8º Congresso Nacional de Mecânica Experimental
Guimarães, 21-23 de Abril, 2010

LONG DECK SUSPENSION BRIDGE MONITORING: THE VISION SYSTEM CALIBRATION PROBLEM

Santos, C.A.[1]; Costa, C.O.[2]; Batista, J.P.[3]

[1] Scientific Instrumentation Centre, National Laboratory for Civil Engineering
2 Scientific Instrumentation Centre, National Laboratory for Civil Engineering
3 Department of Electrical Engineering and Computers, University of Coimbra



ABSTRACT

The Structural Health Monitoring (SHM) is an emergent powerful diagnostic tool, which can be used to identify and to prevent possible failures of the various components that comprise the infrastructure. In the particular case of a suspension bridge, the measurement of the vertical and the transversal displacements plays an important role for its safety evaluation. Taking into account the restrictions usually found on these structures, an enhanced solution comprises a non-contact measurement system, specifically a vision-based measurement system. The paper describes a methodology to perform the vision system calibration which can be carried out in-situ, while the deck is moving, and requires little effort and a minimum set of information. Results related to the performance evaluation, obtained by simulation, are presented and they show that even in an environmental severely affected by noise it is possible to obtain a standard accuracy better than 10 mm.

1 - INTRODUCTION

1.1 - Background

In Civil Engineering, to design an infrastructure with a comfortable degree of safety at minimal costs is a great challenge. Usually, the design is performed taking into account information that does not represent completely the truth state, since there exist always a certain level of uncertainty about the materials properties as well as the constructions techniques. Actually, the application of new materials in innovative designs is an issue which implies some risk and the differences in construction are one of the larger influences on how the structure deforms. Likewise, the materials degradation along the time as well as a drastic change in the operation conditions (e.g. earthquakes, excessive load) can drive the structure to an abnormal behaviour.

Structural Health Monitoring (SHM) is a powerful diagnostic tool which can be a good help to overcome most of those issues. Specifically, SHM can be used to identify and to prevent failures of the multiple components that comprise the infrastructure, to improve codes for the construction, to calibrate the mathematical models and to optimize the infrastructure design. Without an efficient monitoring system, a component failure can cause irreversible damages in the infrastructure and, eventually, a possible loss of human lives. SHM is the key to increase the confidence of infrastructure owners and

users, and to protect it from serious damages.

To achieve these goals, it is necessary to measure several quantities, some of them related to the external actions that act over the structure (e.g. wind, load) and others related to the respective structure response (e.g. deformation, displacement). In the particular case of a long span suspension bridge, the knowledge about the vertical and the transversal displacements plays an important role for the structure safety assessment, namely at the middle span where, usually, these are more significant and can reach a couple of meters. However, in this case, the traditional displacements transducers and measurement techniques cannot be used since there is not a fixed point in the neighbourhood of the part of the deck to be monitored. Moreover, as said before, the displacements of the deck can reach a high amplitude (more than one meter), which makes them impracticable for this kind of application. A common solution is to measure the acceleration, and sometimes the velocity, and estimate the displacement by mathematical integration. Though, this solution has many drawbacks since the degree of accuracy depends on various factors, such as the sampling rate, the data record length, the drift and the offset of the electrical signal, etc. All these perturbations that are disturbing the data are also integrated.

Considering the restrictions usually found in this kind of structure, an enhanced solution is to use a non-contact measurement system with dynamic response, accuracy and amplitude range well-suited to the physical phenomenon to measure. By now, only a few measurement systems satisfies these requirements, albeit partially. This set, basically, includes the systems based on: i) GPS (*Global Position System*); ii) radar technology; and iii) optical devices (vision metrology).

The measurement systems based on GPS are established in the large structures from long time ago (Nakamura 2000; Çelebi et al. 2002; Brown et al. 2006). Currently, using differential positioning techniques (Real Time Kinematic – RTK), these systems can provide measurements with accuracy of 10 mm in the horizontal direction and 20 mm in the vertical direction at a sampling rate of 20 Hz (Brown et al. 2006). However, these systems are not suitable to measure the displacements of the deck bridge because they are particularly sensitive to the signal multipath effects and, as consequence, the position accuracy is deteriorated. In the particular case of a suspension bridge, where the steel is the main material, the signal multipath is amplified significantly by the deck, cables and towers. Moreover, the GPS signal, and consequently the accuracy, is also affected by the propagation delays induced by dry air, water vapour and other particles (Solheim and Vivekanandan 1999).

Another non-contact measurement system is based on the radar principle (Bernardini et al. 2007). The system sends a microwave signal and waits for an echo returned by the obstacles (targets) found in the propagation path. The time propagation is related with the straight distance between the target and the measurement system. Using a differential interferometric technique, one measurement system is enough to measure the distance to multiple targets. Nonetheless, to identify clearly the targets, the minimum distance between any two targets in the direction of the signal propagation must be higher than 0.5 m. Otherwise, the reflected signal from the targets will appear scrambled at the receiver, making impossible to identify the contribution of each target. This type of system can operate at 500 m of distance, with an accuracy ranging between 10 μm and 100 μm at 40 Hz of sampling rate. However, taking into account the material type and the distance between the trusses that comprise the deck bridge, it is easy to conclude that this type of system is not suitable for this application. Furthermore, since the radar system measures the distance between the target and the measuring unit, in the direction of the signal propagation (radial), from the measurement data (1D) it is not possible to

discriminate the contribution of each displacement component (3D).

Despite of some limitations (e.g. free line of view between the camera and the target), for large structures built with steel, like a deck bridge, the vision-based systems are the solution that gathers more advantageous. In this way, a vision-based measurement system is being developed, that aims to measure the displacements of large amplitude, such as those observed on long span deck suspension bridges, with a standard accuracy of 10 mm, resolution of 1 mm and a minimum sampling rate of 20 Hz.

1.2 - Vision-based systems

One of the first non-contact measurement systems, based on optical devices, was developed at the LNEC, in 1969, and was used to monitor the deck displacements of the bridge 25 de Abril, over the river Tejo, in Lisbon (Marécos et al. 1969). The system comprised a target, fixed on the deck, and an optical system fixed on pier's base. Each target comprised a set of two lamps, powered with low voltage to radiate a near infrared concentrated beam, and spaced by 20'' in vertical and horizontal directions. One of the lamps was used for tracking, while the other was used for the system calibration and for automatic replacement when the first failed. The optical system comprised a lens with 640 mm of focal length, an optical filter and four photodiodes mounted in a bridge configuration. The unbalanced signal generated by the sensor's bridge was feed as input to a servo-controlled motor system, to providing two degrees of rotation (tilt and azimuth) to follow the target. This system was in service almost a decade and was a helpful to study the dynamic bridge behaviour. Due to the poor dynamic response, provided by the moving parts, this system cannot fit the high demands of the new SHM system.

Almost 25 years later, in 1993, comes out a true vision-based system. Taking advantage of the emergent off-the-shelf video cameras and parallel processing capability offered by the transputer technology, Stephen et al. (1993) developed a new measurement system for monitoring the deck's displacements of the Humber bridge (main span with 1410 m). Without moving parts, this system could offer high sampling rate and required less maintenance. The target comprised two concentric reflective rings in a black board and the system calibration was done based on the knowledge of the diagonal target length and the respective length in the image. Considering that the target had no capability to produce radiation this system couldn't operate in the night period or in poor visibility conditions. To overcome this issue, Macdonald et al. (1997), introduces some improvements to the previous system, namely the target illumination with lamps.

Since then, a few vision-based systems were developed to measure the displacements of structures, although only for a brief test purpose and not for continuous way structure health monitoring. Olaszek (1999) proposed a solution, based on a complex optical mechanism, to focus in the same sensor's video camera two targets at different field of view. One of the targets was for reference (stationary) and the other was for monitoring the structure. Thus, combining the two target's information it was possible to eliminate the undesired vibration produced at the camera. Wahbeh et al. (2003) used a target built with two light emitting diodes (LED), to radiate in the red wavelength spectrum, and spaced by 10''. Similarly to the first system mentioned, only one LED was used for tracking, being powered permanently; the other was used for the calibration task. Conversely to the others solutions, the camera was mounted at the centre of the tower structure, at the deck level. However, as the authors reported, this location compromised the data since it was affected by the high vibration produced by the on-going traffic.

Recently, Lee and Shinozuka (2006) and Fukuda et al. (2009) proposed a measurement system which includes a new calibration method and different target geometry. The target comprises four white spots disposed with a known geometry

over a black board. The spot's positions establish a 2D coordinates reference system to calibrate the camera's rotation, albeit partially.

1.3 - Some remarks

Considering the characteristics of the actual vision-based systems, some remarks are justified. They are basically two and are related with the target type and with the calibration methodology. In fact, a passive target limits the measurement system usefulness to the daylight period and the image's contrast is dependent of the visibility conditions. As a result, the image processing is more difficult to execute and the recovering of the target's coordinates is more prone to error.

The calibration method employed by the referred to above vision-based systems is not suitable since, in the majority, it is done using the knowledge of the distance between two points in the target and the respective distance in the image. Actually, apart others sources of error, this method works only if the deck bridge displacements occurs on a plan parallel to the sensor's video camera; otherwise, there is an embedded error associated with the target displacement. It is known that the vertical and the transversal deck displacements have much higher amplitude than the longitudinal, which justifies their importance to the SHM. However, the existence of a small displacement in the longitudinal direction is enough to generate an error on the vertical and/or transversal direction with amplitude higher than the required accuracy value. In fact, as showed on Fig. 1, the longitudinal displacement (z_w) of the deck induces in the camera *C* an apparent displacement in the vertical direction, as given by Eq. (1).

$$y_w = z_w \cdot \frac{T_Y}{T_Z + z_w} \cong z_w \cdot \frac{T_Y}{T_Z} \qquad (1)$$

In the particular case of the bridge 25 de Abril, with $T_Y = 70$ m and $T_Z = 500$ m, the Eq. (1) converts to $y_w \cong 0.14 \cdot z_w$. This means that to satisfy the accuracy required, the longitudinal displacement must be less than 71.4 mm. Wahbeh et al. (2003) fixed the camera at the deck level ($T_Y = 0$) but, as they reported, this was a bad decision. Thus, it is essential to measure the longitudinal displacement to achieve the high level accuracy required. Though, to measure the 3D displacements it is necessary a minimum of two cameras, tracking the same target.

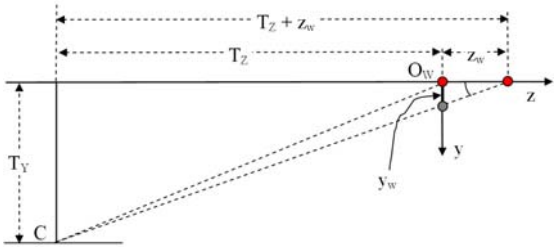

Fig. 1 – The longitudinal displacement of the deck (z_w) induces in the camera *C* an apparent vertical displacement (y_w).

Furthermore, the calibration method employed does not take into account the rotation of the camera related to the deck bridge coordinates system. The deck bridge displacements are measured in the coordinate system of the camera which, certainly, has a different orientation of the deck's coordinate system. One of the systems tried to calibrate the rotation, but only partially, since the calibration is performed on the plan and not in the space.

2 - VISION SYSTEM SETUP

To fulfill the goals mentioned before, namely the high level accuracy, it is essential to perform a genuine vision system calibration which is not carried out by the referred to above vision systems. Another issue is the target type and shape, whereas a good design can be a helpful to recover the image's coordinates of the targets with sub-pixel accuracy and, as a result, to reduce the global error.

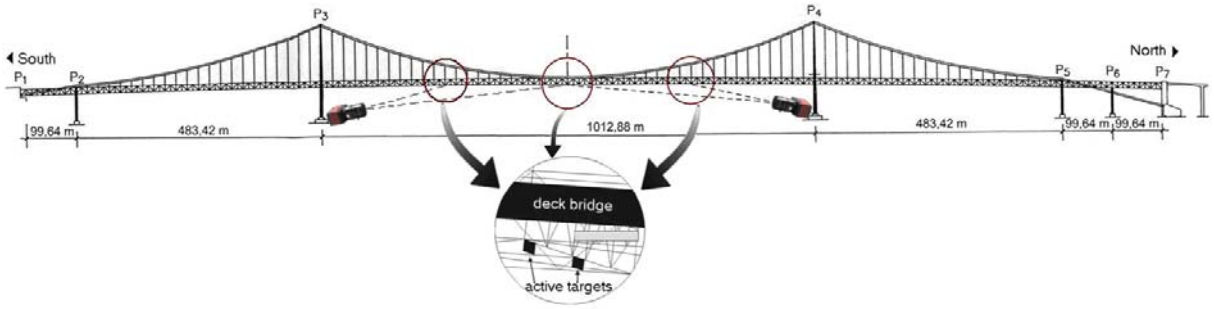

Fig. 2 - Working principle of the vision-based system, applied for measuring the deck displacements of the suspension bridge 25 de Abril over the river Tagus, in Lisbon.

Tacking into account the limitations of the knew vision systems, we are developing a vision-based measurement system for monitoring the displacements of long deck suspension bridges, with high accuracy and to operate in continuous way. The system setup, basically, comprises a set of active targets and a minimum of two industrial digital video cameras, with high pixel resolution, coupled to long focal length lenses (600 mm) (Santos et al. 2009). The target, built with low power LED, radiates a cross shape with near infrared wavelength, in a controlled mode according to the visibility conditions (Santos et al. 2007).

The targets are fixed to the deck (monitoring points) whereas the cameras, focused on the targets, are mounted at the piers base (reference points), as showed in Fig. 2. The deck displacements measurement comprises: i) the images capture from a synchronised set of cameras; ii) the images processing, by digital techniques, to get the coordinates of the targets on each frame with sub-pixel accuracy; iii) the calculation of the targets 3D absolute position, embedding the image coordinates of the targets obtained by all cameras on a triangulation technique; and iv) the calculation of the displacement of the deck bridge between two different time samples.

3 - CALIBRATION METHODOLOGY

3.1 - Affine camera model

Among the various camera models available, the central perspective projection is the most frequently used by the computer vision community. However, according to Hartley and Zisserman (2003), the perspective camera model can be approximated, with slight loss, by an affine camera model if: i) the depth relief is small compared to the average depth, i.e. the points (targets) to be projected in the image plan are almost in the same plan, when compared to the average distance between the camera and the set of points; and ii) the distance of any point to the optical axis is small. These two conditions are usually satisfied when the distance between the camera and the set of points is large, as is the case. Thus, assuming no lens distortion, a 3D point positioned on the deck with homogeneous coordinates $\tilde{P}_i = [X_i, Y_i, Z_i, 1]^T_{i=1,\ldots n}$ is projected into the image plane at the coordinates $\tilde{p}_i = [x_i, y_i, 1]^T_{i=1,\ldots,n}$, according to Eq. (2):

$$\lambda \cdot \underbrace{\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}}_{\tilde{p}_i} = \underbrace{\begin{bmatrix} \tilde{m}_{11} & \tilde{m}_{12} & \tilde{m}_{13} & \tilde{m}_{14} \\ \tilde{m}_{21} & \tilde{m}_{22} & \tilde{m}_{23} & \tilde{m}_{24} \\ 0 & 0 & 0 & \tilde{m}_{34} \end{bmatrix}}_{\tilde{M}} \cdot \underbrace{\begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}}_{\tilde{P}_i} \quad (2)$$

where λ is a non zero scale factor and $\tilde{M}$ represents the camera projection matrix. The matrix $\tilde{M}$ can be decomposed into intrinsic and extrinsic matrix parameters, as showed by Eq. (3):

$$\tilde{M} = \begin{bmatrix} K_{2x2} & 0 \\ 0^T & 1 \end{bmatrix} \cdot \begin{bmatrix} R_{2x3} & T_{2x1} \\ 0^T & 1 \end{bmatrix} \quad (3)$$

where K represents the intrinsic parameters (focal length, skew, aspect ratio and principal point coordinates), R is the rotation matrix and T the translation vector. Making a transformation such as

$m_{ij} = \frac{\widetilde{m}_{ij}}{\widetilde{m}_{34}}$ ($\lambda = m_{34} = 1$), Eq. (2) can be rewritten as:

$$\underbrace{\begin{bmatrix} x_i \\ y_i \end{bmatrix}}_{p_i} = \underbrace{\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix}}_{M} \cdot \underbrace{\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix}}_{P_i} + \underbrace{\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}}_{t} \quad (4)$$

Eq. (4) is a linear mapping, in inhomogeneous coordinates, composed with a translation, whereas $t = [t_1 \quad t_2]^T$ is the image of the origin of the deck reference system (in pixels). Thus, Eq. (4) can simply be written as:

$$\underbrace{\begin{bmatrix} x_i - x_0 \\ y_i - y_0 \end{bmatrix}}_{\Delta p_i} = \underbrace{\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix}}_{M} \cdot \underbrace{\begin{bmatrix} X_i - X_0 \\ Y_i - Y_0 \\ Z_i - Z_0 \end{bmatrix}}_{\Delta P_i} \quad (5)$$

$$\Delta p_i = M \cdot \Delta P_i$$

where $p_0 = [x_0 \quad y_0]^T$ is the image of the point $P_0 = [X_0 \quad Y_0 \quad Z_0]^T$. Thus, one can choose p_0 and P_0 as the origin of the coordinates system in the image and in the deck, respectively.

3.2 - Calibration methodology

The vision system calibration aims to find the parameters for the mathematical model (M) of each camera that matches the projection of the 3D points coordinates into the respective 2D image points coordinates.

With the proposed calibration methodology it is not necessary to know the absolute 3D coordinates of the targets and it is possible to use several frames per camera viewing a reduced set of control points[1]. Additionally, the calibration can be performed automatically and in an autonomous mode. Though, to allow Euclidean metric measurements, the distance between any two neighbour targets is required. The measurement of the distance between the targets can easily be performed with cheap measurement tools.

The methodology developed is based on the factorization method proposed by Tomasi and Kanade (1992).

Let's assume m control points, describing an object shape, that are fixed on the deck bridge and all are viewed by the n cameras. Further, let's assume that each camera have captured q frames (synchronously) while the deck bridge was moving and the region of the deck where the set of targets were fixed behaves like a rigid body.

Since in an affine camera the centroid of a set of 3D points is mapped to the centroid of their image projections (Hartley and Zisserman 2003) and the relative position of each 3D point in the object does not change over time, we can obtain an *average frame* per each camera, as represented by Eq. (6), which preserves the image of the object.

$$\underbrace{\begin{bmatrix} \hat{x}_{11} & \cdots & \hat{x}_{1m} \\ \hat{y}_{11} & \cdots & \hat{y}_{1m} \\ \vdots & \ddots & \vdots \\ \hat{x}_{n1} & \cdots & \hat{x}_{nm} \\ \hat{y}_{n1} & \cdots & \hat{y}_{nm} \end{bmatrix}}_{\hat{W}} = \frac{1}{q} \cdot \begin{bmatrix} \sum_{i=1}^{q} x_{i11} & \cdots & \sum_{i=1}^{q} x_{i1m} \\ \sum_{i=1}^{q} y_{i11} & \cdots & \sum_{i=1}^{q} y_{i1m} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{q} x_{in1} & \cdots & \sum_{i=1}^{q} x_{inm} \\ \sum_{i=1}^{q} y_{in1} & \cdots & \sum_{i=1}^{q} y_{inm} \end{bmatrix} \quad (6)$$

where (x_{ijk}, y_{ijk}) represents the coordinates of the target k in the frame i captured by the camera j. The image coordinates of each point, in the average frame, matches the projection of the 3D centroid of the points cluster, created by the motion of that point while being captured. The matrix $\hat{W}$ is called the *measurement matrix*.

The effect produced by averaging the set of frames is a reduction of the noise effects which is assumed to be Gaussian with zero mean and frame independent. Since the principal point is not defined in an affine camera, we can select the origin for the image coordinate system as the centroid of its 2D projections, obtained as:

[1] Henceforth, it is supposed that the targets coordinates are know and consequently we will refer the target as a point.

$$\begin{bmatrix} \bar{x}_1 \\ \bar{y}_1 \\ \vdots \\ \bar{x}_n \\ \bar{y}_n \end{bmatrix} = \frac{1}{m} \cdot \begin{bmatrix} \sum_{i=1}^{m} \hat{x}_{1i} \\ \sum_{i=1}^{m} \hat{y}_{1i} \\ \vdots \\ \sum_{i=1}^{m} \hat{x}_{1i} \\ \sum_{i=1}^{m} \hat{y}_{1i} \end{bmatrix} \tag{7}$$

This implies that the origin of the deck reference system is established at the centroid of the 3D points. On the other hand, since the deck motion is not controlled, we do not know[2] where is located the origin of the reference system. As a result of this decision, the image point's coordinates, in the new image coordinate system, are related with the 3D points coordinates by the Eq. (8), which has a similar form as Eq. (5):

$$\underbrace{\begin{bmatrix} \hat{x}_{11}-\bar{x}_1 & \cdots & \hat{x}_{1m}-\bar{x}_1 \\ \hat{y}_{11}-\bar{y}_1 & \cdots & \hat{y}_{1m}-\bar{y}_1 \\ \vdots & \ddots & \vdots \\ \hat{x}_{n1}-\bar{x}_n & \cdots & \hat{x}_{nm}-\bar{x}_n \\ \hat{y}_{n1}-\bar{y}_n & \cdots & \hat{y}_{nm}-\bar{y}_n \end{bmatrix}}_{W} = \underbrace{\begin{bmatrix} M_1 \\ \vdots \\ M_n \end{bmatrix}}_{M} \cdot \underbrace{\begin{bmatrix} P_1 & \cdots & P_m \end{bmatrix}}_{P}$$

$$W_{2nxm} = M_{2nx3} \cdot P_{3xm} \tag{8}$$

where M_i is the projection matrix of the camera i and the coordinates of $P_{i=1\ldots m}$ are referred to the centroid of the 3D points (origin of the deck reference system). The matrix W is called the *registered measurement matrix* and contains only measurement data, whereas the right side is the product of a *2n* x *3* matrix M (called the *motion matrix*) and a *3* x *m* matrix P (called the *object shape matrix*). This fact implies that the rank of $M \leq 3$ and, consequently, the same apply to W. Thus, the matrix W can be decomposed, using the single value decomposition (SVD), as $W = U \cdot D \cdot V^T$, where U and V are orthogonal matrices and D is a diagonal matrix containing the singular values of W, such as $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_m$. Since the rank of $W \leq 3$, only the first three single values (λ_1, λ_2 and λ_3) may be different from zero. However, if the coordinates of the image points are perturbed by noise then the rank of W can be higher and, since we are only interested in the three major singular values, the others can be forced to zero to guarantee the rank condition, as soon as the ratio between λ_3 and λ_4 is sufficiently large. Thus, assuming that $\widetilde{U}$ represents the first three columns of U, $\widetilde{D} = diag(\lambda_1, \lambda_2, \lambda_3, 0, \cdots 0)$ and $\widetilde{V}$ represents the first three columns of V, an estimation for the camera matrices (M) and for the targets position (P), is given as:

$$\underbrace{\begin{bmatrix} M_1 \\ \vdots \\ M_n \end{bmatrix}}_{M_{2nx3}} = \widetilde{U}_{2nx3} \cdot \sqrt{\widetilde{D}_{3x3}}$$

$$\underbrace{\begin{bmatrix} P_1 \\ \vdots \\ P_m \end{bmatrix}^T}_{P_{3xm}} = \sqrt{\widetilde{D}_{3x3}} \cdot \widetilde{V}^T_{mx3} \tag{9}$$

This estimation is only determined up to an affine transformation, since any arbitrary nonsingular matrix H satisfies the condition expressed by Eq. (9). In fact, $W = (M \cdot H) \cdot (H^{-1} \cdot P) = M \cdot P$ satisfies as well as, which involves to find the transformation H_{3x3}. The calculation of H is supported by the fact that the rotation matrix of each camera (R_{2x3}) is orthogonal, i.e. the norm of each line is unitary and the cross product between the lines is zero. Thus, each camera projection matrix provides three constrains, described as:

$$\begin{cases} m_{11} \cdot (H \cdot H^T) \cdot m_{11}^T = 1 \\ m_{12} \cdot (H \cdot H^T) \cdot m_{12}^T = 1 \\ m_{11} \cdot (H \cdot H^T) \cdot m_{12}^T = 0 \\ \vdots \\ m_{n1} \cdot (H \cdot H^T) \cdot m_{n1}^T = 1 \\ m_{n2} \cdot (H \cdot H^T) \cdot m_{n2}^T = 1 \\ m_{n1} \cdot (H \cdot H^T) \cdot m_{n2}^T = 0 \end{cases} \tag{10}$$

where the unknowns are the elements of the symmetric matrix $H \cdot H^T$ and m_{ij} represents the line j of the matrix M_i. The solution for $H \cdot H^T$ can be calculated by using SVD and the solution for H by Cholesky factorization. However, this last step requires that the solution of $H \cdot H^T$ be positive definite which, in general, is satisfied with three or more

[2] In the end of the calibration, if desired, it is possible to move de origin of the coordinate system to another point.

cameras. When this condition is not satisfied, the solution of H can still be found by non-linear optimization. Having found the matrix H, we calculate the new camera matrices and the new targets position (up to a scale factor) as:

$$\begin{cases} M' = M \cdot H \\ P' = H^{-1} \cdot P \end{cases} \quad (11)$$

To measure the deck bridge displacements, we need to find the scale factor (λ), such that the reconstructed object dimensions are approximately similar to the real dimensions. For this purpose, a single distance between two targets would be enough. Although, since the coordinates of the points projected in the image may be corrupted by noise, using as many distances as available allows the estimation of an average scale factor $(\bar{\lambda})$ and the reduction of this source of error. Assuming that the distance between any two 3D points in the deck is known, the scale factor is determined as the ratio between the real distance (measured) and the distance (calculated) between the respective reconstructed points (P'). Thus, likewise in the Eq. (11), one must multiply the camera's matrix by the inverse of the scale factor:

$$\begin{cases} M'' = M' \cdot \bar{\lambda}^{-1} \\ P'' = \bar{\lambda} \cdot P' \end{cases} \quad (12)$$

By now, we have an approximated Euclidean shape reconstruction that can be improved by a non-linear optimization process, just like the Levenberg-Marquardt (LM) algorithm. The error function (ε) to be minimized by the LM algorithm is:

$$\varepsilon = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} \underbrace{\left\| p_{ij} - \hat{p}_{ij}\left(\hat{M}_i, \hat{P}_j\right) \right\|^2}_{coordinates\ error} \right) + w \cdot \sum_{k=1}^{C_2^m} \left(\sum_{q=k+1}^{m} \underbrace{\left\| D_{kq} - \left(\hat{P}_k - \hat{P}_q\right) \right\|^2}_{distance\ error} \right) \quad (13)$$

where p_{ij} represents the coordinates of the point j in the image of the camera i, $\hat{p}_{ij}$ is the image coordinates of the reconstructed world point $\hat{P}_j$, projected according to the estimated matrix model $\hat{M}_i$, both under optimization, w is a weight of confidence, C_2^m is the number of combinations of any two 3D deck targets ($\hat{P}_k$ and $\hat{P}_q$) and D_{kq} is the respective real distance. It is advantageous to include the maximum number of constrains in the optimization process. Thus, using the distances obtained from all possible combinations of points one gets a rigid net of connections between points, that clearly enhance the final solution.

The error function comprises two contributions of errors: i) the error due to the projection of the reconstructed deck point in the image of each camera; and ii) the distance error between either two reconstructed points (P''). The observations related with the image coordinates can be affected by noise and, as consequence, are less feasible than the distances which are accurately measured. Thus, since the LM algorithm allows the use of different weights according to the confidence we have in the observations, we choose a weight ratio of 1/10, i.e. $w = 10$. The entries to optimize by the LM are the points ($\hat{P}_j$) and the cameras matrices ($\hat{M}_i$); the initial values are those obtained up to now, i.e. $\hat{M}^0 = M''$ and $\hat{P}^0 = P''$.

The transformation H, found before, was determined up to a rotation of the deck reference system, since the orientation of the reference system is arbitrary. In this context, it is necessary to rotate the deck reference system to a desired and known orientation. To fulfil this goal one needs to find a B_{3x3} rotation matrix that rotates the actual reference system such as the y-axes is aligned in the vertical direction and the x-axes in the transversal direction of the deck bridge. For this purpose, a minimum of three targets are required, disposed such as they establish an orthogonal reference system with the desired orientation. These targets can be positioned having as reference the orientation of the main truss of the deck bridge. The Eq. (14) represents the camera matrices and the reconstructed

targets position obtained at the end of the calibration task.

$$\begin{cases} M''' = \hat{M} \cdot B \\ P''' = B^{-1} \cdot \hat{P} \end{cases} \quad (14)$$

As said before, the origin of the deck reference system is placed at the centroid defined by all points captured under the deck bridge motion. Considering that the main goal is to measure the displacements of the deck bridge, the position of the reference system is secondary and does not settle a particular problem to the measurement system, as soon as it is under the central zone of the deck motion. However, if desired and since the cameras matrices are known, it is possible to move the reference system to another position (e.g. to the object centroid position related to the first captured frame).

4 - PERFORMANCE EVALUATION

4.1 - System configuration

In Fig 3 we present two camera's layout with two (L_1 – S–N) and four cameras (L_2 – SW–SW–NE–NW) which, among others, were used to evaluate the proposed calibration methodology.

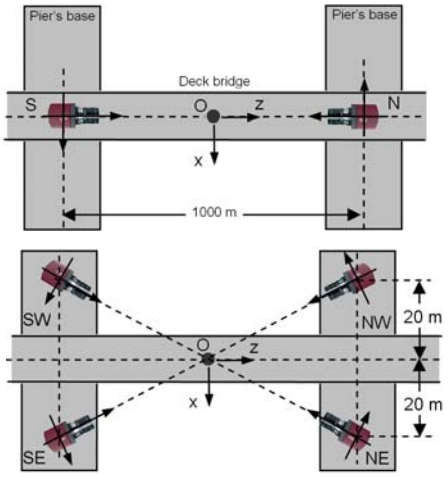

Fig. 3 – Measurement system layout with two cameras and with four cameras.

Taking into account the characteristics of the bridge used as model (bridge 25 de Abril over the river Tagus), the localization of the cameras were firstly constrained by the available physical support to place the cameras and secondly by the dimensions of the piers base. For the tests it was assumed that: i) the mid span of the deck bridge is 500 m length; ii) the height from the deck bridge to the piers base is 100 m; and iii) the piers base is 40 m wide[3]. The reference system for the deck bridge is a direct orthonormal basis with origin in *O* and the z-axis aligned with the longitudinal direction of the deck. In the same way, in each of the cameras is established a direct orthonormal coordinate system with the *z*-axis perpendicular to the image plan and pointing in direction of the origin of the deck's reference system.

4.2 - Vision system calibration

The calibration was carried out according to the proposed methodology, considering the synchronous capture of 1000 images while the deck bridge was moving. For that, it was created a 3D simulated deck trajectory, whereas each displacement's direction is a three part's combination: i) large amplitude with low frequency; ii) low amplitude with high frequency; and iii) a random component with low amplitude (Fig. 4). The goal was to get a simulated trajectory which gathers the long period motion and the instantaneous vibratory movements. The vision system calibration was performed assuming a set of sixteen targets fixed to the deck bridge and visible by all the cameras involved on the test.

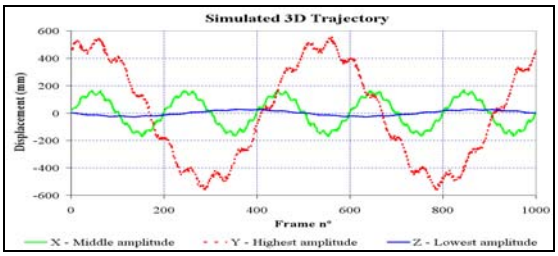

Fig. 4 – Simulated 3D deck trajectory created for calibration purposes.

The targets were distributed taking into account the bridge geometry and avoiding a possible target overlap in the images, as shown in Fig. 5. They form the shape (contour) of a parallelepiped with the sides parallel to the axes of the

[3] These parameters values are approximated of the real values of the bridge 25 de Abril over the river Tagus.

reference system (*O*) and the centre of mass matching the origin. For each image captured, independent Gaussian noise (in pixels) was added to the image coordinates with zero mean and different levels of standard deviation (SD)[4] according to the test performed.

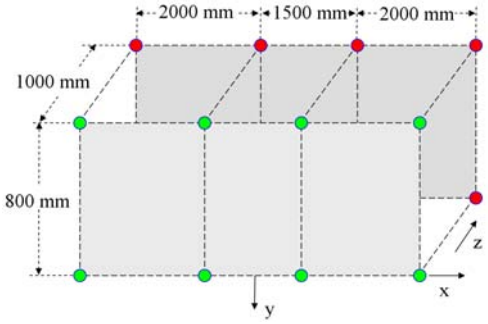

Fig. 5 – Targets distribution and distance between the adjacent targets, for calibration purposes.

4.3 - Deck bridge monitoring

To measure the 3D displacements of the deck bridge, a minimum of two cameras is required. Regarding the number of targets, a single tracking target is enough to measure the deck displacements. However, using as much targets as available allows to reduce the noise effect. Assuming that we have *m* targets and *n* cameras, recalling Eq. (8), the 3D world coordinates of the *m* targets are given as:

$$\underbrace{\begin{bmatrix} X_1 & \cdots & X_m \\ Y_1 & \cdots & Y_m \\ Z_1 & \cdots & Z_m \end{bmatrix}}_{X_{3\times m}} = \underbrace{\begin{bmatrix} M_1 \\ \vdots \\ M_n \end{bmatrix}^{+}}_{M_{2n\times 3}} \cdot \underbrace{\begin{bmatrix} x_{11}-\bar{x}_1 & \cdots & x_{1m}-\bar{x}_1 \\ y_{11}-\bar{y}_1 & \cdots & y_{1m}-\bar{y}_1 \\ \vdots & \ddots & \vdots \\ x_{n1}-\bar{x}_n & \cdots & x_{nm}-\bar{x}_n \\ y_{n1}-\bar{y}_n & \cdots & y_{nm}-\bar{y}_n \end{bmatrix}}_{W_{2n\times m}} \quad (15)$$

where $M_{i=1,\ldots,n}$ is the camera matrix[5] of the camera *i*, $\left[x_{ij} \quad y_{ij}\right]^T_{i=1,\ldots,n;\ j=1,\ldots,m}$ are the coordinates of the target *j* in the image of the camera *i* and $\left[\bar{x}_i \quad \bar{y}_i\right]^T_{i=1,\ldots,n}$ are the origin of the system coordinates at the camera *i*, obtained by Eq. (7), at the calibration stage.

Considering that the calibration of the vision system was carried out as previously described, the monitoring test was performed tracking 20000 positions of the deck bridge, obtained randomly and assuming a Gaussian distribution with zero mean and a standard deviation of 1000 mm in the transversal direction, 2000 mm in the vertical direction and 500 mm in the longitudinal direction. The tested conditions are more demanding than in a real scenario because in this case, the displacements between two successive instants are much smooth and of smaller amplitude. The noise, added to the image coordinates (in pixels), was assumed to be Gaussian with zero mean and standard deviation variable according to the test.

Fig. 6 shows the absolute mean deviation obtained with different number of targets and under different noise levels at the calibration and at the monitoring stages. As expected, the absolute mean deviation in *XY* plan and in *XYZ* decreases as the number of targets and/or of cameras increase, being more significant in the case of *XYZ*. The absolute mean deviation in *XYZ* is much higher than in *XY* because the measurement of the longitudinal displacement is more difficult to measure and, as result, more prone to error. In fact, the displacement error in the longitudinal direction represents the most important portion on the *XYZ* deviation.

As showed, even under severe noise conditions and large amplitude deck displacements, the required accuracy in *XY* plan is satisfied in a layout with two cameras and two monitoring targets or with four cameras and one monitoring target. Using sixteen targets with four camera's layout, one gets an absolute mean deviation in *XYZ* lower than 10 mm. However, tests performed with a typical stereo camera configuration (two cameras placed on a single piers base, SE–SW or NE–NW) allowed to verify that this layout is not suitable to measure the displacements at long distance, as is the case. This evidence shows that the absolute mean deviation value depends not only on

[4] Henceforth, the reference to a number of pixels of noise means a Gaussian distribution with zero mean and a standard deviation with that number of pixels.

[5] $\left[M\right]^+$ means pseudo-inverse of *M*.

the number of targets and cameras but also on the camera's layout.

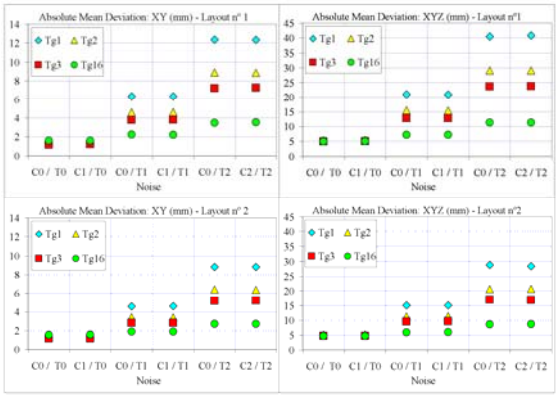

Fig. 6 – Deviation on *XY* plan and *XYZ* obtained with the two layouts, as function of the noise level, at the calibration (*C*) and at the monitoring (*T*) stages, and of number of targets (T_g)[6].

Fig. 7 presents a 3D simulated deck translation trajectory, representative of 10000 positions of the deck bridge, created following an identical principle as referred to the vision system calibration trajectory (Fig. 4).

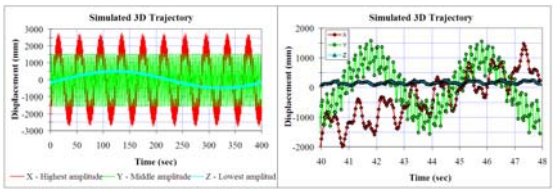

Fig. 7 – Simulated 3D deck bridge trajectory created for monitoring purposes.

Fig. 8 shows the absolute deviation in *XY* and in *XYZ*, obtained with sixteen monitoring targets under 1 pixel of noise, in calibration and in monitoring stages, for the two layouts. The results obtained confirm that the absolute peak value of deviation in *XY* is less than 10 mm and the absolute mean value is less than 2 mm. Likewise, the absolute mean deviation in *XYZ* is also lower than 10 mm, although some points show high deviation. As one can see, the absolute mean value deviation obtained in this test is slightly smaller than those presented on figure 5 (for the same noise conditions) because, in this case, the displacements are of less amplitude.

[6] C_m/T_n means standard deviation noise of *m* pixels in calibration and *n* pixels in tracking (monitoring). $T_g x$ means *x* targets used on the monitoring stage.

The main conclusions are: i) the noise effect in the calibration stage has a negligible contribution in the deviation value of the displacement, confirming the robustness of the proposed calibration methodology; ii) using an adequate number of targets and/or of cameras is possible to get an absolute mean deviation lower than 10 mm in the *XY* plan; iii) for the same test conditions the deviation in *XYZ* is almost three times higher than in *XY*. Regardless of this evidence, the displacements in the transversal (*X*) and vertical (*Y*) directions, usually, are much higher than in the longitudinal direction (*Z*) and, as a consequence, they play the most important role in the bridge monitoring.

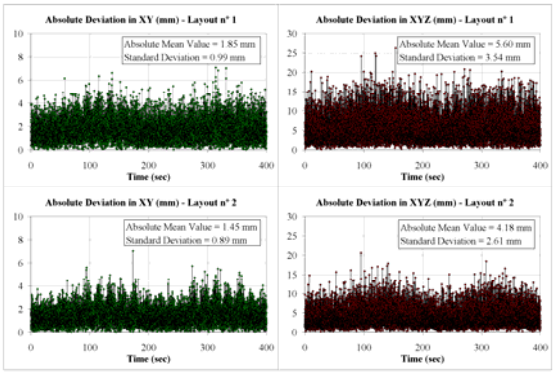

Fig. 8 – Absolute deviation in *XY* and in *XYZ* obtained with 1 pixel of noise in the calibration and in the monitoring stages.

5 - CONCLUSIONS

The paper describes a measurement vision-based system suitable for structure health monitoring application of large structures, in particular long deck suspension bridges, where the vertical and the transversal displacements of the deck can reach a couple of meters and the access to some parts of the structure is difficult. It is proposed a methodology to calibrate the vision system *in-situ* while the deck bridge is in motion. To fulfil this task, a minimum of two cameras, viewing a set of targets fixed to the deck and the 3D distance between any two targets are the only source of data that is required.

The results obtained from simulated tests, carried out with different cameras

layout and with different noise levels, showed the good feasibility and the robustness of the proposed methodology. This conclusion is supported by the high accuracy value in the measurement of the deck bridge displacements, even under severe noise conditions. Actually, using two cameras tracking a minimum of two targets or four cameras tracking one target, under abnormal noise conditions (standard deviation of 2 pixels), the measurement system can satisfy an accuracy in *XY* (vertical and transversal directions) better than 10 mm. Further, the accuracy value can be improved increasing the number of cameras and/or of targets.

Taking into account the achieved results, we may conclude that the vision system is more adequate for the deck bridge monitoring than *GPS* and radar based systems.

6 - REFERÊNCIAS

Bernardini, G. Pasquale, G. Bicci, A. Marra, A. Coppi, F. Ricci, P. Pieraccini, M. 2007. Microwave interferometer for ambient vibration measurement on civil engineering structures: 1. Principles of the radar technique and laboratory tests, EVACES'07, Experimental Vibration Analysis for Civil Engineering Structures.

Brown, C. Robertsm G. Meng, X. 2006. When bridges move: GPS-based deflection monitoring, Sensors, 4, p. 16–19.

Çelebi, M. Eeri, M. Sanli, A. 2002. GPS in pioneering dynamic monitoring of long-period structures. Earthquake Spectra; 18 (1), p. 47-61.

Fukuda, Y. Feng, M. Shinosuka, M. 2009. Cost-effective vision-based system for monitoring dynamic response of civil engineering structures, Structural Control and Health Monitoring, DOI: 10.1002/stc.360.

Hartley, R. Zisserman, A. 2003. Multiple view geometry in computer vision, 2nd Edition, Cambridge University Press.

Lee, J. Shinozuka, M. 2006. A vision-based system for remote sensing of bridge displacement, NDT&E International, 39, p. 425–431.

Macdonald, J. Taylor, C. Thomas, B. Dagless, E. 1998. Real-time remote monitoring of dynamic displacements by computer vision, 6th Society for Earthquake and Civil Engineering Dynamics Conference, p. 389–396.

Marécos, J. Castanheira, M. Trigo, J. 1969. Field observation of Tagus river suspension bridge, Journal of the Structural Division, Proceedings of the American Society of Civil Engineering.

Nakamura, S. 2000. GPS measurement of wind-induced suspension bridge girder displacements, Journal of Structural Engineering, 126 (12), p. 1413–1419.

Olaszek, P. 1999. Investigation of the dynamic characteristics of bridge structures using a computer vision method, Measurement, 25 (3), p. 227–236.

Santos, C. Oliveira Costa, C. Batista, J. 2007. Disponibilidade ao longo do tempo de um sistema óptico de medição de deslocamentos da P25A, ICM'2007, A Instrumentação Científica e a Metrologia Aplicadas à Engenharia Civil, LNEC.

Santos, C. Oliveira Costa, C. Batista, J. 2009. Sistema de medição de deslocamentos verticais e transversais de tabuleiros de pontes suspensas ou atirantadas por meio de dispositivos ópticos, ASCP'09, 1º Congresso de Segurança e Conservação de Pontes, II, p. 125–132.

Solheim, F. Vivekanandan, J. 1999. Propagation delays induced in GPS signals by dry air, water vapour, hydrometeors, and other particulates, Journal of Geophysical Research, 104 (D8), p. 9663–9670.

Stephen, G. Brownjohn, J. Taylor, C. 1993. Measurement of static and dynamic displacement from visual monitoring of the Humber bridge, Engineering Structures, 15 (3), p. 197–208.

Tomasi, C. Kanade, T. 1992. Shape and motion from image streams under orthography: a factorization method, International Journal of Computer Vision, 9 (2), p. 137–154.

Wahbeh, A. Caffrey, P. Masri, S. 2003. A Vision-based approach for the direct measurement of displacements in vibrating systems, Institute of Physics Publishing, Smart Materials and Structures, 12, p. 785–794.