Differential
Equations
& Applications
Volume 3, Number 4 (2011), 555–564

A VARIATIONAL APPROACH TO CAMERA MOTION SMOOTHING

LUIS ALVAREZ, LUIS GOMEZ, PEDRO HENRIQUEZ AND LUIS MAZORRA

Dedicated to Professor Jesús Ildefonso Díaz
on the occasion of his 60th birthday

(*Communicated by M. Pašić*)

Abstract. In this paper we study a variational problem derived from a computer vision application: video camera calibration with smoothing constraint. By video camera calibration we mean to estimate the location, orientation and lens zoom-setting of the camera for each video frame taking into account image visible features. To simplify the problem we assume that the camera is mounted on a tripod, in such case, for each frame captured at time t, the calibration is provided by 3 parameters : (1) $P(t)$ (PAN) which represents the tripod vertical axis rotation, (2) $T(t)$ (TILT) which represents the tripod horizontal axis rotation and (3) $Z(t)$ (CAMERA ZOOM) the camera lens zoom setting. The calibration function $t \to \mathbf{u}(t) = (P(t), T(t), Z(t))$ is obtained as the minima of an energy function $I[\mathbf{u}]$. In this paper we study the existence of minima of such energy function as well as the solutions of the associated Euler-Lagrange equations.

1. Introduction

In this paper we deal with video camera calibration with smoothing constraint. We focus our attention on the case of video cameras mounted on a tripod. In such case, for each time t of the video sequence, camera calibration configuration is provided by 3 parameters : (1) $P(t)$ (PAN) which represents the tripod vertical axis rotation, (2) $T(t)$ (TILT) which represents the tripod horizontal axis rotation and (3) $Z(t)$ (CAMERA ZOOM) the camera lens zoom setting. From these time-dependent functions we can easily deduce (using tripod information), for each time t, the camera calibration, that is, the camera intrinsic and extrinsic parameters which determine the position of the camera in the $3D$ space and the way $3D$ objects are projected in the camera projection plane (the CCD in the case of digital cameras).

Human being visual system is very sensitive to motion, and small perturbations over time of $P(t)$, $T(t)$, and $Z(t)$ values produce small oscillations in the camera motion disturbing the observer. To remove such perturbations is a critical issue in applications like the inclusion of artificial graphic objects in real video sequence scenarios. To illustrate this phenomenon, in http://www.ctim.es/demo102 we show a real video

Mathematics subject classification (2010): 35J20, 35J50, 35J55, 35J60.
Keywords and phrases: variational methods, camera calibration, Euler-Lagrange equations.
This research has partially been supported by the MICINN project reference MTM2010-17615 (Ministerio de Ciencia e Innovación. Spain).

Paper DEA-03-34

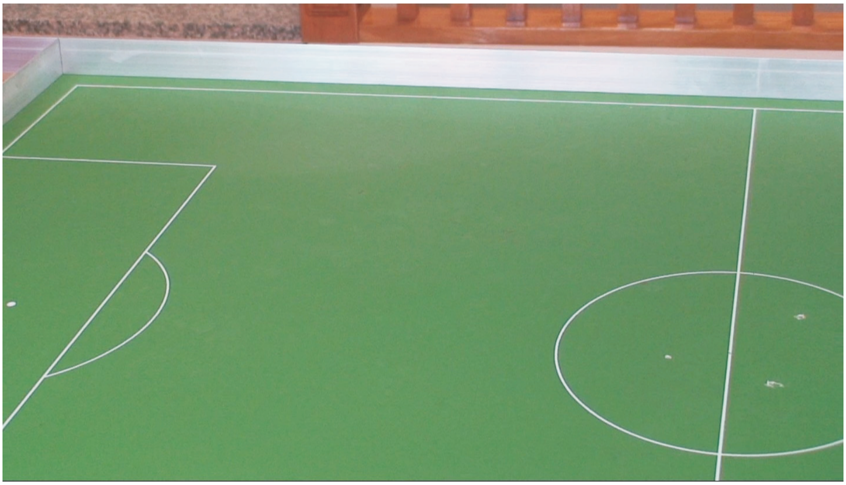

Figure 1: One video frame of a typical application scenario. The primitives used for calibration are the white lines and cercles detected in the image.

sequence where some graphic objects have been included. In this video we illustrate the main practical problem we deal with, that is, standard calibration techniques which do not take into account the expected time regularity of $\mathbf{u}(t) = (P(t), T(t), Z(t))$ can introduce a significant noise in the camera motion estimation over time.

In this paper we propose to include a smoothing constraint in the estimation of $\mathbf{u}(t)$ by minimizing the following energy :

$$I[\mathbf{u}] = \int_{t_0}^{t_1} \left(P'(t)^2 + T'(t)^2 + Z'(t)^2 + \alpha F(\mathbf{u}(t), t)\right) dt, \tag{1}$$

where $[t_0, t_1]$ is the time interval, $\alpha \geqslant 0$ is a weight to balance the different components of the energy and, $F(x_1, x_2, x_3, t) \geqslant 0$ is a standard calibration function which forces, for each time t, that the projection of $3D$ points be close to primitives detected in the image. In fact, when no time regularization is used, $\mathbf{u}(t)$ is usually estimated by minimizing $F(\mathbf{u}(t), t)$ independently for each time t. So, by using the proposed variational model (1), we introduce a time regularity condition in the video camera calibration procedure.

The Euler-Lagrange equations associated to energy (1) yields to the following non-linear system of differential equations:

$$\begin{cases} -P''(t) + \alpha \frac{\partial F}{\partial x_1}(P(t), T(t), Z(t), t) = 0 & \text{in } (t_0, t_1), \\ -T''(t) + \alpha \frac{\partial F}{\partial x_2}(P(t), T(t), Z(t), t) = 0 & \text{in } (t_0, t_1), \\ -Z''(t) + \alpha \frac{\partial F}{\partial x_3}(P(t), T(t), Z(t), t) = 0 & \text{in } (t_0, t_1), \end{cases}$$

with adequate boundary conditions. The paper is organized as follows: in section 2, we briefly introduce the basic mathematical models we use for camera calibration. In section 3, we analyze the calibration function $F(x_1,x_2,x_3,t)$. In section 4, we show the existence of minimizers of energy functional (1) and that such minimizers are solutions of the associated Euler-Lagrange equations. Finally, in section 5, we present some experiments to illustrate the smoothing performance of the proposed method.

2. Geometry and calibration of cameras mounted on a tripod

A tripod is defined by two unitary rotation axes $\bar{e}^0 = \left(\bar{e}^0_x, \bar{e}^0_y, \bar{e}^0_z\right)^T$ and $\bar{e}^1 = \left(\bar{e}^1_x, \bar{e}^1_y, \bar{e}^1_z\right)^T$ and a center of rotation $\bar{X}_0 \in \mathbb{R}^3$. Pan parameter $P(t)$ determines the tripod rotation angle with respect to $\bar{e}^0$. We denote by $R\left(\bar{e}^0, P(t)\right)$ the associated rotation matrix. In a similar way we define $R\left(\bar{e}^1, T(t)\right)$. The general rotation matrix generated by tripod motion is a composition of the above rotation matrix. Given a 3D point $\bar{X}$ the point transformation induced by the tripod motion can be expressed as:

$$\bar{X}\left(P(t),T(t)\right) = \bar{X}_0 + R\left(\bar{e}^0,P(t)\right)R\left(\bar{e}^1,T(t)\right)\left(\bar{X}-\bar{X}_0\right). \tag{2}$$

We use the basic pinhole model to modelize the way the 3D scene is projected in the 2D image projection plane. Such projection is expressed in projective coordinates as a 4×3 projection matrix $\mathscr{P}(\mathbf{u}(t))$ defined by

$$\mathscr{P}(\mathbf{u}(t)) \equiv A\left(Z(t)\right)R_0\left[Id,-c^0\right]\begin{pmatrix} R\left(P(t),T(t)\right) & \bar{t}\left(P(t),T(t)\right) \\ 0 & 1 \end{pmatrix}, \tag{3}$$

where:

$$R\left(P(t),T(t)\right) \equiv R\left(\bar{e}^0,P(t)\right)R\left(\bar{e}^1,T(t)\right), \tag{4}$$

$$\bar{t}\left(P(t),T(t)\right) = \bar{X}_0 - R\left(P(t),T(t)\right)\bar{X}_0, \tag{5}$$

$$A\left(Z(t)\right) = \begin{pmatrix} Z(t) & 0 & x_c \\ 0 & r\cdot Z(t) & y_c \\ 0 & 0 & 1 \end{pmatrix}, \tag{6}$$

$$R_0 = \begin{pmatrix} r^0_{00} & r^0_{01} & r^0_{02} \\ r^0_{10} & r^0_{11} & r^0_{12} \\ r^0_{20} & r^0_{21} & r^0_{22} \end{pmatrix}, \tag{7}$$

$$\left[Id,-c^0\right] = \begin{pmatrix} 1 & 0 & 0 & -\bar{c}^0_x \\ 0 & 1 & 0 & -\bar{c}^0_y \\ 0 & 0 & 1 & -\bar{c}^0_z \end{pmatrix}. \tag{8}$$

R_0 and $\bar{c}^0$ correspond to the initial tripod rotation and translation. For more details about the pinhole model see, for instance, [3], [4].

3. The calibration function $F(\mathbf{u}(t),t)$

To calibrate a camera frame is to obtain the associated $\mathbf{u}(t) = (P(t),T(t),Z(t))$ parameters which determine the projection matrix $\mathscr{P}(\mathbf{u}(t))$. To estimate $\mathbf{u}(t)$ we use the "observable" information we can obtain in the image. For instance in the image of fig. 1, we can see some parts of the white lines and circles of the soccer court. In general, we note by $\Omega(t) \subset \mathbb{R}^2$ the finite collection of visible curves in the image we use to calibrate frame t. We assume that we know the actual 3D location of $\Omega(t)$ in the real scene. That is, for any curve $\tilde{s}(.) \in \Omega(t)$, we know its actual $3D$ location $\tilde{S}(.)$. Then, the calibration function $F(\mathbf{u}(t),t)$ is defined as

$$F(\mathbf{u}(t),t) \equiv \sum_{\tilde{s}\in\Omega(t)} \oint_{\tilde{s}} \text{distance}(\mathscr{P}(\mathbf{u}(t))\tilde{S}(.),\tilde{s}(q))^2 dq, \tag{9}$$

where distance$(C,\mathbf{x})$ is the usual euclidean distance between a point $\mathbf{x}$ and a curve C. We observe that for any $\mathbf{u}(t)$, $F(\mathbf{u}(t),t) \geqslant 0$ and, the smaller $F(\mathbf{u}(t),t)$, the better is the matching between the 3D scene and its 2D image projection. The usual way to calibrate a camera (that is to obtain $\mathbf{u}(t)$) is by minimizing the calibration function $F(\mathbf{u}(t),t)$. We also observe that $F(\mathbf{u}(t),t)$ is non convex, strongly nonlinear and, in general, we can not expect uniqueness of the minima because it is strongly dependent on the geometry of observable curves $\Omega(t)$ (in fact, in some cases $\Omega(t)$ could be empty).

4. Variational formulation of video calibration problem

We observe that when we minimize the calibration function $F(\mathbf{u}(t),t)$ with respect to $\mathbf{u}(t)$ no assumption is made concerning the regularity in time t of $\mathbf{u}(t)$. To add such regularity condition to the calibration model we propose to minimize the functional

$$I[\mathbf{w}] = \int_{t_0}^{t_1} L(D\mathbf{w}(t),\mathbf{w}(t),t)dt, \tag{10}$$

where $[t_0,t_1]$ is the time interval,

$$L(p,z,t) = \|p\|^2 + \alpha F(z,t) \tag{11}$$

is a weight to balance the different components of the energy and $\alpha \geqslant 0$.

Next we will show the existence of minimizer of $I[.]$. Let be

$$\mathscr{A} = \big\{\mathbf{w} \in W^{1,2}((t_0,t_1);\mathbb{R}^3) \text{ such that } \mathbf{w}(t_0) = (P_0,T_0,Z_0) \text{ and } \mathbf{w}(t_1) = (P_1,T_1,Z_1)\big\}.$$

THEOREM 1. (existence of minimizer) *There exists* $\mathbf{u} \in \mathscr{A}$ *solving*

$$I[\mathbf{u}] = \min_{\mathbf{w}\in\mathscr{A}} I[\mathbf{w}].$$

Proof. To show the existence of minimizer we use the following classical result that, for instance, is presented in [2].

THEOREM A (see Evans [2, Theorem 5, p.453]) *Assume that L satisfies the coercivity inequality*

$$L(p,z,x) \geqslant \alpha \|p\|^q - \beta \tag{12}$$

for constants $\alpha > 0$, $\beta \geqslant 0$ *and* $q > 1$ *and is convex in the variable* p. *Suppose also the admissible set* A *is nonempty. Then there exists* $u \in A$ *solving*

$$I[\mathbf{u}] = \min_{\mathbf{w} \in \mathscr{A}} I[\mathbf{w}].$$

In our case, we observe that $L(p,z,t)$ is convex with respect to p. On the other hand, since $F(z,t) \geqslant 0$ and $\alpha \geqslant 0$, $L(p,z,t)$ satisfies the coercivity inequality

$$L(p,z,t) \geqslant \|p\|^2$$

and therefore the condition (12) is satisfied with $q = 2$ and $\beta = 0$. On the other hand obviously $\mathscr{A}$ is not empty ($\mathscr{A}$ contains simple linear functions). So then, statement of the Theorem 1 follows by straight application of the above result. □

Next, we study if the minima of $I[.]$ are solutions of the associated Euler-Lagrange system. Using the classical theory we need to show some growth conditions on $L(p,z,t)$. First we observe that, in practice, the projection image is given by a rectangle $[0,a] \times [0,b]$ and then the observed curves $\Omega(t)$ are included in such rectangles. In practice, we are interested in estimating the distance function "distance$(\mathscr{P}(z)\tilde{S}(.),\tilde{s}(q))$" when the curve intercepts the image rectangle $[0,a] \times [0,b]$. Therefore, without loss of generality, we can change the distance(.) function in (9) by

$$\text{distance}_M(x,y) = \begin{cases} \text{distance}(x,y) & \text{if } \text{distance}(x,y) \leqslant M, \\ M & \text{if } \text{distance}(x,y) > M, \end{cases}$$

where $M = \sqrt{a^2 + b^2}$. We define the modified calibration function as

$$F_M(z,t) \equiv \oint_{\Omega(t)} \text{distance}_M(\mathscr{P}(z)\tilde{S}(.),\tilde{s}(q))^2 dq$$

and we can state the following result.

THEOREM 2. (solution of Euler-Lagrange system) *If* $\Omega(t)$ *is composed by a finite number of curves which length uniformly bounded in* $[t_0,t_1]$, *then the calibration function* $\mathbf{u}(t) = (P(t),T(t),Z(t))$ *satisfies*

$$I_M[\mathbf{u}] = \min_{\mathscr{A}} I_M[\mathbf{w}]$$

and it is a weak solution of the system

$$\begin{cases} -P''(t) + \alpha \frac{\partial F_M}{\partial x_1}(\mathbf{u}(t),t) = 0 & \textit{in } (t_0,t_1), \\ -T''(t) + \alpha \frac{\partial F_M}{\partial x_2}(\mathbf{u}(t),t) = 0 & \textit{in } (t_0,t_1), \\ -Z''(t) + \alpha \frac{\partial F_M}{\partial x_3}(\mathbf{u}(t),t) = 0 & \textit{in } (t_0,t_1), \end{cases} \tag{13}$$

where

$$I_M[\mathbf{w}] = \int_{t_0}^{t_1} L_M(D\mathbf{w}(t), \mathbf{w}(t), t)dt,$$

and

$$L_M(p,z,t) = \|p\|^2 + \alpha F_M(z,t). \tag{14}$$

Proof. To show the result we use the following classical result that, for instance, is presented in [2].

THEOREM B (see Evans [2, Theorem 7, p.454]) *Assume L verifies the growth conditions*

$$\begin{cases} \|L(p,z,x)\| \leqslant C\big(\|p\|^q + \|z\|^q + 1\big), \\ \|D_pL(p,z,x)\| \leqslant C\big(\|p\|^{q-1} + \|z\|^{q-1} + 1\big), \\ \|D_zL(p,z,x)\| \leqslant C\big(\|p\|^{q-1} + \|z\|^{q-1} + 1\big) \end{cases} \tag{15}$$

for constants $C > 0$ and $q > 1$ and $\mathbf{u} \in \mathscr{A}$ satisfies

$$I[\mathbf{u}] = \min_{\mathscr{A}} I[\mathbf{w}].$$

Then u is a weak solution of (13).

In our case, we apply the above theorem to L_M defined in (14). First we observe that, since the length of curves of $\Omega(t)$ are uniformly bounded in $[t_0,t_1]$, and the function distance$_M(.,.)$ is bounded then the function $F_M(z,t)$ is bounded in $[t_0,t_1]$. On the other hand the two first components of vector z are angles so $F_M(z,t)$ is periodic with respect to z_x and z_y and it has bounded derivatives with respect to z. So we can deduce that there exists $C > 0$ such that

$$\begin{cases} \|L_M(p,z,t)\| \leqslant C\big(\|p\|^2 + 1\big), \\ \|D_pL_M(p,z,t)\| \leqslant C\|p\|, \\ \|D_zL_M(p,z,t)\| \leqslant C, \end{cases}$$

and therefore the statement of the theorem 2 follows by straight application of the above result. □

5. Experiments

To illustrate the performance of the proposed variational model we compare the results obtained for a real video sequence using the minimization of the calibration function $F(\mathbf{u}(t),t)$ independently for each time t and the results obtained with the proposed variational model. In figures 2-7 we show the plots of the obtained $P(t)$, $T(t)$ and $Z(t)$ using both methods as well as their difference.

To observe by visual inspection the quality of the variational method, in

http://www.ctim.es/demo102

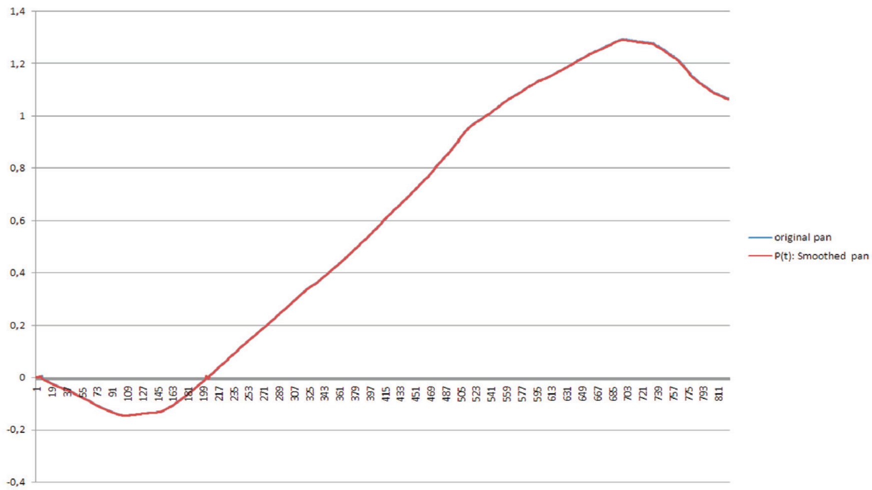

Figure 2: $P(t)$ obtained minimizing $F(\mathbf{u}(t),t)$ independently for each time t and $P(t)$ obtained minimizing $I[\mathbf{u}]$

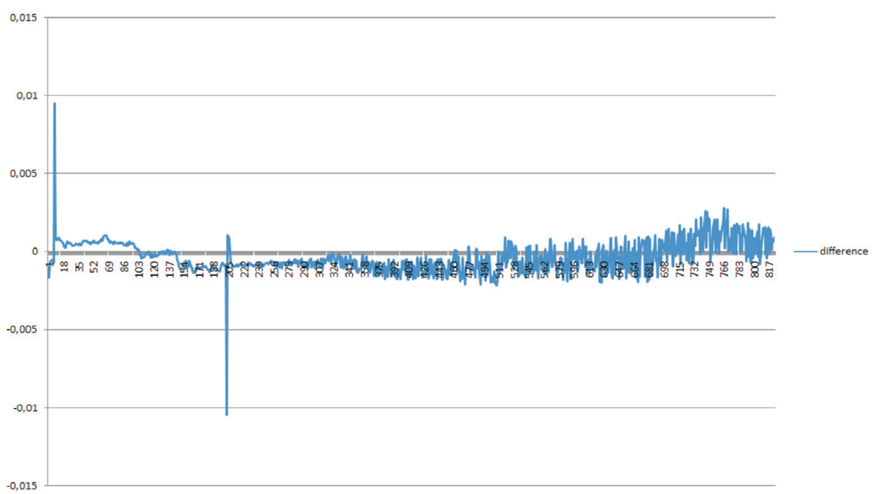

Figure 3: Plot of the difference between the estimation of $P(t)$ using both methods

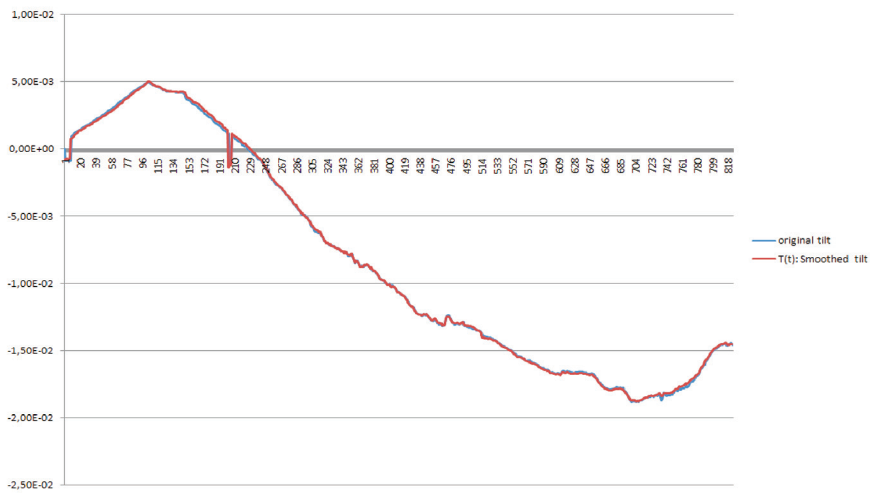

Figure 4: $T(t)$ obtained minimizing $F(\mathbf{u}(t),t)$ independently for each time t and $P(t)$ obtained minimizing $T[\mathbf{u}]$

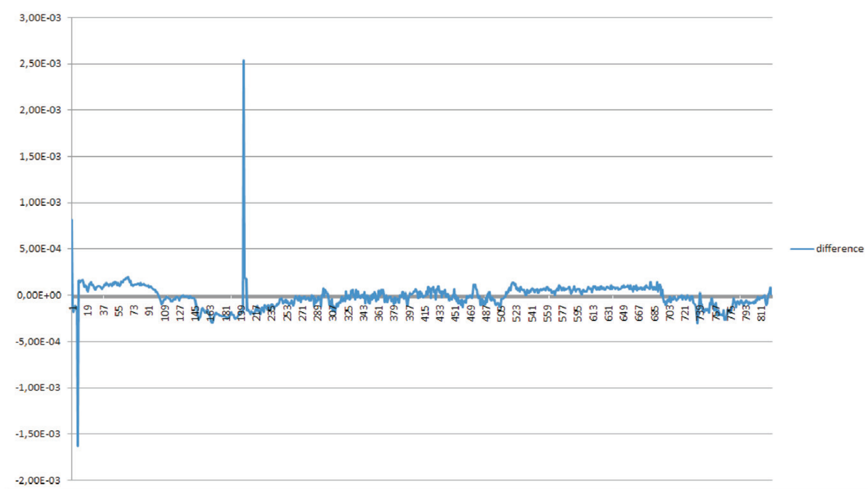

Figure 5: Plot of the difference between the estimation of $T(t)$ using both methods

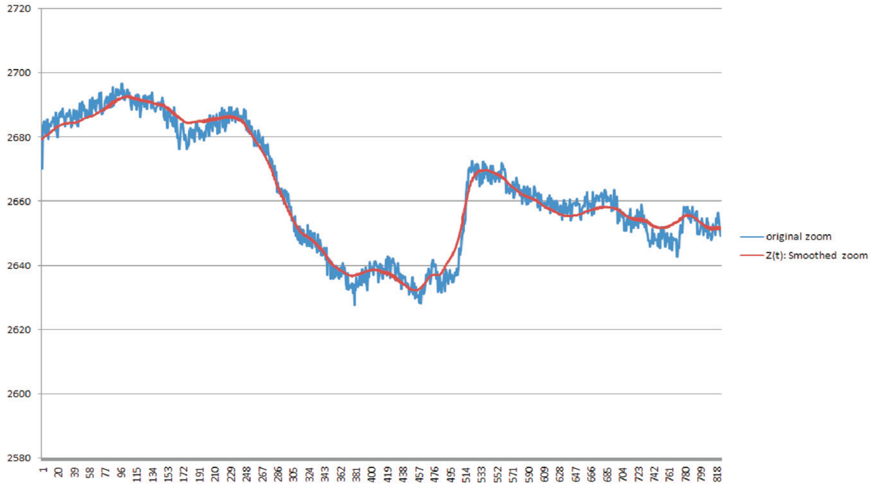

Figure 6: $Z(t)$ obtained minimizing $F(\mathbf{u}(t),t)$ independently for each time t and $Z(t)$ obtained minimizing $I[\mathbf{u}]$

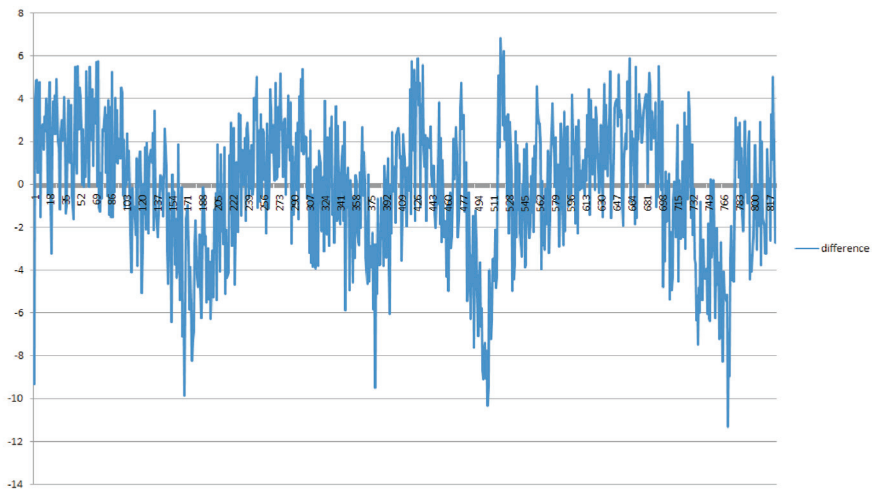

Figure 7: Plot of the difference between the estimation of $Z(t)$ using both methods

we present the video sequence we use in the experiments where some artificial objects have been included in the sequence using the obtained calibration information (once the camera is calibrated we can easily include artificial objects in the scene using computer graphic techniques). We can observe the quality of the calibration by looking at the small graphic object oscillation we can detect visually in the video. It can be appreciated that, as the calibration improves, the observed graphic object oscillations are considerably attenuated or removed. We note that the obtained results are very promising and by using the proposed variational technique we strongly reduce such disturbing oscillations affecting the included graphic objects.

REFERENCES

[1] J. I. DÍAZ, *Nonlinear partial differential equations and free boundaries*, Vol. I, Elliptic equations, Pitman Advanced Publishing Program, N. 106, 1985.

[2] L. E. EVANS, *Partial Differential Equations*, AMS, 1985.

[3] O. FAUGERAS, *Three-Dimensional Computer Vision*, MIT Press, 1993.

[4] O. FAUGERAS, Q.-T. LUONG, T. PAPADOPOULO, *The geometry of multiple images*, MIT Press, 2001.

[5] M. GIAQUINTA, S. HILDEBRANDT, *Calculus of Variations*, Vol 1-2, Springer, 1996.

[6] E. ZEIDLER, *Nonlinear Functional Analysis and its Applications*, Part III, Springer, 1984.

[7] Z. ZHANG, *A flexible new technique for camera calibration*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (11), 2000, 1330–1334.

(Received August 24, 2011)

Luis Alvarez
CTIM. Departamento de Informática y Sistemas
Univ. Las Palmas de G.C., Campus de Tafira
35017 Las Palmas de G.C.
Spain
e-mail: lalvarez@ctim.es

Luis Gomez
CTIM. Departamento de Ingeniería Electrónica y Automática
Univ. Las Palmas de G.C., Campus de Tafira
35017 Las Palmas de G.C.
Spain
e-mail: lgomez@ctim.es

Pedro Henriquez
CTIM. Departamento de Informática y Sistemas
Univ. Las Palmas de G.C., Campus de Tafira
35017 Las Palmas de G.C.
Spain
e-mail: phenriquez@ctim.es

Luis Mazorra
CTIM. Departamento de Informática y Sistemas
Univ. Las Palmas de G.C., Campus de Tafira
35017 Las Palmas de G.C.
Spain
e-mail: lmazorra@ctim.es