (Collewet and al., 2010).

The first column in Figure 6 shows the different scenes. The second represents the corresponding histograms. The third and the fourth column illustrate, respectively, the translational and the rotational positioning errors during the visual servoing scheme.

5 CONCLUSIONS

In this paper we focused on the importance of global visual features in visual servoing applications.

We found that when the used global feature is the whole image luminance the mobile robot takes so much time to reach its desired pose, therefore we proposed a new approach to achieve fast and real-time visual servoing tasks. This approach is based on new global feature which is the luminance of a random distribution of image points. To demonstrate the efficiency of this new method our works were, firstly, realized on a virtual platform of VRML then on a real mobile robot. To get the convergence of the robot we have turned the visual servoing problem into an optimization problem. Thus, we have used

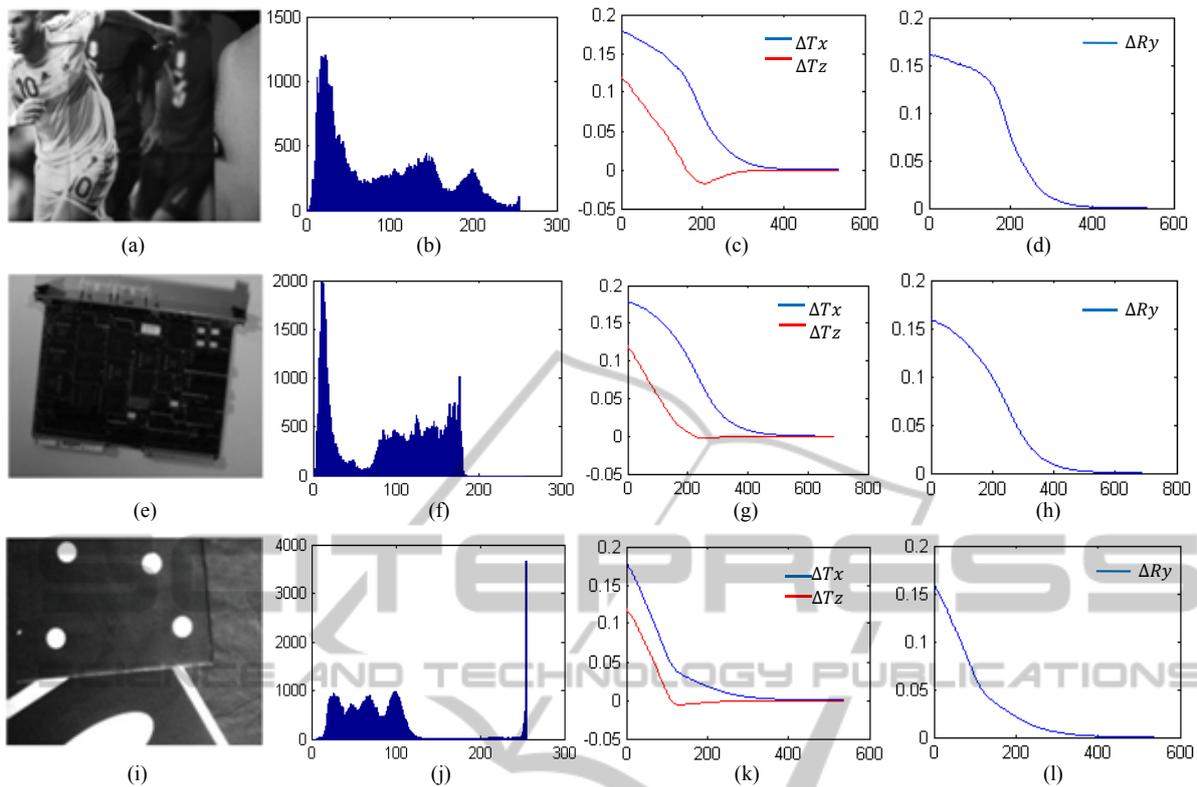


Figure 6: Results of our approach in different cases of scenes. First column: scenes considered, second column: corresponding histograms, third column: translational positioning errors in meter (x axis in frame number), fourth column: rotational positioning errors in radian (x axis in frame number).

the control law based on the minimization of a cost function since that ensures the convergence in the case of global visual features.

The new feature has proved to be able to ensure good and fast convergence of the mobile robot even in the case of low textured scenes. As it is global, it does not require any matching nor tracking step and there is no image processing step.

Future works can be intended to verify the robustness of our approach with respect to partial occlusions and large illumination changes.

REFERENCES

- Abdul Hafez A., Achar S., and Jawahar, C., 2008. Visual servoing based on Gaussian mixture models. *In IEEE Int. Conf. on Robotics and Automation*, Pasadena, California, pp 3225–3230.
- Abdul Hafez, A., and Jawahar, C., 2006. Improvement to the minimization of hybrid error functions for pose alignment, *in IEEE Int. conf. on Automation Robotics, Control and Vision*, Singapore.
- Abdul Hafez, A.H., and Jawahar, C., 2007. Visual Servoing by Optimization of a 2D/3D Hybrid

- Objective Function. *In IEEE International Conference on Robotics and Automation*, pp 1691 – 1696, Roma.
- Chaumette, F., and Hutchinson, S., 2003. Application of moment invariants to visual servoing. *In IEEE Int. Conf. on Robotics and Automation*, pp 4276-4281, Taiwan.
- Chaumette, F., and Hutchinson, S., 2006. Visual servo control, Part I: Basic approaches. *In IEEE Robotics and Automation Magazine*, 13(4):82–90.
- Chaumette, F., and Hutchinson, S., 2007. Visual servo control, part II: Advanced approaches”. *IEEE Robotics and Automation Magazine*, vol. 14, no. 1.
- Chaumette, F., and Hutchinson, S., 2008. Visual servoing and visual tracking”. *In B. Siciliano and O. Khatib, editors, Handbook of Robotics*, chapter 24, pp. 563–583. Springer.
- Collewet, C., Marchand, E., and Chaumette, F., 2008. Visual servoing set free from image processing. *In IEEE Int. Conf. on Robotics and Automation*, CA, pp. 81–86, Pasadena.
- Collewet, C., Marchand, E., and Chaumette, F., 2010. Luminance: a new visual feature for visual servoing. *In Visual Servoing via Advanced Numerical Methods LNCIS 401, Springer-Verlag (Ed)*, 71–90.
- Collewet, C., and Marchand, E., 2011. Photometric visual servoing. *In IEEE Trans. on Robotic*, Vol. 27, No.4. pp. 828-834.

- Dame, A., Marchand, E., 2009. Entropy-based visual servoing. In *IEEE ICRA'09*, pp. 707–713, Kobe, Japan.
- Espiau, B., Chaumette, F., and Rives, P., 1992. A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, 8(3):313-326.
- Marchand, E., 2007. Control camera and light source positions using image gradient information. In *IEEE Int. Conf. on Robotics and Automation, ICRA'07*, pages 417–422, Roma, Italia.
- Marchand, E., and Collewet, C., 2010. Using image gradient as a visual feature for visual servoing. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp 5687-5692, Taiwan.

