Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Dynamical Effects of Vision-Based Position Measurement

Edouard Laroche, Julien Delavigne

Image Sciences, Computer Sciences and Remote Sensing Laboratory,
CNRS, University of Strasbourg 1,
BP 10413, 67412 Illkirch cedex, France,
Phone: 00 33 3 90 24 44 68, Fax: 00 33 3 90 24 44 80;
laroche@lsiit.u-strasbg.fr

Abstract: In vision-based control schemes, a position measurement is computed from image processing. This kind of sensor sensibly differs from the classical physical sensors for which the data can be filtered before being sampled. This paper is dedicated to the analysis of the dynamical effects of this measurement system. A model is derived allowing to emphasize the behavior in the high frequencies, with better accuracy than the models available. When the camera is used as a sensor for a mechanical system with a continuous-time model, we show that a discrete-time model can be derived that accounts accurately for the camera effect. When considering identification of the continuous-time model of the system, it is shown that accurate results can be obtained with output measurement provided by camera by considering the input data as provided by a first-order hold instead of a zero-order hold. It is also shown that accurate results can be obtained even if the eigen modes are close to the Shannon frequency and when neglecting modes at higher frequency than the Shannon frequency.

Keywords: camera, position sensor, continuous-time identification, dynamical model

1. INTRODUCTION

Vision systems with real-time image processing have been developping during the last decades and are now able to provide measurements allowing to estimate the position of an object at relatively high frequency. Based on this measurement, specific control schemes have been developed for controlling robotic manipulators (Corke [1996]). Andersson [1989] developed a robot playing ping-pong based on a 60 Hz stereo vision system. Nakabo et al. [2000] used a 1000 Hz vision chip in a visual loop and obtain astonishing fast results. Those approaches yielded interesting results in various fields, such as medical robotics, as shown by Krupa et al. [2003] and Ginhoux et al. [2005].

When dealing with the model of a camera, it is classical to consider the projective model whose parameters can be estimated from a calibration procedure (Brown et al. [2005]). This model is static and concerns the formation of the image of the 3D object. When the object is moving and if its displacements during an integration period of the image are significant, the measure provided by image processing is altered. Very few papers deal with this issue. To our knowledge, the only model was proposed by Ranftl et al. [2007], who implemented visual servoing for a fast ultrasonic actuator. More often, vibrations are considered as an external perturbation and its effects must be compensated for, such as in Luna et al. [2006].

The aim of this paper is to investigate the dynamical effects of the vision-based measurement. A mathematical model will be developed in Section 2, that allows to understand the effects of the camera as a position sensor. Based on this model, the dynamical effects of the camera when identifying a dynamical system will be investigated in Section 3. It will be show that an accurate discrete-time model can be derived. In the case of continuous-time identification, we will evaluate the effects of the camera in two cases: first when identifying a flexible system with eigen frequencies close to the Shannon frequency, second when identifying a system including high dynamics that are neglected in the identification model.

2. VISION-BASED POSITION MEASUREMENT

2.1 Description

Image processing allows to compute the position in the image of some elements of interest. For example, a LED placed at the tip of an instrument held by a robot yield a marker in the image and the center of mass of the marker can be computed in real-time and used for control. This kind of control schemes, also called *visual servoing* in the Robotics community, received great interest during the two last decades (see Hutchinson et al. [1996]).

Classically, the elements of interest are simple circular markers and the goal of the image processing is to compute the coordinates of their centers of mass in the image. In order to limit the computation cost for real-time implementation, this is generally done with the following method in three steps:

- a Sobel filter, i.e. a spacial high-pass filter, allows to emphasize the transition between the background and the spot;

- a threshold is applied in order to select only the pixels with high gradient;
- the center of mass of the remaining pixels is computed as an estimation of the center of the marker.

This computation is done in a limited region of interest where only one marker is to be found. Some variations can be considered: it is possible to detect the projection of a laser held by a robotic arm, such as in Krupa et al. [2003]. The marker can also be a simple black point placed on a white background as in Gangloff and de Mathelin [2003], Bachta et al. [2007]. In 2D visual servoing, the goal is to control directly the position of the markers in the image. Several markers can be combined in order to compute the configuration of a manipulator and allowing 3D visual servoing. Features of higher order can be also considered. For example, the position of a curved needle was considered as a portion of a circle and located with image processing by Nageotte et al. [2005]; visual servoing of spheres is presented by Fomena and Chaumette [2007]. Active vision systems, based on the projection of structured light can also be used for estimation of 3D surfaces as done by Albitar et al. [2007].

In the present paper, we focus on the characterization of the dynamical effects of the camera-based measurement system. The black-and-white camera considered herein has an integration period of $T = 20$ ms, corresponding to the classical frequency of 50 Hz. One image is acquired every T period. The camera resolution is 300×420 pixels. We consider the case of a circular marker of 30 pixels of diameter emitting constant light and placed over a black background. The process of the formation of a grey-level image can be considered as a continuous-time state system where the state vector contains the intensity of each pixel. The state equation, giving the rate of variation of the luminosity of each pixel, depends on the marker position. Such a model have been implemented; the obtained image for a fast displacement of the marker, corresponding to a circular trajectory with a diameter of 160 pixels at 24 Hz, is presented in Fig. 1, in addition with the trajectory of the center of the marker. The contour and the measurement obtained from image processing are given in Fig. 2. Notice that a color camera can be modeled in the same way by considering independently the three color components.

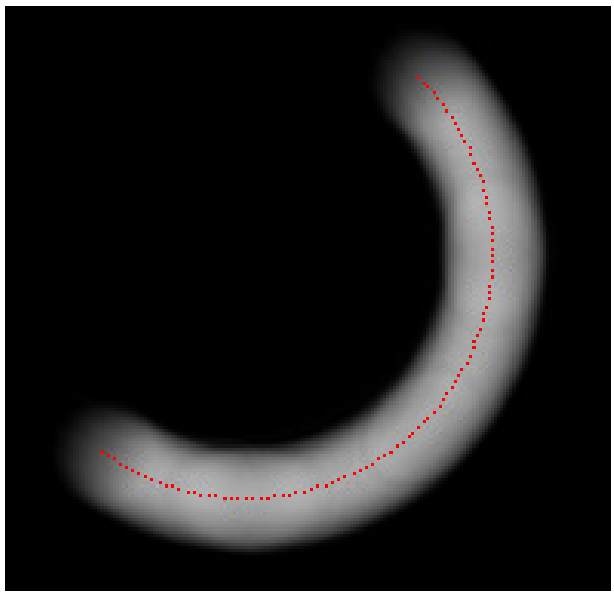

Fig. 1. Image produced by a fast circular displacement of the marker and the trajectory of the marker

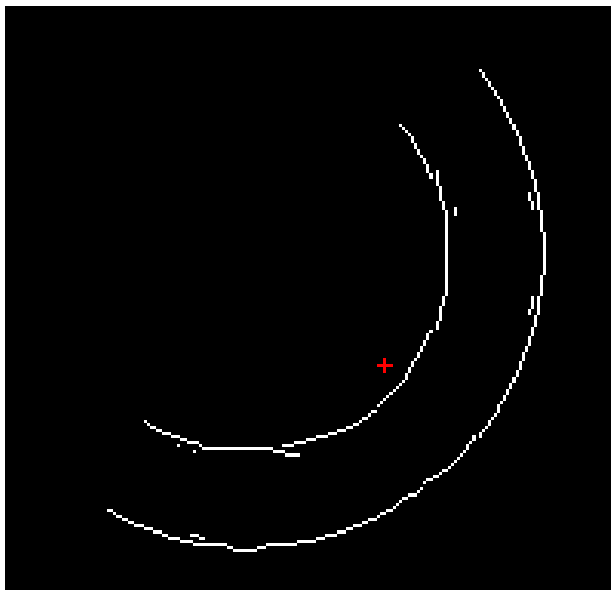

Fig. 2. Gradient image and its center of mass (+)

2.2 Modeling

Let consider the following assumptions concerning the image integration:

A1. the size of the marker is small compared to the displacement during an integration period, so that the gradient image is close to the original trajectory;
A2. the trajectory in the image of the marker do not intersect itself over an integration period (i.e. for $(t_1, t_2) \in [t_k, t_{k+1}]^2$, $(x_M(t_1), y_M(t_1)) = (x_M(t_2), y_M(t_2)) \Rightarrow t_1 = t_2$);
A3. the camera resolution is high so that spacial discretization effects can be neglected;
A4. the amplitude of the speed of the marker in the image is constant during an integration period.

Under assumptions A1 to A3, the measurement is equal to the center of mass of the curve depicted by the center of the marker. Denoting M a point of the curve located between M_{k-1} and M_k that are the positions at respectively t_{k-1} and t_k; denoting O the origin of the image, the center of mass G_k of the portion of curve of length l_k is defined by:

$$\int_{\mathrm{M}_{k-1}}^{\mathrm{M}_k} \overrightarrow{\mathrm{OM}}\,\mathrm{d}l = l_k\,\overrightarrow{\mathrm{OG}_k}. \qquad (1)$$

If assumption A1 is violated, the center of mass of the contour will be displaced for rotating trajectories. Indeed, the outer part of the contour has a longer length than the inner part, which produces a shift of the center of mass in the centrifugal direction. If assumption A2 is violated, some pixels can be found that correspond to the same position of the marker at two diferent times. These pixels would be counted once in the image processing whereas they would be counted twice in Eq. (1), once for each time;

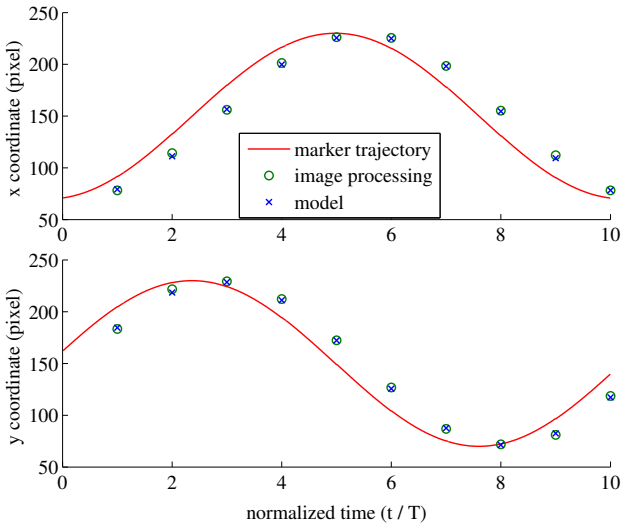

Fig. 3. Trajectories provided by the measurement based on image processing and the simplified model

therefore, violation of assumption A2 yields a shift in the results of the image processing compared with Eq. (1).

Performing a change of variable, relationship (1) rewrites:

$$\int_{t_{k-1}}^{t_k} \overrightarrow{\mathrm{OM}} \sqrt{\dot{x}_M^2(t) + \dot{y}_M^2(t)}\, \mathrm{d}t = l_k \overrightarrow{\mathrm{OG}_k}, \tag{2}$$

with:

$$l_k = \int_{t_{k-1}}^{t_k} \sqrt{\dot{x}_M^2(t) + \dot{y}_M^2(t)}\, \mathrm{d}t. \tag{3}$$

Under assumption A4, the relationship simplifies and the coordinates of G_k write:

$$x_{Gk} = \frac{1}{T} \int_{t_{k-1}}^{t_k} x_M(t)\, \mathrm{d}t, \tag{4}$$

$$y_{Gk} = \frac{1}{T} \int_{t_{k-1}}^{t_k} y_M(t)\, \mathrm{d}t. \tag{5}$$

The measurement provided by image processing can then be considered as the mean value over a sampling period. The remainder of the paper relies on this assumption and is therefore valid for any sensor providing a mean value of a quantity over an integration period.

The measurements provided by this model were computed on a circular trajectory at 5 Hz and compared with the measurements obtained by image processing applied to simulated images. In Fig. 3, exhibiting the obtained results, one can see that the model reproduces nicely the measurements obtained by image processing. The measurement error can be considered as a random variable and the signal-to-noise ratio (SNR) is high.

2.3 Implementable models

Denoting $y(t)$ one entry of the measurement vector containing the coordinates of the markers, image processing, applied to the image available at time $t_k = kT$, provides the measurement w_k that is considered as the average of $y(t)$ on an integration period:

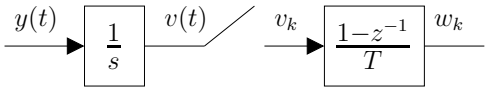

Fig. 4. Model of the camera-based measurement system (first version of Model #1)

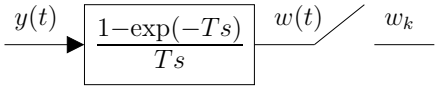

Fig. 5. Second version of Model #1

$$w_k = \frac{1}{T} \int_{t_{k-1}}^{t_k} y(t)\mathrm{d}t. \tag{6}$$

A first model is easily derived by introducing an intermediate variable:

$$v(t) = \int_0^t y(\tau)\mathrm{d}\tau. \tag{7}$$

Denoting $v(z)$ the Z-transform of the sampled-time signal $v_k = v(t_k)$, one can write $w(z) = H(z)v(z)$ with:

$$H(z) = \frac{1 - z^{-1}}{T}. \tag{8}$$

Therefore, the camera-based position measurement can be simulated with three elements as shown in Fig. 4: a continuous-time integrator, a sampler and a discrete-time first-order transfer function. This model will be referred as Model #1 in the sequel.

A similar model can also be derived. Let us introduce a fictitious continuous-time signal:

$$w(t) = \frac{1}{T} \int_{t-T}^{t} y(\tau)\mathrm{d}\tau, \tag{9}$$

such that $w_k = w(t_k)$. Derivation of the equation yields:

$$\dot{w}(t) = \frac{1}{T}(y(t) - y(t-T)). \tag{10}$$

Using the Laplace transform, the equation writes:

$$\dot{w}(s) = \frac{1}{T}(1 - \exp(-Ts))y(s). \tag{11}$$

Therefore, one have $w(s) = \tilde{H}(s)\, y(s)$ with:

$$\tilde{H}(s) = \frac{1 - \exp(-Ts)}{Ts} \tag{12}$$

This latter model, represented in Fig. 5, is equivalent to Model #1 as $z = \mathcal{Z}\mathcal{L}^{-1}(\exp(Ts))$, $\mathcal{Z}$ denoting the Z-transform and $\mathcal{L}$ the Laplace transform. This latter formulation suits for frequency analysis whereas the previous one is convenient for simulation.

2.4 Effects on measurement

The Bode diagram of Model #1 of the camera is presented in Fig. 6 in addition with two other approximated models:

- Model #2 $((1 + z^{-1})/2)$, presented in dashed line, is the mean-value of the position at the present sampling instant and the position at the previous instant. It was used by Ranftl et al. [2007] for controlling a high-speed ultrasonic motor;

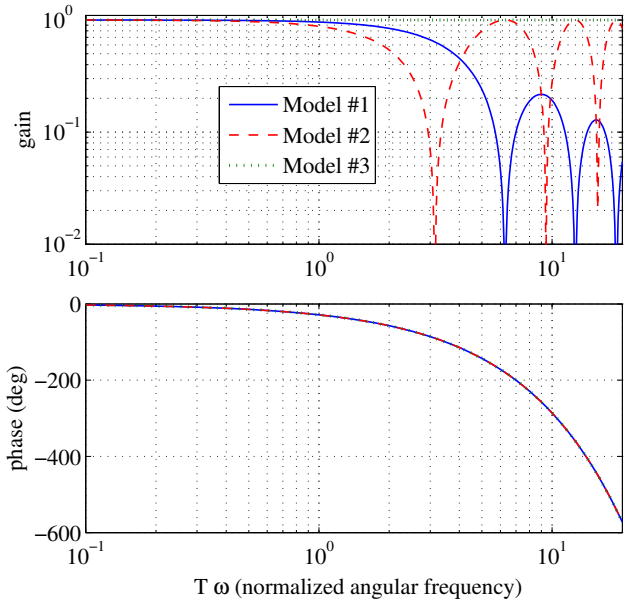

Fig. 6. Bode Diagrams of the models (plain: Model #1; dashed: Model #2; dotted: Model #3

- Model #3 ($\exp(-T\ s/2)$), presented in dotted line, is a delay of half of the period.

One can notice that the effects of the camera are negligible for angular frequencies lower than $\frac{1}{T}$. For higher frequencies, the three models have the same phase but different gains. Model #3 has a constant gain. The gain of Model #2 is zero for frequencies $\frac{1}{T}(\frac{1}{2}+k)$, $k \in \mathbb{N}$ whereas the gain of Model #1 is zero at frequencies $\frac{1}{T}(1+k)$, $k \in \mathbb{N}$. Moreover, Model #2 do not reproduce the gain attenuation at the high frequencies. This gain attenuation is very important as it is not possible to implement anti-aliasing filter; this gain attenuation works as a natural anti-aliasing filter.

3. EFFECTS ON MODEL IDENTIFICATION

3.1 Discrete-time model

When a continuous-time system $G(s)$ is fed with a ZOH and when the data are normally sampled, the equivalent discrete-time model writes:

$$G(z) = \frac{z-1}{z}\mathcal{Z}\,\mathcal{L}^{-1}\left(\frac{G(s)}{s}\right). \tag{13}$$

When, in addition, the measurement is done through a camera, Model #1 allows to derive the discrete-time model of the system:

$$\tilde{G}(z) = \frac{(z-1)^2}{Tz^2}\mathcal{Z}\,\mathcal{L}^{-1}\left(\frac{G(s)}{s^2}\right), \tag{14}$$

which corresponds to the equivalent model with a first-order holder (FOH) and a conventionally sampled measurement.

Consider for instance a second-order model:

$$G_0(s) = \frac{K\omega_0^2}{s^2+2\xi\omega_0 s+\omega_0^2} \tag{15}$$

Its discrete-time model with camera and ZOH, obtained from (14), writes:

$$\tilde{G}_0(z) = z^{-1}\frac{\beta_0+\beta_1 z^{-1}+\beta_2 z^{-2}}{1-\alpha_1 z^{-1}+\alpha_2 z^{-2}} \tag{16}$$

with:

$$\alpha_1 = 2\exp(-\xi\omega_0 T)\cos(wT), \tag{17}$$

$$\alpha_2 = \exp(-2\xi\omega_0 T), \tag{18}$$

$$\beta_0 = K\left(\frac{2\xi}{\omega_0 T}(\alpha_1-1)+1-g\right), \tag{19}$$

$$\beta_1 = K\left(\frac{2\xi}{\omega_0 T}(1-\alpha_1-\alpha_2)-\alpha_1+2g\right), \tag{20}$$

$$\beta_2 = K\left(\alpha_2(\frac{2\xi}{\omega_0 T}+1)-g\right), \tag{21}$$

where:

$$w = \omega_0\sqrt{1-\xi^2} \tag{22}$$

and:

$$g = \left(\frac{2\xi}{\omega_0 T}\cos(wT)+\frac{1-2\xi^2}{wT}\sin(wT)\right)\exp(-\xi\omega_0 T). \tag{23}$$

Therefore, an accurate discrete-time model can be identified without difficulty with one delay and order two at both numerator and denominator [1]. Notice that this model relies on 5 parameters compared to the 3 parameters of the continuous-time model (when accounting for a static unitary gain); therefore, it is not easy to obtain the parameters of the continuous-time model from discrete-time identification. The sequel of this section considers the identification of the continuous-time model $G_0(s)$.

3.2 Continuous-time model identification

Continuous-time model identification algorithms has been developed more recently that their discrete-time counterparts. Nowadays, many algorithms are available in toolboxes such as Contsid [2] and Captain [3]. As their discrete-time counterparts, they consider different kinds of noise models (output error, ARX, ARMAX...). As the data are sampled, it is possible to consider that the input data were obtained either with a ZOH or a FOH; the output being simply sampled. Following the analysis of the previous subsection, we propose to improve the accuracy of the identification results based on data obtained with a ZOH at the input and camera measurement by considering the data as produced with a FOH and a conventional sampling.

In order to evaluate the accuracy that can be reached by identification with the camera, identification of the continuous-time model $G_0(s)$ with $K=1$, $w_0=10$ rad/s and $\xi=0.1$ was processed with the srivc method available in the Contsid toolbox Garnier et al. [2007]. This method is based on instrumental variables and considers an output error model problem; the Poisson filters for estimating the signal derivatives being adjusted iteratively. The input

[1] The System Identification Toolbox, developed by L. Ljung et al. is available with Matlab can be used for discrete-time model identification.
[2] The Contsid toolbox, developed by H. Garnier et al. is available at http://www.iris.cran.uhp-nancy.fr/contsid.
[3] The Captain toolbox, developed by P. Young et al. is available at http://www.es.lancs.ac.uk/cres/captain/

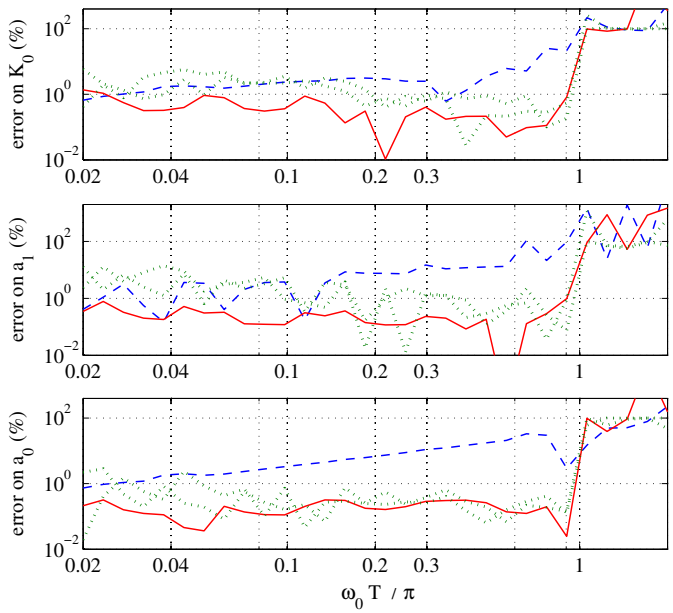

Fig. 7. Variations of the estimation errors (in percent of the actual values) with respect to the sampling period (plain: with camera and FOH (i); dashed: with camera and ZOH (ii); dots: with camera, FOH and noise of SNR=10 (iii, 2 samples))

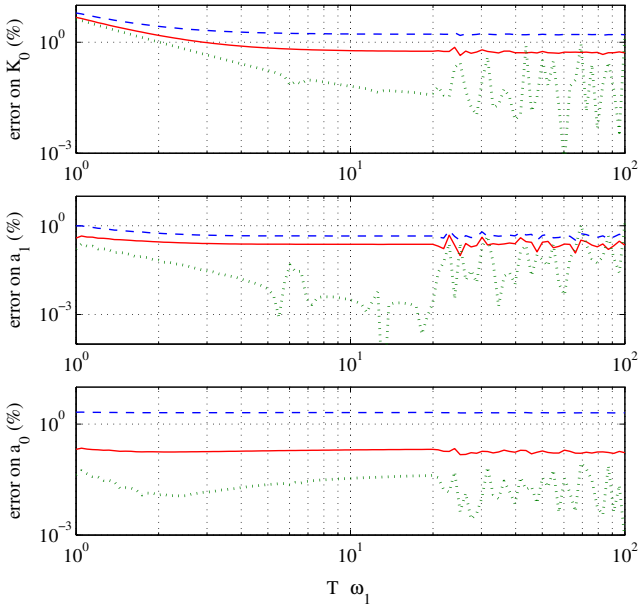

Fig. 8. Estimation errors (in percent of the actual values) due to high frequency dynamics (plain: with camera and FOH; dashed: with camera and ZOH; dots: without camera

excitation was chosen as a pseudo-random binary sequence of order 9 and length 511, held contant over a sample period of length T (ZOH). The output measurements are simulated with the model of Fig. 4.

Identification was processed in three configurations:

i. by considering the data as provided by a FOH in order to account for the camera;
ii. by considering the data as provided by a ZOH;
iii. in the same case as in (i) but with an output additive noise with SNR ratio of 10; two noise samples being considered.

In order to analyse the impact of the sampling frequency with respect to the eigen-frequencies of the system, identification was done for different sampling periods T. The model being rewritten:

$$G_0(s) = \frac{K}{s^2 + a_1 s + a_0}, \tag{24}$$

the relative errors of the three parameters $K = 1$, $a_1 = 2\xi\omega_0$ and $a_0 = \omega_0^2$ are given in percentage in Fig. 7. One can see that if the Shannon condition is satisfied (i.e. $T\omega_0 < \pi$), it is possible to obtain estimation results with good accuracy (less than 1% without noise and arround several percents with noise, provided that identification is done with the FOH argument. Indeed, when neglecting the effects of the camera and considering data as obtained with ZOH, a significant shift appears, mainly on $a_0 = w_0^2$; this shift increases as the sampling period T increases.

3.3 Effects of high-frequency dynamics

When identifying a model of a flexible structure, the identified model will account for the modes at frequencies lower than the Shannon frequency and the higher modes will be neglected. This will result in identification errors on the identification model. In order to evaluate the errors caused by the high frequency modes, tests were processed with a model of order 4:

$$G(s) = G_0(s)G_1(s) \tag{25}$$

where $G_0(s)$ is given in (15) and contain the low frequency part of the system and:

$$G_1(s) = \frac{\omega_1^2}{s^2 + 2\xi_1\omega_1 s + \omega_1^2} \tag{26}$$

contains the high frequency part of the system. With $T = 20$ ms and $\xi_1 = 0.1$, identification is done with the same input signal as in the previous case for different values of ω_1. For comparison purposes, three cases are considered, the results being given in Fig. 8:

i. measurements done by camera and identification data considered as obtained from a FOH (plain line);
ii. measurements done by camera and identification data considered as obtained from a ZOH (dashed line);
iii. conventional measurement (no camera) and identification data considered as obtained from a ZOH (dotted line).

One can see that $G(s)$ can be identified with good accuracy (relative accuracy of 1% on the parameters) with a camera, provided that $T\omega_1$ is large enough. Significant improvements are obtained when using the FOH argument instead of the ZOH argument, specially for parameter $a_0 = \omega_0^2$.

4. CONCLUSION

An original model was proposed in this paper in order to account for the dynamical effects induced by a vision-based measurement device. We have shown that the effects of the camera on the measurement is similar to the effects of a ZOH on the input of a system. The different assumptions on which this model relies were also exhibited. When

processing identification of a continuous-time model with input data produced by a ZOH and measurement provided by vision, it is necessary to choose the FOH argument in the identification method, in order to obtain accurate results.

Based on this model, we were able to show that it is possible to identify a continuous-time model of a system with eigenfrequencies close to the Shannon frequency with good accuracy provided that the data are considered as obtained with a FOH instead of a ZOH. It is also possible to identify the lower-frequency part of a system containing high-frequency dynamics even if the high-frequency modes are relatively close to the Shannon frequency.

REFERENCES

C. Albitar, P. Graebling, and C. Doignon. Design of a monochromatic pattern for a robust structured light coding. In *IEEE Int. Conf. Image Process. ICIP'07*, San Antonio, USA, September 2007.

R.L. Andersson. Dynamic sensing in a ping-pong playing robot. *IEEE Trans. Robot. Autom.*, 5(6):728 – 739, 1989.

W. Bachta, P. Renaud, E. Laroche, J.Gangloff, and A. Forgione. Progress towards a robotized heart stabilizer design and control. In *Int. Conf. Intelligent Robots and Sytems (IROS)*, San Diego, Ca, USA, 2007.

M. Brown, A. Majumder, and R. Yang. Camera-based calibration techniques for seamless multiprojector displays. *IEEE Trans. Visual. Comput. Graphics*, 11(2): 193 – 206, 2005.

P.I. Corke. *Visual control of robots*. Research Studies Press Ltd., Taunton, Somerset, U.K., 1996.

R.T. Fomena and F. Chaumette. Visual servoing from spheres with paracatadioptric cameras. In *Int. Conf. on Advanced Robotics, ICAR'07*, Jeju, Korea, August 2007.

J. Gangloff and M. de Mathelin. High speed visual servoing of a 6 DOF manipulator using multivariable predictive control. *Advanced Robotics*, 17(10):993–1021, 2003.

H. Garnier, M. Gilson, P.C. Young, and E. Huselstein. An optimal IV technique for identifying continuous-time transfer function model of multiple input systems. *Control Engineering Practice*, 15(4):471–486, 2007. doi: 10.1016/j.conengprac.2006.09.004.

R. Ginhoux, J. Gangloff, M. de Mathelin, L. Soler, M. Arenas Sanchez, and J. Marescaux. Active filtering of physiological motion in robotized surgery using predictive control. *IEEE Trans. Robot.*, 21(1):67–79, 2005.

S. Hutchinson, G. D. Hager, and P. I. Corke. A tutorial on visual servo control. *IEEE Trans. Robot. Autom.*, 12 (5):651–670, 1996.

A. Krupa, J. Gangloff, C. Doignon, M. de Mathelin, G. Morel, J. Leroy, and L. Soler. Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. *IEEE Trans. Robot. Autom.*, 19(5):842 – 853, 2003.

C.A. Luna, M. Mazo, J.L. Lazaro, J.F. Vazquez, J. Urena, S.E. Palazuelos, J.J. Garcia, F. Espinoza, and E. Santiso. Method to measure the rotation angles in vibrating systems. *IEEE Trans. Instrum. Measurement*, 55(1): 232–239, 2006. doi: 10.1109/TIM.2005.860868.

F. Nageotte, C. Doignon, M. de Mathelin, P. Zanne, and L. Soler. Circular needle and needle-holder localization for computer-aided suturing in laparoscopic surgery. In *SPIE Int. Symp. Medical Imaging*, USA, San Diego, February 2005.

Y. Nakabo, M. Ishikawa, H. Toyoda, and S. Mizuno. 1 ms column parallel vision system and its application of high speed target tracking. In *Int. Conf. Robot. Autom.*, volume 1, pages 650 – 655, 2000.

A. Ranftl, L. Cuvillon, J. Gangloff, and J. van der Sloten. High speed visual servoing with ultrasonic motors. In *Int. Conf. Robot. Autom. (ICRA)*, April 2007.