Using image recognition to automate video analysis of physical processes

Jaime Riera,[a] Juan A. Monsoriu, and Marcos H. Giménez
Departamento de Física Aplicada, Universidad Politécnica de Valencia, E-46071 Valencia, Spain

José L. Hueso and Juan R. Torregrosa
Departamento de Matemática Aplicada, Universidad Politécnica de Valencia, E-46071 Valencia, Spain

(Received 7 October 2002; accepted 1 April 2003)

Multimedia technologies and video analysis allow the design of low-cost physical measurement and data acquisition systems. The number of pixels in the images limits their spatial resolution, whereas their time resolution depends on the number of frames recorded per second. Both characteristics are determined by the video recording system used. We have developed a laboratory system with these characteristics where important improvements have been reached by using image recognition to automate video analysis. In the present work we first examine several image recognition techniques and evaluate them from the point of view of their application to measurement systems. Then we describe the proposed system and the methodology followed in the measurement process. An image of an experimental object is recorded and used as a filter, while the sequence of images of that object in motion become the input frames for the recognition process. Finally, we discuss the results obtained by this measurement process and compare them with those obtained by using traditional measurement techniques.
[DOI: 10.1119/1.1578066]

I. INTRODUCTION

New communication and information technologies allow the investigator to solve problems in easier and cheaper ways than traditional methods. In particular, digital simulation extends and improves the efficiency of earlier analog simulations.[1,2]

We have developed a laboratory measurement system based on image digitization. The system ''captures'' the physical phenomenon, uses its digitized image as a basis for qualitative analysis, and performs quantitative analysis by means of dedicated software that extracts the measurements starting from the points clicked by the user.[3] This process is in some way the inverse of that of digital simulation: instead of creating a digital image that simulates reality, we analyze the image of the real process in order to extract physical parameters. From the didactic point of view, the system tries to integrate three ideas: real experience, qualitative models, and quantitative models, reducing the ''distance'' between the experiment and its mathematical description. Recently, articles that use this technique have been published in this journal.[4–6]

In order to automate the measurements and improve the efficiency of the system, we have turned to image recognition techniques. The idea of recognition and analysis of an object from its digital image has a remote antecedent in the VanderLugt correlator[7] based on optical filtering in the Fourier domain, or in the adjoint transform correlator that uses optical processing in the image plane.[8] In both cases, the method is based on the correlation between a target image of the object to be detected and the input scene that contains that object. The correlation provides a measurement of the similarity in terms of the mean quadratic difference between the correlated patterns. One shortcoming of this method, however, is that it is very sensitive to deformations of the detected object.

It would be desirable in some applications to find methods providing distortion invariant recognition, especially rotation and changes in scale. Solutions have been proposed, based on the Mellin transform or on the orthogonal decomposition of the function representing the target image.[9]

Among the several possibilities for object detection, we choose a method that provides the required resolution level within the available computation time. The methodology we propose consists of performing physical experiments which are easy to control, and employs a relatively simple detection algorithm, and then applies the developed method to situations with higher noise level.

In this paper we analyze the results obtained by applying our system to laboratory experiments where conditions are nearly ideal so the recognition algorithms can be simple. We have used standard linear correlation[7] as a basis for the detection technique due to its easy implementation and very low computational cost. As an alternative technique, we apply nonlinear correlation based on the binary decomposition SONG (Sliced Orthogonal Nonlinear Generalized)[10] algorithm that is applicable to gray-scale images.

II. DESCRIPTION OF THE MEASUREMENT SYSTEM

The measuring process consists of four steps: (a) the images of the experiment are recorded by means of a video camera, (b) the analog signal of the camera is digitized and transferred to a digital video file in the computer, (c) the object is detected in each image, and (d) the information is analyzed and processed by means of suitable software.

A. Time variable

Video recording systems provide information at a regular time rate. The output of the video camera consists of a series of snapshots taken at regular time intervals that can then be examined to reveal the qualitative characteristics of some physical phenomenon. This qualitative analysis can be refined to a quantitative one. Modern multimedia technologies allow us to digitize the analog video signal and transfer it to the computer as a digital file consisting of a sequence of still pictures.

The digital camera used in the experiments described in this paper was a Panasonic NV-DS15EG, with an exposure time of 1/2000 s and rate of 25 frames/s providing a time resolution of 0.04 s. If greater time resolution is required, then the two fields that integrate the frame (the one with the odd-numbered lines and the one with the even-numbered ones) can be analyzed before the image is reconstructed, providing a resolution of 0.02 s.

As a test for the possible variability of the time interval, we recorded images of the display of a digital counter. The detected error was negligible, less than 1 μs. The use of long exposure times (greater than 1/100 s) may be the reason why some authors[6] report different results.

B. Space variable

The location of each object is determined from its position in the frame. The coordinate origin is placed in the upper left corner of each frame. The pixel (the smallest piece of information of the graphical image) is taken as the unit of length.

The European standard PAL (Phase Alternation by Line) system generates 720×576 pixel images. We have taken 512×512 pixel windows for the information analysis, as most of the image recognition techniques require using matrices whose dimensions are integer powers of 2.

In order to obtain the scale factor, we use a reference image of known size and determine its length in pixels. This reference must be placed in the same plane in which the movement takes place to achieve the desirable accuracy. In Sec. IV we explain in more detail how to obtain the scale factor in each situation.

III. IMAGE RECOGNITION TECHNIQUES AND EXPERIMENT DESIGN

The recognition of shapes in images allows us to determine if an input scene contains a specified target object. This method is widely used in many automatic processes, such as control processes, bank security, and robotics. Most of the methods for recognition of two-dimensional objects are based on the correlation between an input scene, $f(x,y)$, and a target object, $h(x,y)$. This correlation is given by

$$g(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x',y')h(x'-x,y'-y)dx'dy'. \quad (1)$$

Therefore, the maximum of the correlation corresponds to the location where the scene is most similar to the target object.

When working with digitalized images, correlation involves double integrations that require much computation time. However, the process is substantially simplified by using the properties of Fourier transforms. It can be easily proved that[11]

$$g(x,y)=F^{-1}\{F\{f(x,y)\}F\{h(-x,-y)\}\}. \quad (2)$$

Therefore, the correlation can be calculated relatively rapidly by means of direct and inverse Fourier transforms based on fast Fourier transform algorithms. The analysis software has been developed in VisualBasic in our laboratory. It runs on Windows platforms.

As a more discriminatory alternative, we have done a binary decomposition of the image in gray levels. We use the decomposition called SONG.[10] From this decomposition, every image in gray levels can be defined as:

$$f(x,y)=\sum_{q=1}^{Q} q e_q(f(x,y)) \quad (3)$$

where $Q=255$ and

$$e_q(f(x,y))=\begin{cases}1 & \text{if } f(x,y)=q\\ 0 & \text{if } f(x,y)\neq q.\end{cases}$$

After binary decomposition, we obtain a linear correlation between the input scene and the object of reference for each gray level. The final result will be the sum of these correlations, which gives rise to nonlinear correlations.

For testing the system, we have designed two experiments with a high degree of control of the variables: (a) the displacement of an object at uniform speed and (b) the fall of an object.

In both cases the video camera was at a distance of 1.5 m and its axis orthogonal to the plane of movement.

The sequence of recorded images was digitalized by using commercial software (Pinnacle Systems, DV500) and stored in AVI files for visualization and RAW binary files for analysis. The RAW files allow us to have the gray level of each pixel in different frames between 0 and 255 (8 bytes).

In order to detect the position of the moving objects, we take each frame as an input scene and use a detached image as the filter for the object of interest. Following the above recognition techniques, we have obtained for each pixel of the frame the correlation between the frame itself and the filter centered to the coordinates (x,y) of that pixel. The representation of the correlation values versus the position (x,y) provides a peak detection, i.e., a function with a maximum located at a target object.

The high degree of control we have on the experimental conditions avoids distortions due to scale changes or rotations and minimizes the presence of noise. In consequence, we do not need high precision in the detection algorithm and so, we can use elementary detection techniques based on linear correlations [Eq. (2)] between the filter and each frame. This results in a methodology easy to implement and with low computational cost.

Although these two experiments are one-dimensional, we analyze the problem in two dimensions in order to approximate the actual conditions of the motion and prepare the system for more complex experiments. We perform the analysis on 512×256 pixel windows for the first experiment, and 128×512 for the second one, thus providing a good compromise between information level and computational cost.

Afterwards, we confirm these results by using nonlinear correlation techniques, with more discrimination power. These techniques decompose the filter and the input images in gray levels, linearly correlate each level, and then add the resulting contributions.[10]

The object measurements obtained with the linear correlation technique are very similar to that of the nonlinear correlation technique, in spite of having performed 255 correlation computations in the second case and only one in the first case.

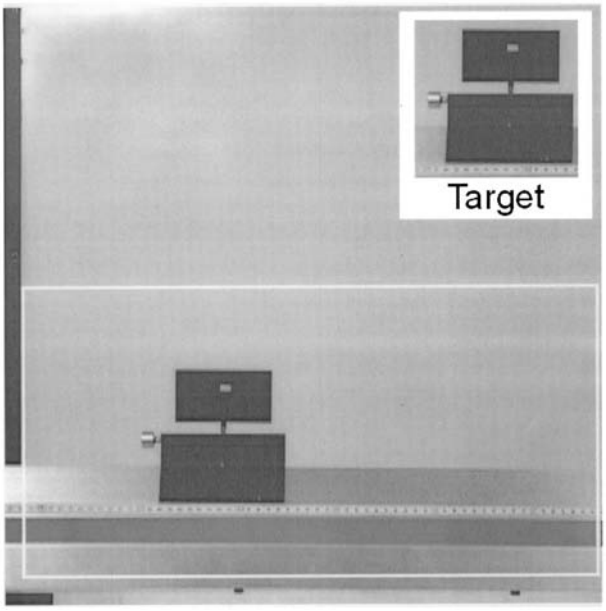

Fig. 1. Input scene, where the detection window has been framed. The object to be detected is shown in the upper-right corner.

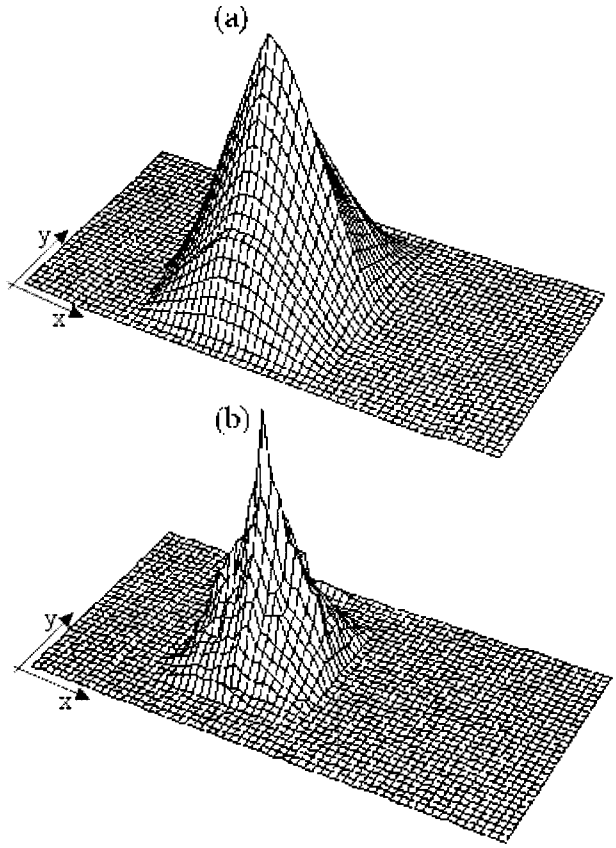

Fig. 2. (a) Linear correlation, and (b) nonlinear correlation between the input scene and the object to be detected shown in Fig. 1. The axes x and y represent the pixel position in the detection window.

IV. RESULTS

A. Displacement of an object at uniform speed

In our first experiment we have recorded the movement of a glider at uniform speed on an air track. Figure 1 shows one input scene and the object to be detected, from which the filter is obtained. It also shows in the frame the detection window that in this experiment has 512×256 pixels. The gray level i has been transformed into the level $(255-i)$ so that the functions $f(x,y)$ and $h(x,y)$ appearing in Eq. (2) will have large values for the dark object.

Figure 2 shows the detection peaks obtained from one frame by using the technique of (a) linear correlation and (b) nonlinear correlation. The second method gives a sharper peak and therefore better resolution of the glider position. The linear correlation technique produces a broader peak due to the noise, in spite of using negative images. However, the results are very similar, within an error of one pixel.

Figure 3 shows the least squares line that fits these data. The coefficients of the equation, also shown in the figure, provide the estimated value of the initial position and the velocity in pixels per second. The velocity in SI units is obtained by applying the scale factor, (499 ± 1) pixels$\Leftrightarrow(590\pm1)$ mm, which is obtained by measuring the distance between two references on the scale graded in mm placed on the track. The fit presents a high degree of correlation and the result, $\nu=(0.641\pm0.004)$ m/s, agrees with that obtained by using a Phywe photogate timer with resolution of 1 ms, within the experimental error.

The parameters and their errors have been obtained according to the standard procedure of minimization of the χ^2 (chi-square) merit function, defined as

$$\chi^2=\sum_i \frac{(Y_i^{\mathrm{exp}}-Y_i^{\mathrm{mod}})^2}{\sigma_i^2} \tag{4}$$

and by assigning an error of 1 pixel to the ordinate values. Y_i^{exp} represents experimental values that are obtained with individual standard deviations σ_i, and Y_i^{mod} are the values predicted on the basis of a theory or a model. If the experimental data are "normally" distributed, and if σ_i is the correct measure for the quality of these data, then χ^2 is a measure for the quality of the fit.

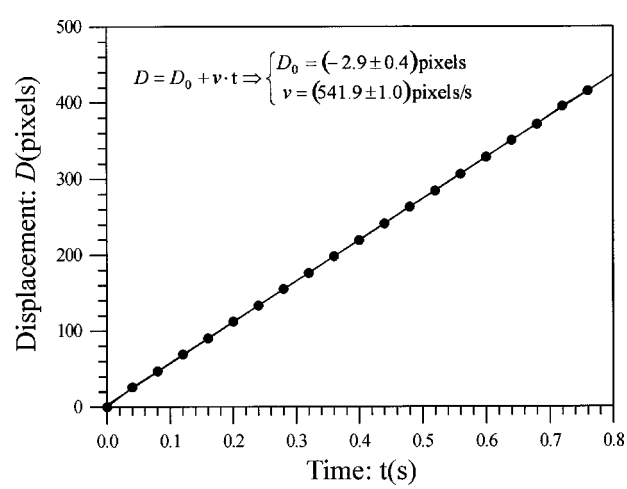

Fig. 3. Displacement vs time representation of the experimental data, and the least squares fit to the data.

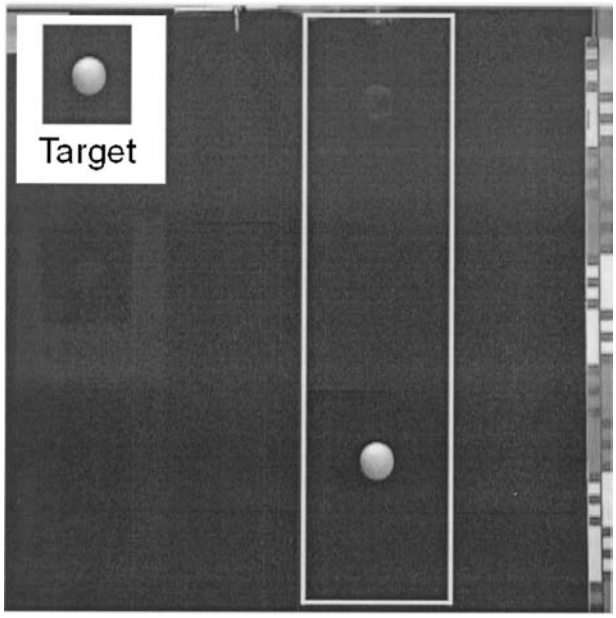

Fig. 4. Input scene, where the detection window has been framed. The object to be detected is shown in the upper-left corner.

B. Movement of a falling object

As the second test of our measurement system, we analyze the movement of a falling ball. The process is analogous to that of the previous experiment. We have first recorded the fall and then processed the images in order to detect the position of the object. Figure 4 shows an input frame and the image of the object to be detected from which the filter is constructed. The scale factor is computed from the reference scale shown on the right of the image. The detection window is shown by a rectangle whose size is 128×512 pixels in this experiment.

Figure 5 shows the detection peaks obtained from an image by using the above-mentioned techniques. As in the former experiment, the techniques based on linear correlation give satisfactory results with the lowest computational cost. The experimental data are represented in Fig. 6 where space is measured in pixels and time in seconds. In this case we fit a second degree curve by the techniques referred to in the former section to get the parameters of the motion.

These parameters measured in pixels are also shown in the figure. The acceleration, computed by applying the appropriate scale factor, (412 ± 1) pixels$\Leftrightarrow(450\pm1)$ mm, obtained from a vertical rule near the ball, is $a=(9.77\pm0.11)$ m/s^2.

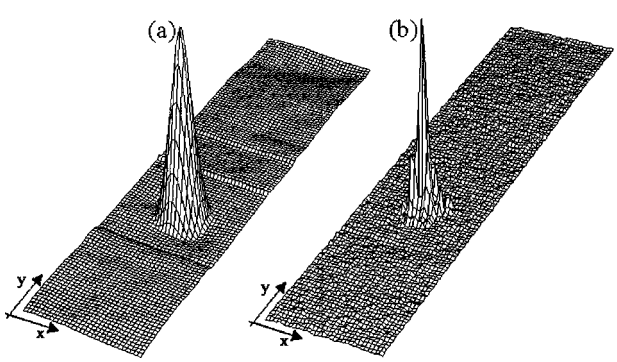

Fig. 5. (a) Linear correlation, and (b) nonlinear correlation between the input scene and the object to be detected shown in Fig. 4. The axes x and y represent the pixel position in the detection window.

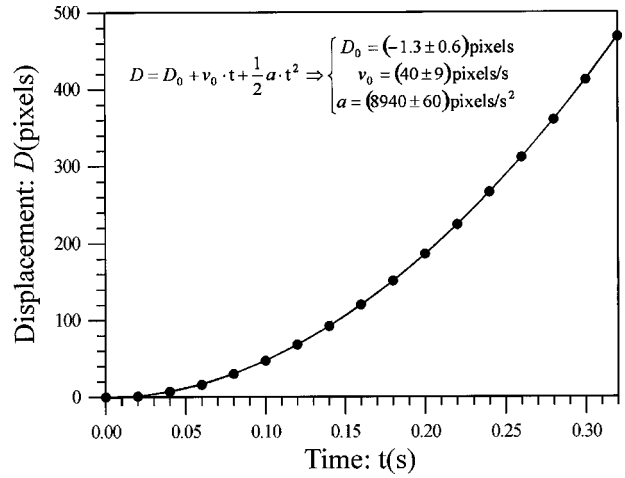

Fig. 6. Displacement vs time representation of the experimental data, and the least squares parabolic fit to the data.

C. Evaluation of results

The results obtained in both experiments using linear correlation recognition techniques agree with those obtained from traditional devices which may be less versatile but more expensive. Moreover, the measurement system associated with our multimedia interface optimizes the pedagogical efficiency of the experiments, because the use of commercial video analysis tools increases the reliability of the measurement process.

Besides its didactic application, the system offers good prospects to be used in technological processes and in experiments with less control of the parameters. However, in some cases, more complex image recognition algorithms will be needed. The use of scale invariant detection techniques will allow the application of generic filters to different experiments.

ACKNOWLEDGMENTS

This work has received financial support from the Generalitat Valenciana (Grant No. GV99-182-1-01) and from the Universidad Politécnica de Valencia (P.I.I. 20000591/P.I.I. 20020632), Spain.

[a]Electronic mail: jriera@fis.upv.es

[1]M. Webb, "Computer-based modelling in school science," School Sci. Rev. **74** (269), 33–47 (1993).

[2]A. Vidaurre, J. Riera, M. H. Giménez, and J. A. Monsoriu, "Contribution of digital simulation in visualizing physics processes," Comp. Appl. Eng. Educ. **10** (1), 45–49 (2002).

[3]J. Riera, M. Boscá, M. H. Giménez, A. Vidaurre, and R. Iñiguez, "Analysis of physical phenomena by means of the digitalization of the image of the process," Proceedings EFING2000 (CD-ROM), La Habana, Cuba, 2000.

[4]W. M. Wehrbein, "Using video analysis to investigate intermediate concepts in classical mechanics," Am. J. Phys. **69**, 818–820 (2001).

[5]R. Cross, "Measurements of the horizontal coefficient of restitution for a superball and a tennis ball," Am. J. Phys. **70** (5), 482–489 (2002).

[6]E. J. Salumbides, J. Maristela, A. Uy, and K. Karremans, "A vision-based motion sensor for undergraduate laboratories," Am. J. Phys. **70** (8), 868–871 (2002).

[7]A. VanderLugt, "Signal detection by complex spatial filtering," IEEE Trans. Inf. Theory **IT-10**, 39–145 (1964).

[8]C. S. Weaver and J. W. Goodman, ''A technique for optically convoluting two functions,'' Appl. Opt. **5**, 1248–1249 (1966).
[9]E. W. Hansen and J. W. Goodman, ''Optical reconstruction from projections via circular harmonic expansion,'' Opt. Commun. **24**, 268–272 (1978).
[10]P. Garcia-Martínez and H. H. Arsenault, ''A Correlation Matrix Representation using Sliced Orthogonal Non-linear Generalized Decomposition,'' Opt. Commun. **174**, 503–515 (2000).
[11]J. D. Gaskill, *Linear Systems, Fourier Transforms and Optics* (Wiley, New York, 1978), p. 196.