Methods and algorithms for video-based multi-point frequency measuring and mapping

David Mas, Belen Ferrer, Pablo Acevedo, Julian Espinosa

PII: S0263-2241(16)00119-6
DOI: http://dx.doi.org/10.1016/j.measurement.2016.02.042
Reference: MEASUR 3828

To appear in: *Measurement*

Received Date: 6 May 2015
Revised Date: 7 January 2016
Accepted Date: 21 February 2016

Please cite this article as: D. Mas, B. Ferrer, P. Acevedo, J. Espinosa, Methods and algorithms for video-based multi-point frequency measuring and mapping, *Measurement* (2016), doi: http://dx.doi.org/10.1016/j.measurement.2016.02.042

Methods and algorithms for video-based multi-point frequency measuring and mapping.

David Mas[1*], Belen Ferrer[2], Pablo Acevedo[1] ,Julian Espinosa[1]

[1]Univ. Inst. Physics Applied to Sciences and Engineering,

[2]Dept. Civil Engineering

University of Alicante, PO Box, 99, 03080 Alicante (Spain)

*david.mas@ua.es

ABSTRACT

Object vibrations and movements can be detected through changes in its luminance. In this paper, we demonstrate that we can obtain the vibration frequency of all vibrating targets in the sequence simultaneously through the analysis of local neighborhoods. The study is completed with a short-time Fourier analysis so that changes in the movement frequencies are also accounted. We also show that this information can be displayed like a color frequency map that can be superimposed to the video sequence providing a whole description of the analyzed sequence. The method can be used to analyze complex structures since their different vibrating parts can be visualized at a glance. The main algorithms, methods and some sequences are freely downloadable so that new applications and procedures can be implemented by the scientific community.

GRAPHICAL ABSTRACT

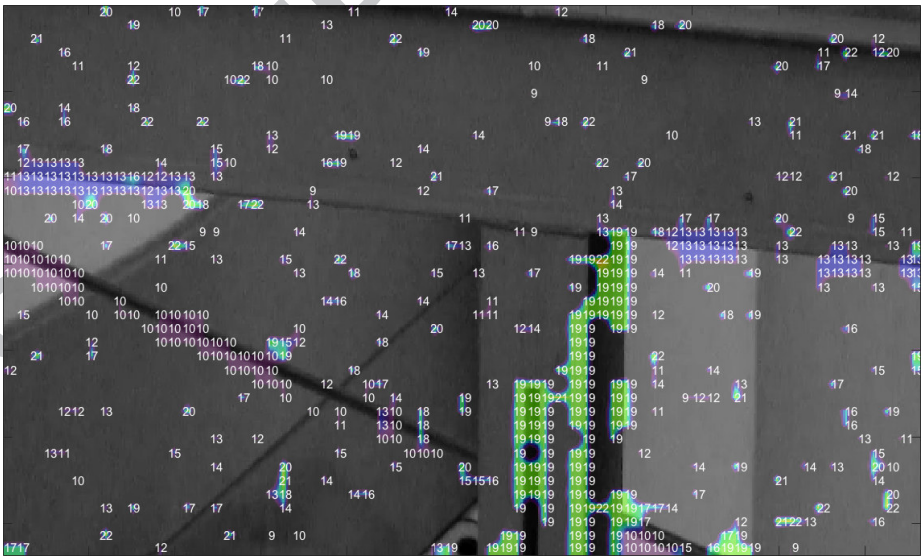

Figure Color map showing the vibration frequencies of the different parts of a pedestrian bridge. Notice the cable, deck and vertical beam vibrating at 10, 13 and 19 Hz respectively.

Keywords: vibration analysis, video processing, multipoint targetless vibrometry.

1.- Introduction

Movement detection and frequency measurements are of great importance in many fields including Physics and Engineering. Contact sensors are often preferred and, among them, accelerometers are the most used for vibration monitoring. Although their really good performance, in some occasions the complexity or accessibility of the specimen to be measured increases the installation cost and may be even hazardous for the test crew safety. Because of this, non-contact methods such as laser vibrometers are becoming more popular and are being applied to new fields [1]. Nevertheless, their high cost is still a big drawback that restricts their use to only some particular cases.

Computer vision systems have been demonstrated to be a reliable alternative to traditional methods. In the last years, thanks to the decreasing prices of high-speed cameras together with the increased capabilities of storage systems and computers, several proposals, methods and applications have appeared in several areas such as Biomedical Sciences, Surveillance or Civil Engineering [2-14]. With independence of the application area or the proposed method, the common hypothesis of these papers is that object movement can be perceived through changes in the light reflected or diffused by a moving target.

In references [8-10], the authors use object recognition techniques in order to determine the vibration of different structures. Although these methods provide accurate results, they require of high contrast images in order to facilitate the image analysis and also tend to be very demanding in computational time and resources. Vibration measurement techniques without object recognition have been also proposed in the literature. In [11], the authors analyze video sequences of objects placed close to a loudspeaker. Video analysis is done through decomposition of the video signal into the local spatial amplitude and phase using a complex-value steerable pyramid filter bank [12]. Through image analysis they are able to speech capturing and recognition even through acoustically isolating windows. The method does not require that the vibrating surface be specular, diffusive or fulfill any special property to obtain a reliable signal. The obtained impacting results received much attention from public and

media. The same approach has been used later by some of the authors to magnify the motion of simple structures and facilitate the modal identification and analysis [13]. Unfortunately, the algorithm presented requires complex image transformations and may result in a slow and heavy algorithm. This problem was later addressed in [14]. There, the authors present a reformulation of the idea in terms of image correlation. They claim that their method is much faster than the original one and is able to resolve movements up to 0.16 px on the CCD sensor. Although this may seem a minor detail, resolutions below the pixel limits are of fundamental importance for measuring small vibrations or from far distances. It results evident that large movements can be easily detected by noticeable changes in the image. On the contrary, small amplitude movements are not so easy to detect since changes in the image will be subtle and affect to very few unconnected pixels. Therefore, reliable methods for vibration measurements need to achieve sub-pixel resolutions.

In [15], the authors demonstrated the possibility of measuring the vibration frequency of objects in a high-speed video sequence just by counting luminance changes in local regions. The object does not have to fulfill any specific requirement and no especial feature is detected or matched. Although the method may not be as fast as the one proposed in [14], it is able to measure oscillation amplitudes of 0.013 px, as we will show below. This means an improvement of one order of magnitude with respect to similar papers cited above without performing any interpolation or image transformation. Additionally, the method is developed without any specific application in mind so it can be easily adapted to different experiments. Unfortunately, this technique is only capable of measuring the movement frequency since no direct connection has been found between pixel luminosity changes and vibration amplitude. Even though, frequency determination with simple and cheap non-contact method can be applied in a lot of areas and problems such as structural engineering, industrial mechanics and modal analysis.

In [15], an interesting idea was outlined: the possibility to extend the method for simultaneous multipoint calculation. This feature is of great interest since it

would permit multiple measurements from a single experiment. Therefore, with a simple excitation, one could measure and compare the vibration frequency between different parts of an object, or even different objects, and obtain valuable information about the system dynamics.

Multipoint measurements have been implemented in Laser Vibrometers through a scanning mechanism but it may result very expensive, slow and measurements are taken asynchronically so, strictly speaking it is not a multiple point device. Other authors have proposed the use of radar interferometry in order to measure vibrations in structures and buildings [16-18]. The method is fast and accurate but it is still expensive and requires of special properties of the vibrating object in order to produce a clean signal. In [19], the authors propose the use of a Kinect sensor to obtain the time variation of vibrating three-dimensional surfaces. The system permits simultaneous measurement of multiple points in real time with a cheap device, but it is limited by the hardware to frequencies below 15 Hz.

In this paper, we propose a simple and low cost method for simultaneously measuring the vibration frequencies in all the points of a captured video sequence. The idea is developed from the algorithms presented in [15], which consisted of selecting a local neighborhood or region of interest (ROI) and applying a multilevel binary threshold to the image within this region. By proper combination and analysis of all binary levels the authors are able to obtain the vibration frequency of the part of the object within the ROI. Here, we exploit the idea and propose to extend the method and to analyze the frequency information in the entire scene and through the whole sequence. Therefore, the frame is subdivided in overlapping ROIs that are individually analyzed. Each ROI contains all the voxels through the sequence and thus allows the calculation of the vibration frequencies of each local neighborhood. With this information, a frequency map of the sequence can be built and represented. As we will see in the examples provided, this allows identifying different vibrating parts of a complex object and even recognize the source of the detected frequencies.

Finally, we will implement a Short-Time Fourier transform algorithm so that we can obtain the frequency changes and variations occurring in a single sequence.

Compared with [19], our proposal is restricted to 2-D movements but it can be easily implemented with any camera. In fact, it has been used both with a high-end and a pocket camera with satisfactory results in both cases. Because of this versatility, the method is mainly limited by the camera configuration. In the following pages, we will show the capabilities of our proposal. The reader will notice that although the same example can be used for all the experiments, we decided to present different vibrating objects so that the performance and capabilities of the method is better understood.

The structure of the paper is organized as follows. In Section 2, we will recall the basic principles and algorithms that allow the calculation of the vibration frequency in a ROI from a high-speed video sequence. In Section 3, we will show the capabilities of the method for detecting and resolving frequency changes in the object within the ROI. In Section 4, we will extend the method to the whole frame to obtain a frequency map identifying different parts of an object vibrating simultaneously at different frequencies. Combination of the mapping option with the short-time Fourier transform will allow us to build a time-varying frequency video map, that is capable of detecting different objects vibrating both simultaneously or at different times. Finally, in the last Section, we will outline the main conclusions.

2.- METHOD

In [20], the authors presented a discussion about the theoretical limits of object tracking with sub-pixel resolution. The paper proposed a theoretical approach to an optimal target design that would maximize the detection probabilities but, unfortunately, the proposal had not realistic possibilities of being implemented in real experiments. However, in [15], the authors showed that part of the condition could be fulfilled for almost any object whose contour is not aligned with the CCD array. With this, they proposed a method that allows obtaining the vibration frequency of a moving object through illumination changes of the

pixels within a ROI in a temporal sequence. Since the results here presented are derived from those in [15], we include a review of the method with many improvements that have been added since it was published.

The method was implemented in the following way: we determined the minimum and maximum luminance levels within each ROI and calculated several equidistant levels between these two. Then, the image within the ROI was thresholded and binarized at each level so that we obtained different binary sequences. The number of white pixels at each frame was then counted and a numerical sequence was obtained for each ROI. Our hypothesis was that any change in the object position can be detected as a local variation in the image luminance, thus it will be seen as a change in the number of active pixels in some (or many) binary levels.

Let us consider a temporal sequence of images $I(i,j,t)$ being i and j the coordinates of the image pixels, and consider that in each frame, we select a rectangular ROI of size (N_x, N_y) located by its central pixel (i_0, j_0) inside the frame. Therefore a signal $B_l(i_0, j_0, t)$ codifying the movement within a ROI at each level l is constructed in the following way:

$$B_l(i_0, j_0, t) = \sum_{m=i_0-N_x/2}^{i_0+N_x/2-1} \sum_{n=j_0-N_y/2}^{j_0+N_y/2-1} \left[I(m,n,t) \right]_l \tag{1}$$

with:

$$\left[I(m,n,t) \right]_l = \begin{cases} 0 \text{ if } I(m,n,t) \geq l \\ 1 \text{ if } I(m,n,t) < l \end{cases} \tag{2}$$

Notice that the signal $B_l(i_0, j_0, t)$ can be further processed and analyzed in order to obtain the main frequency peaks or other information about the movement. In Figure 1, we show a video sequence of a vibrating tuning fork designed for vibrating at a frequency of 440 Hz. We also show there its correspondent binary version thresholded at level 128. By running the video, one can see some changes in the gray-scale sequence, but not recognizable movement. Conversely, in the binary sequence, periodic changes can be clearly appreciated thus

revealing the movement pattern of the tuning fork. As we showed in [15], simple counting of active (white) pixels through the sequence, as it is expressed in eq. (1) provides the vibration frequency of the object within the displayed ROI.

Figure 1 (video). a) Extreme of a tuning-fork prong vibrating at 440 Hz. b) Sequence binarized at level 128. (See the video at https://goo.gl/2vdmg4)

With independence of the number of levels, two different options can be followed when implementing the method. The first one consists of taking the absolute maximum and minimum of the ROI through the whole the sequence and applying the same thresholds for all frames. This approach implies loading all the frames in the computer memory, which results in a very inefficient algorithm. The alternative approach consists of calculating the threshold levels according to the ROI's maximum and minimum intensity in each frame. In this case, the sub-index l in expression (1) refers to the threshold ordinal (i.e. first binary level, second, etc.) and not the binary level itself. This approach has the advantages that is faster and automatically adapts to the general illumination over the scene. In any case, we tried both implementations and results did not differ very much.

The signal obtained from equation (1) through any of the procedures just described tends to be very noisy. Any change in the scene, independently of its cause, will be detected through the luminance variation, thus degrading the signals. Fortunately, if the movement is repeated, a peak will appear above the noise level, and the frequency of the movement will be determined, provided that the signal to noise ratio is high enough.

In order to increase the signal quality, several procedures were implemented. The first one consisted of selecting the ROI size in order to be sure that majority of changing pixels belong to the vibrating object and not to the noisy background. Proper election of a ROI depends on many issues, and a general rule is difficult to establish. In general, the ROI must be selected in a high contrast region and be large enough to contain moving pixels and also static parts to provide contrast in each binary level. As a rule of thumb, we recommend to consider ROIs that are double size that the image detail whose movement is being tracked i.e. In the case we are observing a bar of 10 pixels width whose expected transversal movement is 2 pixels, a scanning area of 20x20 pixels should fit. In this case, almost have of the pixels are occupied by the moving objects and the other half by the static background

Second procedure consisted of convenient combination of the information obtained for each binary level. The movement of the object will probably affect to different brightness levels. Therefore, the numerical signals obtained from the binary sequences in equation (1) will show some redundancy that can be exploited to improve the signal and decrease the noise. In [15], the authors took eight binary levels between the minimum and maximum luminance levels and then calculated the average of the eight frequency spectra obtained for each ROI. Although other options can be explored, we also take 8 levels in all the examples that are described below. Unfortunately, for subtle vibrations, some binary sequences may only contain noise and thus degrade the result when all the signals are combined. Therefore, instead of just calculating the mean value of the signals, we decided to calculate their weighted sum so that the signals carrying relevant information are enhanced whereas the others are attenuated. Calculation of the weight coefficients must be simple enough to not increase the computational cost, which may be very high. According to this, we decided to implement the variance of the signal as the weight function. Since all signal energies have been normalized to the unit, higher variances imply lower signal-to-noise ratio and thus a noisier signal [21]. Therefore, if we consider the binary sequences $B_i(t)$ in one ROI, the final signal describing the activity in the area is:

$$S(t)=\sum_{l=L_{Min}}^{L_{Max}}\frac{B_l(t)}{\sigma^2\left[B_l(t)\right]} \quad (3)$$

with σ^2 being the variance of the signal $B_l(t)$, and L_{Min} and L_{Max} the minimum and maximum luminance levels respectively. This final signal will be processed in order to obtain the frequency of the movement.

All the processes just explained have been implemented using our own code in Matlab and is freely downloadable from [22]. This software permits selection and calculation of vibration frequencies in four regions simultaneously. Although copyright laws restrict commercial use [23], we encourage the readers to use our software and ask the community to check and improve our proposals.

3.- METHOD VALIDATION: MEASUREMENT OF A SINGLE ROI

In order to show the performance of the method, we have measured the movement of a complex pattern situated on a vibrating platform. In our case, we use a "Vibe-Tribe Troll" vibrating loudspeaker connected to a computer. The speaker is installed upside-down on an isolating Sorbothane foam. On the moving platform, we put a frame with an attached target showing a QR-code. Finally, a ceramic shear ICP accelerometer with a sensitivity of 1000 mV/g and a frequency range from 0.5 to 3 kHz was glued at the top of the frame in order to check the vibration frequency. Additionally, since our method does not provide the amplitude of the movement and due the stationary character of this movement, we will use signal from accelerometer to obtain, through double integration, an estimation of the movement, and infer the sub-pixel accuracy.

The video sequence was captured in non-compressed AVI format with an AOS X-PRI camera, working at 1000 fps and situated at 50 cm from the target. A 50 W standard halogen lamp connected to a stabilized AC/DC converter illuminated the scene. In Figure 2, we show the scene with the QR target on it. The target side is of 40 mm length with a frame size of 800x600 thus giving a resolution of 3.60

px/mm. The vibrating platform was set to a frequency of 330 Hz. We also show in the figure the ROI used for the calculation with our method.

Since the signal registered by the accelerometer was clean and soft, we decided to adjust it to a sum of sines and, from that, obtain an estimation of the movement amplitude. The simplest expression that provided a correlation coefficient above 0.95 was a sum of two sinuses so we adjusted our data to:

$$A(t) = a_1 \times \sin\left(b_1 \times t + c_1\right) + a_2 \times \sin\left(b_2 \times t + c_2\right) \tag{4}$$

with A being the acceleration in g units and t the time. The result of the fit was:

$$\begin{aligned} a_1 &= 1.598 \ \ (1.592,\ 1.603) \text{ g units} \\ b_1 &= 2073 \ \ (2073,\ 2074) \text{ rad/s} \\ c_1 &= 68.74 \ \ (67.51,\ 69.98) \text{ rad} \\ \\ a_2 &= 0.2079 \ \ (0.2023,\ 0.2136) \text{ g units} \\ b_2 &= 4146 \ \ (4144,\ 4149) \text{ rad/s} \\ c_2 &= 138.2 \ \ (128.8,\ 147.5) \text{ s} \\ \\ R^2 &= 0.99690 \\ RMSE &= 0.06412 \end{aligned} \tag{5}$$

with the amount between brackets being the confidence intervals at 95%, Notice that $b_2 = 2 \times b_1$ so result in (5) is just showing the fundamental (2073 rad/s or equivalently, 329,93 Hz) and first harmonic (659,70 Hz) of the oscillating movement. Equation (5) can be easily double-integrated so, substituting the obtained coefficients, we obtain that the movement amplitude is of $3.6 \mu m$.

In Figure 3 we show the signal obtained according to equation (3) from a ROI of 50x50 px in the center of the QR code. Notice that this signal is composed by adding the number of active pixels within the ROI in different thresholded levels. Therefore, the signal carries the information about the movement but localized peaks may not have a direct correspondence with the movement itself. In Figure 4, we show the frequency spectrum obtained through our method and from the

acceleration signal. We can appreciate that, as we predicted before, the signal coming from the video is noisier than that coming from the accelerometer. The reader should also notice the presence of low frequency components that is masked in the accelerometer signal by the higher frequencies. (Fig. 4c) In the Figure 4b, we show a magnification of the low frequency region of the acceleration pectrum with is eight orders of magnitude lower than the one represented in the general view. One can see there the presence of the peak at 7.8 Hz that has been detected by the camera, although its amplitude in signal from accelerometer is much lower and noisier. Use of seismic accelerometers may improve the performance of the system in the low frequencies range, but even so, it is known that accelerometers response depends on the frequency squared so the dominant high frequency peaks mask the low frequencies components. This effect was already observed previously [24] and may be further explored or neglected depending on the a priori information about the particular system that is being measured.

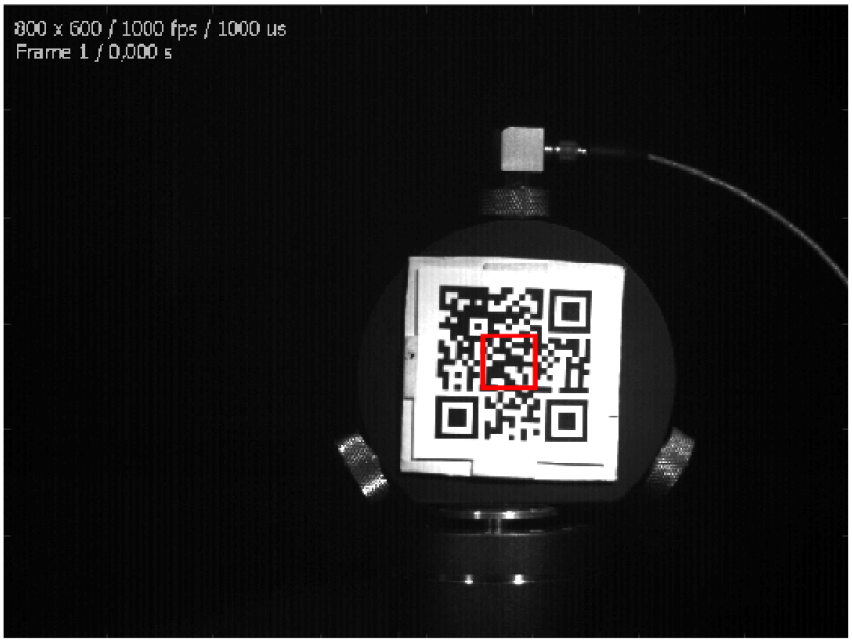

Figure 2. QR-code on the top of the vibrating speaker. The red square marks the analyzed ROI of 50x50 px.

Finally, we would like to underline that, according to the resolution of the system and the movement amplitude, our method is capable to detect vibrations of 0.012 px of amplitude without applying any interpolation, transform or recognition method over the images. This resolution is one order of magnitude higher than the one obtained by the one claimed in [14].

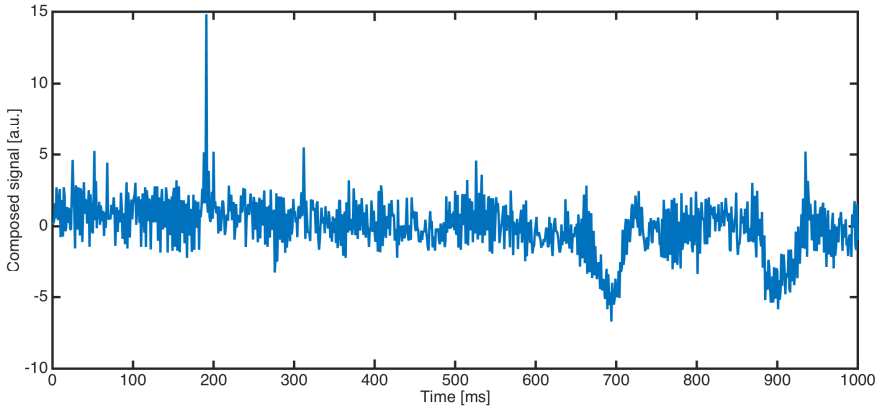

Figure 3. Signal obtained applying eq. (3) to the ROI selected in Figure 2

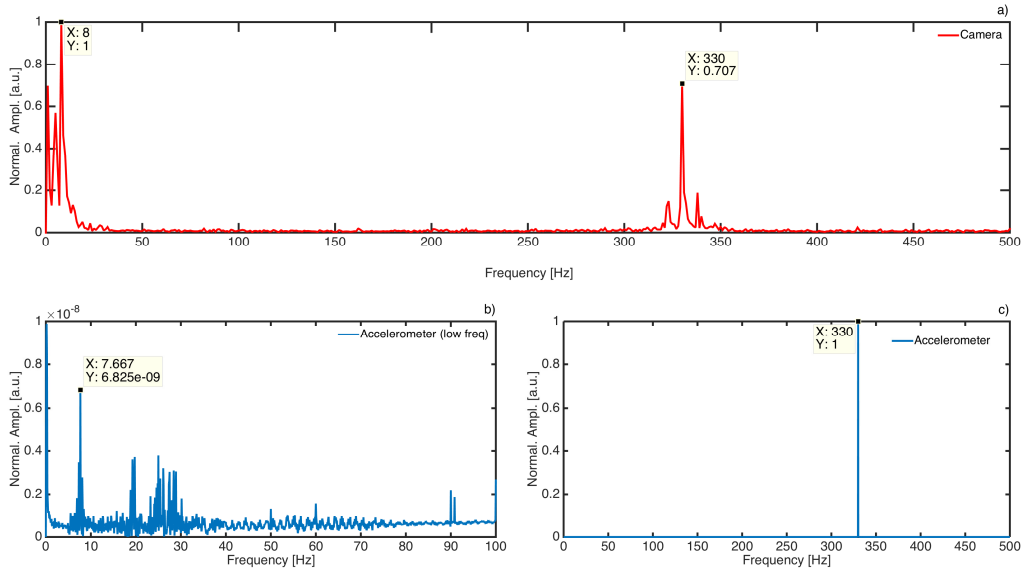

Figure 4. a) Fourier spectrum obtained with our method b) Low components of the Fourier spectrum obtained with the accelerometer. Notice the vertical scale. c) Fourier spectrum obtained from the accelerometer in the range [0-500 Hz]

The proposed method can be slightly modified so that additional results can be obtained from the binarized sequences. Up to know, we have implicitly assumed

that the frequency of the movement is constant, but this is not a requirement for our purpose to work. In our algorithm, the Fourier transform can be substituted by a short time Fourier transform (STFT) [25] and thus, changing frequencies inside a ROI can be measured in the same sequence.

In short, the STFT is a particularization of the general Fourier transform where the function to be transformed is multiplied by a temporal window function, which is non-zero during a short time. Therefore, the Fourier transform is calculated as the window is displaced along the time axis. The result is a two-dimensional time-frequency (τ,ω) representation:

$$STFT\{S(t)\}(\tau,\omega)\equiv X(\tau,\omega)=\int_{-\infty}^{\infty}S(t)W(t-\tau)\exp(-j\omega t)dt\ ; \qquad (6)$$

where $W(t)$ is the time window function, and $S(t)$ is the function to be transformed, which is obtained following equation (3). In the discrete time case, the above written expression converts into:

$$STFT\{S[t_k]\}(t_i,\omega)\equiv X(t_i,\omega)=\sum_{k=-\infty}^{\infty}S[t_k]W[t_k-t_i]\exp(-j\omega t_k)\ ; \qquad (7)$$

being $S[t_i]$ the signal obtained for the ROI evaluated at the frame corresponding to the instant t_i.

One of the issues of the STFT is related to the time-frequency resolution limitations. The shortest the time-window is, the higher time resolution, but, since we have fewer bins, the frequency resolution will be worse. And vice versa: the longer the time window is, the poorer temporal resolution is obtained. Therefore, windows width selection depends on the particular experiment and the expected results. Additionally, the window is usually selected with smooth borders in order to avoid the Gibbs phenomenon. In our case, we took a Hann window [25].

In Figure 5, we show a time-frequency chart obtained for the QR code and ROI already shown above in Figure 2. The target has been excited during two seconds with two frequencies consecutively. During the first second the frequency of

vibration was 330 Hz and then was suddenly changed to 445 Hz. The calculation of the STFT has been done in Matlab using the *tfrstft* function from the Time Frequency Toolbox [26] with the predefined parameters. The window function that has been used was the one suggested by the authors:

$$W(t_k) = 0.54 - 0.46\cos\left(\frac{2\pi t_k}{T+1}\right); \qquad (8)$$

with T being the total time length of the signal.

The picture clearly shows in yellowish colors the detection of the two frequencies and the time span while they are acting on the target. One can also see the presence of a low frequency band, just as it happened in the previous case.

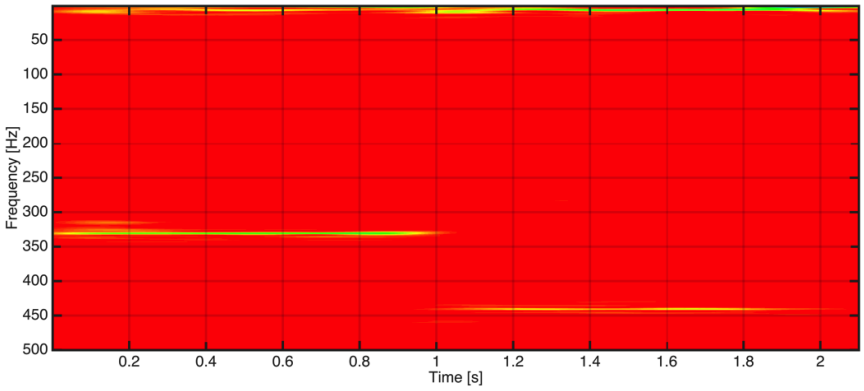

Figure 5. Time frequency chart obtained for the ROI in Figure 2 when two consecutive excitation frequencies (330 and 445 Hz) are applied. Yellowish colors show the frequency and the duration of the frequency peaks. Color scale is arbitrary so it has not been added to the map.

3.- CALCULATION AND REPRESENTATION OF FREQUENCY MAPS

Up to now, the procedure has been applied on a single region of interest. However, it can be extended to different regions in the scene so that we can obtain simultaneous measurement of different areas vibrating at different frequencies or even different objects. Moreover, the whole image can be divided in adjacent blocks and thus perform a local analysis of the movement. Overlapping of the regions may be adjusted depending on the computer capabilities: less overlapping implies fewer regions on the image so the calculation will be faster but less accurate. Each of these local neighborhoods is

analyzed according to the method exposed above, i.e. multilevel thresholding, pixel counting, signal composition and Fourier analysis of the resulting signal.

With this, each region is reduced to a vector containing the frequency spectrum obtained for this area. Therefore, we have a matrix containing the frequency spectrum for all the ROIs positions. With this information we construct a map for each frequency component. Notice that the smaller ROIs we select, the more accurate information will be obtained from the scene, but also the larger will be the resulting matrix.

One must bear in mind that the final aim of full-frame frequency analysis is to provide information about vibrating objects in a scene at a glance. Therefore, we decided to select, for each region the highest peak and represent the dominant frequency at each ROI defined in the scene.

In the following figures, we present an example of the method applied for a complex image analysis. In Figure 6, we display a pedestrian bridge inside a University building. A person jumping excited the bridge and caused its free oscillation for some seconds. For this example, we took a sequence of 2.5 s captured at 120 fps. After the visualization of the sequence, only a sequence of 1 s after the person falls down is analyzed. This provides a frequency resolution of 1 Hz, which is enough for our proposal but may be not accurate for modal analysis of the structure. We would like to underline here that the resolution of the method is linked to the ratio between the number of samples of the sequence and the Nyquist limit. Therefore, higher accuracy is just obtained by taking longer video sequences,.

In this experiment, the sequence was captured with a pocket camera CASIO Exilim ZR-1000. Since this is a low-cost camera, the only option for capturing was AVI-JPEG format so some noise from the compression algorithm is introduced in the sequence.

In Figure 7, we show a section of the bridge with a superimposed frequency color map with the corresponding numerical value. Notice that, because of the limitations on image quality and resolution of the low cost camera, the zoom and the view has changed with respect to Figure 6. According to the ROI size criterion above sketched about the size of the ROI and the vibrating elements in the scene, local neighborhoods have been taken of 8x8 px (Block size) with an overlap of 6 px in both vertical and horizontal directions.

In t Figure 7 we can observe that the region around the cable, deck and vertical beam provide clear results. Majority of local border neighborhoods around these areas give a consistent result. Additionally, we can appreciate some noise in the image, which may come from jpeg compression noise.

Figure 6. General view of the pedestrian bridge used for the experiment. A person jumping in the center excites the structure. An accelerometer in the upper part of the deck will register the main vibration.

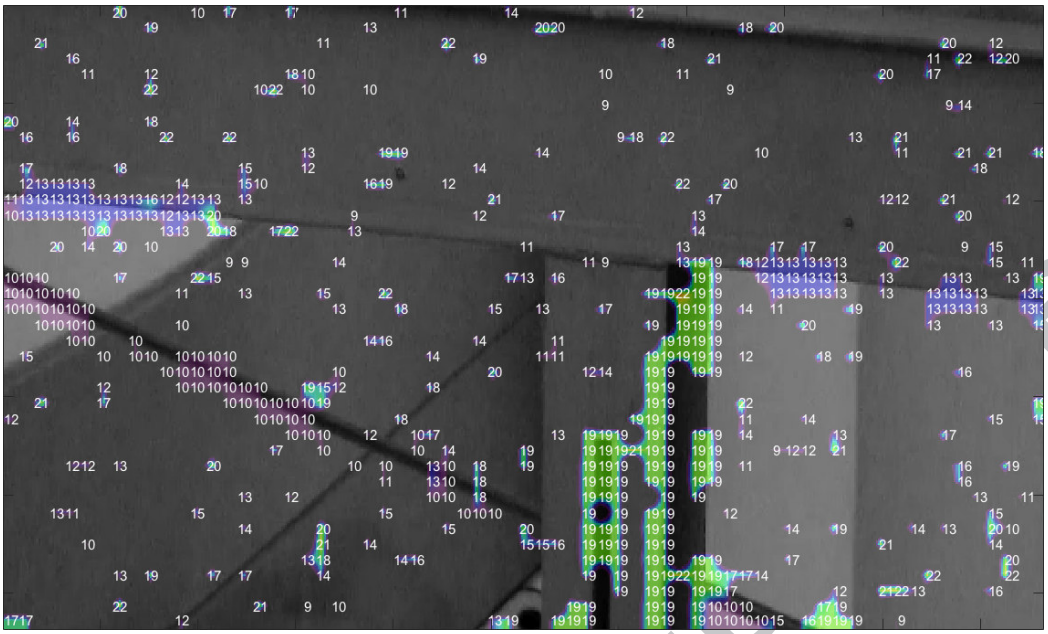

Figure 7. Color map showing the vibration frequencies of the different parts of the bridge. Notice the three main parts vibrating at 10, 13 and 19 Hz, respectively

The main drawback of this representation is that each ROI is considered independently and only the main peak is selected and depicted, with independence of its value or the value of the surrounding neighborhoods' peaks. Nevertheless, this procedure can reveal "false" peaks coming from noise, transient phenomena or any other non-controlled origin. Therefore, only the stronger peaks, whose height is above 5 times the standard deviation of the spectrum signal, are considered. Even so, spurious peaks still may appear in the map, as we can appreciate in Figure 7. These peaks seem not to be associated to any specific object so they can be ignored or even filtered for cleaner representations.

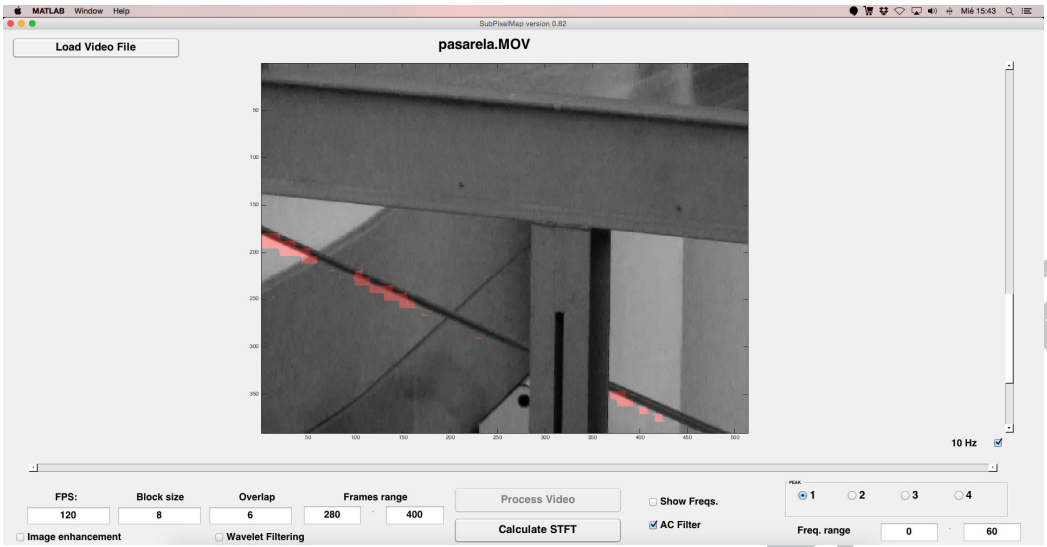

Figure 8 (video). Video sequence with the areas vibrating at a specific frequency highlighted in red. Notice the three main areas, cable, deck and vertical beam, vibrating at 10, 13 and 19 Hz, in accordance with the map in Figure 6. (See the video at https://goo.gl/D6OsaN)

In order to better understand the results in Figure 7, we have represented in Figure 8 the peaks location within the frame for each frequency. Since information is arranged in a three-dimensional matrix, the representation is displayed in a video sequence where each frame displays the map for one specific frequency. In this video, one can clearly see now the vibrational structure of bridge: the cable is vibrating at 10 Hz while the deck is vibrating at 13 Hz. The vertical beam attached to the deck is vibrating at 19 Hz. For the remaining frequencies up to the Nyquist limit at 60 Hz, only spurious dots appear but since they are isolated and not-connected to any significant part of the bridge, they may be classified as noise and neglected.

In the same experiment, the vibration was also measured with a uniaxial accelerometer in the center of the deck. In Figure 9, its frequency spectrum up to 100 Hz is depicted. Notice that, as it appears in the picture, in this example we are considering 120 frames at 120 fps so our accuracy is of ± 1 Hz. Therefore, we can conclude that the three main vibration frequencies detected by the accelerometer coincide with the peaks detected by our method.

We would like to underline that, since the accelerometer is a single point device, all frequencies are registered at the same location and one cannot distinguish the

origin of the vibration. The method proposed here is not only able to register the vibration, but also to locate its origin, thus providing valuable information about the structure behavior.

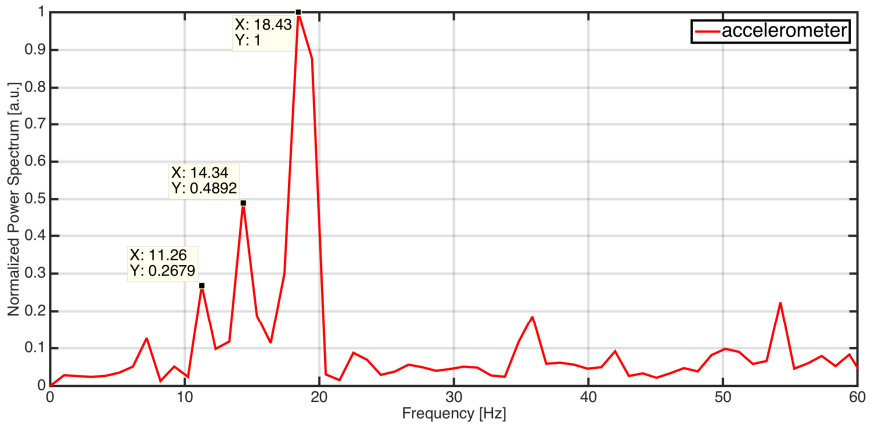

Figure 9. Accelerometer signal obtained from the bridge vibration. Notice the coincidence between the three peaks detected in the center of the deck and those obtained for the cable, deck and beam respectively in Figures 6 and 7.

Finally, and going one step further, we can combine both the frequency mapping feature with the time-frequency analysis in order to detect the changing vibration of an object and construct a dynamic frequency map. To this end, we have selected a different object, so that different frequencies can be excited at different times.

The object selected is a guitar and the sequence has been captured with an AOS XPRI camera at 1000 fps. In the video depicted on Figure 10, we show three strings of a guitar that are played sequentially. A black card behind the strings covers the resonance box of the guitar in order to avoid that its vibration overlaps the vibration of the strings and distorts the map.

In this case, the map has been calculated using ROIs of 6x6 px with an overlapping of 3 px. It has been overlapped to the video so that one can see the frequency value together with the movement, and thus obtain complete information about the scene at a glance. At the beginning of the sequence, there is some localized noise in the background, but when the strings start to vibrate,

this noise is cancelled and the correct frequencies appear. Notice that the ROIs are displaced form the object. This is because the algorithm takes a sub-sequence of frames to calculate the short-time Fourier transform and the peak is assigned to an average location.

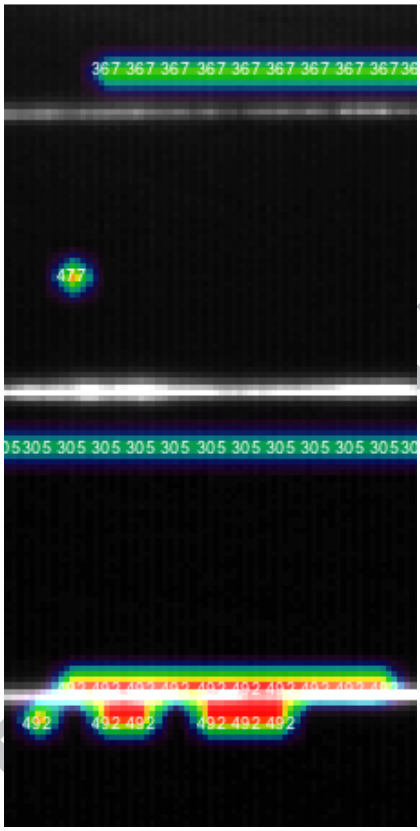

Figure 10 (video). Sequence showing the time-frequency map of three guitar strings vibrating at 367, 305 and 492 Hz, respectively (the guitar is not in tune).
(See the video at https://goo.gl/Pnf707)

Again, we would like to emphasize that the method is able to measure different vibrations that are happening simultaneously, and detect where and even when this vibration is produced, which is not possible with a microphone, or an accelerometer or even a laser vibrometer.

CONCLUSION

In this paper, we have developed a method for simultaneous multipoint measurement of vibration frequencies through the analysis of a high-speed video sequence. A complete algorithm for multipoint application is described and modifications to measure time-varying frequencies are also proposed. A convenient application of our method allows plotting a full-frame vibration map describing the frequency of vibration of all the objects in a scene.

Our method is demonstrated through different examples, also showing some of its potential applications. As we said through the text, with our proposal we are able to measure not only the vibration frequency but also to separate the element or even the part of it that is vibrating with a certain frequency and segment it from other vibrating parts or even elements vibrating at a different moment.

The method has been implemented with scientific mid-end cameras and also with pocket cameras, with relatively poor quality and image-compression noise. Therefore, it showed to be versatile and robust against some noise level in the image.

In order to facilitate the analysis and further development of our work, the main algorithms for multi-ROI measurement is freely disposable in Matlab format at [20]. The algorithms for calculating and displaying the maps are still under development but beta versions will also be distributed under personal requirement to the authors.

Finally, we want to underline that although other more precise methods exist, they are limited to single point measurements (accelerometer, laser vibrometer) or to large areas integration (microphone). Therefore, our proposal may be a reliable complement to these devices and provide useful information about complex problems.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Spanish Ministerio de Economía y Competitividad through the project BIA2011-22704. B. Ferrer acknowledges the support of the University of Alicante through the project GRE 13-10.

REFERENCES

[1] Tabatabai, H., Oliver, D. E., Rohrbaugh, J. W., & Papadopoulos, C. (2013). Novel Applications of Laser Doppler Vibration Measurements to Medical Imaging. *Sensing and Imaging: An International Journal*, *14*(1-2), 13–28. http://doi.org/10.1007/s11220-013-0077-1

[2] Poh, M.-Z., McDuff, D. J., & Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Optics Express, 18(10), 10762-10774. http://doi.org/10.1364/OE.18.010762

[3] Krishnan, C., Washabaugh, E. P., & Seetharaman, Y. (2014). A low cost real-time motion tracking approach using webcam technology. Journal of Biomechanics, 48(3), 544–548. http://doi.org/10.1016/j.jbiomech.2014.11.048

[4] Wang, Z., Kieu, H., Nguyen, H., & Le, M. (2015). Digital image correlation in experimental mechanics and image registration in computer vision: Similarities, differences and complements. Optics and Lasers in Engineering, 65, 18–27. http://doi.org/10.1016/j.optlaseng.2014.04.002

[5] Lee, J.J. and Shinozuka, M. (2006), A vision-based system for remote sensing of bridge displacement, NDT & E Int, 39(5), 425-431; http://dx.doi.org/10.1016/j.ndteint.2005.12.003

[6] Ji, Y.F. and Chang, C.C. (2008), Nontarget image-based technique for small cable vibration measurement, J. Bridge Eng., ASCE, 13(1), 34-42; http://dx.doi.org/10.1061/(ASCE)1084-0702(2008)13:1(34)

[7] Caetano, E., Silva, S. and Bateira, J. (2011), A vision system for vibration monitoring of civil engineering structures, Exp. Tech., 35(4), 74-82; http://dx.doi.org/10.1111/j.1747-1567.2010.00653.x

[8] Mas, D., Espinosa, J., Roig, A. B., Ferrer, B., Perez, J., & Illueca, C. (2012). Measurement of wide frequency range structural microvibrations with a pocket digital camera and sub-pixel techniques. *Applied Optics*, *51*(14), 2664-2671. http://doi.org/10.1364/AO.51.002664

[9] Fukuda, Y., Feng, M. Q., Narita, Y., Kaneko, S., & Tanaka, T. (2013). Vision-Based Displacement Sensor for Monitoring Dynamic Response Using Robust

Object Search Algorithm. *IEEE Sensors Journal*, *13*(12), 4725–4732. http://doi.org/10.1109/JSEN.2013.2273309

[10] Busca, G., Cigada, A., Mazzoleni, P., & Zappa, E. (2013). Vibration Monitoring of Multiple Bridge Points by Means of a Unique Vision-Based Measuring System. *Experimental Mechanics*, *54*(2), 255–271. http://doi.org/10.1007/s11340-013-9784-8

[11] Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G. J., Durand, F., & Freeman, W. T. (2014). The visual microphone. *ACM Transactions on Graphics*, *33*(4), 1–10. http://doi.org/10.1145/2601097.2601119

[12] Simoncelli, E. P., & Freeman, W. T. (1995). Steerable pyramid: a flexible architecture for multi-scale derivative computation. In *IEEE International Conference on Image Processing* (Vol. 3, pp. 444–447). IEEE. Retrieved from http://doi.ieeecomputersociety.org/10.1109/ICIP.1995.537667

[13] Chen, J. G., Wadhwa, N., Cha, Y.-J., Durand, F., Freeman, W. T., & Buyukozturk, O. (2015). Modal identification of simple structures with high-speed video using motion magnification. *Journal of Sound and Vibration*, *345*, 58–71. http://doi.org/10.1016/j.jsv.2015.01.024

[14] Wang, Z., Nguyen, H., & Quisberth, J. (2014). Audio extraction from silent high-speed video using an optical technique. *Optical Engineering*, *53*(11), 110502. http://doi.org/10.1117/1.OE.53.11.110502

[15] Ferrer, B., Espinosa, J., Roig, A. B., Perez, J., & Mas, D. (2013). Vibration frequency measurement using a local multithreshold technique. *Optics Express*, *21*(22), 26198-26208. http://doi.org/10.1364/OE.21.026198

[16] Gentile, C. and Bernardini, G. (2010), An interferometric radar for non-contact measurement of deflections on civil engineering structures: laboratory and full-scale tests, J. Struct. Infrastruct. Eng., 6(5), 521-534; http://dx.doi.org/10.1080/15732470903068557

[17] Gentile, C. (2010), Deflection measurement on vibrating stay cables by non-contact microwave interferometer, NDT & E Int., 43(3), 231-240; http://dx.doi.org/10.1016/j.ndteint.2009.11.007

[18] Luzi, G., Crosetto, M., & Cuevas-González, M. (2014). A radar-based monitoring of the Collserola tower (Barcelona). *Mechanical Systems and Signal Processing*, *49*(1-2), 234–248. http://doi.org/10.1016/j.ymssp.2014.04.019

[19] Shen, H., He, B., Zhang, J., & Chen, S. (2015). Obtaining four-dimensional vibration information for vibrating surfaces with a Kinect sensor. *Measurement*, *65*, 149–165. http://doi.org/10.1016/j.measurement.2014.12.019

[20] Mas, D., Ferrer, B., Sheridan, J. T., & Espinosa, J. (2012). Resolution limits to object tracking with subpixel accuracy. Optics Letters, 37(23), 4877–4879. http://doi.org/10.1364/OL.37.004877

[21] Smith, S.W. (1997) The Scientist and Engineer's Guide to Digital Signal Processing, California Technical Publishing, San Diego, CA, USA.

[22] The main program, auxiliary files and a video example can be downloaded from http://goo.gl/g8jHdg

[23] Espinosa, J., Ferrer, B., Mas, D., Perez, J., Roig, A.B., Universidad de Alicante "Método y sistema para medir vibraciones" Patent pending nº P201300498 (05-23-2013)

[24] Ferrer, B., Espinosa, J., Perez, J., Ivorra, S., Mas, D., 2011. Optical Scanning for Structural Vibration Measurement. Res. Nondestruct. Eval. 22, 61–75. http://doi.org/10.1080/09349847.2010.519137

[25] References to the Short Fourier Transform can be found elsewhere. A general description and properties can be followed from http://en.wikipedia.org/wiki/Short-time_Fourier_transform

[26] Auger, F., Lemoine, O., Gonçalves, P., Flandrin, P. The Time-Frequency Toolbox, at http://tftb.nongnu.org

Highlights

- Simultaneous multipoint measurement of vibration frequency
- Calculation and display of a frequency map over a scene
- Time-frequency analysis of video sequences
- Analysis of complex-object's vibration at a glance
- Easy implementation
- Free access to main algorithms
- Wide applicability.