Virtual Visual Sensors and Their Application in Structural Health Monitoring

Yi-Zhe Song[a], Chris R. Bowen[c], H Alicia Kim[c],
Aydin Nassehi[c], Julian Padget[b], Nick Gathercole[c]

[a]School of Electronic Engineering and Computer Science, Queen Mary, University of London, London E1 4NS, UK
[b]Department of Computer Science, University of Bath, Bath, BA2 7AY, UK
[c]Department of Mechanical Engineering, University of Bath, Bath, BA2 7AY, UK

ABSTRACT

Wireless Sensor Networks (WSN) are being increasingly accepted as an effective tool for Structural Health Monitoring (SHM). The ability to deploy a wireless array of sensors efficiently and effectively is a key factor in SHM. Sensor installation and management can be difficult in practice for a variety of reasons: a hostile environment, high labour costs and bandwidth limitations. We present and evaluate a proof-of-concept application of Virtual Visual Sensors (VVS) to the well-known engineering problem of the cantilever beam, as a convenient physical sensor substitute for certain problems and environments. We demonstrate the effectiveness of VVS as a means to achieve non-destructive evaluation (NDE). Major benefits of VVS are its non-invasive nature, ease of installation and cost effectiveness. The novelty of VVS lies in the combination of *marker extraction* with *visual tracking* realized by modern computer-vision algorithms. We demonstrate that by deploying a collection of VVS on an oscillating structure, its modal shapes and frequencies can be readily extracted from a sequence of video images. Subsequently, we perform damage detection and localization by means of a wavelet-based analysis. The contributions of this paper are: (i) use of a sub-pixel accuracy marker extraction algorithm to construct virtual sensors in the spatial domain; (ii) embedding dynamic marker linking within a tracking-by-correspondence paradigm that offers benefits in computational efficiency and registration accuracy over traditional tracking-by-searching systems; (iii) validation of VVS in the context of a SHM application.

Keywords: structural health monitoring, non-destructive damage detection, computer vision, wavelets

1. INTRODUCTION

Structural Health Monitoring (SHM) has been an active research area in recent decades and is considered to be a problem of major importance due to its potential benefits in terms of safety, reliability and economy. The ability to detect damage in a structure is a key component of any modern SHM system and Non-Destructive Evaluation (NDE) is an important element because of its convenient and non-invasive nature. In particular, NDE methods based on modal parameters and shapes obtained using vibration-based testing have been shown to be effective.[1, 2] The cornerstone of vibration-based damage detection lies in the dual-mapping of modal parameters with changes in the physical properties of a structure as a result of damage. Essentially, such mapping transfers the feature space in which damage detection algorithms operate from that of mass, damping and stiffness to one constructed from modal frequencies,[3] modal damping and modal shapes.[4, 5] Among these modal properties, modal shapes are widely used as an input for damage detection algorithms.

Methods for effectively acquiring structural modal shapes have been developed by many researchers. This ranges from small scale modal shape acquisition units such as laser doppler vibrometers,[6] to Wireless Sensor Networks (WSN),[7, 8] which are commonly regarded as a standard for large scale SHM applications.[9–11] Initially, sensor networks used relatively large sensing nodes that were physically linked to each other and to a base station.[7] Advances in sensor technologies and micro-electro-mechanical systems (MEMS) have facilitated the mass production of small sensors, often using wireless communication. This has led to the concept of WSNs that have the potential to resolve the flexibility and scalability obstacles of traditional sensor networks. Much research has been conducted in the field of WSN design such as network topology,[12] power consumption,[13] communication[14] and optimisation.[15] Progress on the theoretical aspects of sensor networks has enabled a surge in the number of practical implementations of sensor networks. Successful multi-disciplinary applications include environmental and habitat monitoring,[16] healthcare,[17] energy monitoring[18] and, in particular, the field of SHM.

As an example, traditional[4] experimental modal analysis uses an array of accelerometers mounted onto a structure to determine its Frequency Response Functions (FRFs), which collectively form modal shapes via various forms of curve fitting. Such approaches have two common drawbacks: (i) adding sensors can add undesirable damping effects on the structure; altering its modal properties[19] and (ii) the total number of deployable sensors is limited by the physical form of the accelerometers. This places a bottleneck on the spatial resolution of the acquired modal shapes, which makes damage localisation less accurate.[2] In addition, if the accelerometers are not wireless, a significant quantity of cabling must also be laid throughout the structure.

The aim of this paper is to demonstrate the feasibility of Virtual Visual Sensors (VVS) as an alternative form of low cost, zero-mass sensor. Their realization is made possible by two factors: (i) recent advances in computer vision, specifically in the areas of saliency detection[20, 21] and tracking[22] that enable features to be extracted and matched in a robust manner and (ii) increased availability of cost-effective camera sensor nodes and developments in Wireless Visual Sensor Networks (WVSN).[23]

In contrast to traditional physical sensors, VVS are implemented as features extracted from images/videos that can be easily transmitted and stored via a WVSN. Installing a VVS simply involves placing/painting markers in the sensing environment and pointing a camera sensor at them. Such a setup has two important benefits: first, managing sensors becomes easy since they are freely available and easily (re-)deployable; second, a single camera node can read many VVS and efficiently transmit the acquired data across a network. In addition, it seems feasible to apply to large-scale structures by aligning multiple cameras to capture the complete scene at a required resolution, then using video synchronization and frame stitching, it should be possible to interrogate an entire structure using the VVS technique. The feasibility of such an approach remains to be fully investigated and is outside the scope of this paper."

Other researchers have also advocated the use of camera nodes for structural damage detection. Patsias and Staszewskiy[24] extracted the edges of a cantilevered beam using a standard wavelet-based algorithm. Discretely stored edge profiles were treated as modal shapes and the authors used a wavelet-based approach to detect damage. Recently, Shi et. al.[25] incorporated cameras into the damage detection process for an aluminium cantilever beam. Modal shapes were obtained via standard edge detection using commercial software and the modal shapes were represented at discrete pixel locations where local jaggedness of edges are naturally inherited.

In summary, the contributions of this paper are in the novel application of computer vision to the domain of structural health monitoring, specifically:

- We examine VVS as a convenient and effective physical sensor substitute; their development parallels recent advances in Wireless Visual Sensor Networks, where camera nodes play increasingly important roles.
- We show that a sub-pixel accuracy Hough Transform based algorithm can extract markers in a robust and accurate manner; our method has significant advantages in computational speed and transmission efficiency against traditional image processing methods such as edge detection.[25]
- We apply a non-rigid point set registration technique[26] to establish temporal linking of markers across video frames. Acquiring marker behaviour in the temporal domain allows us to calculate their displacement, velocity and acceleration. By implementing tracking-by-correspondence, our method can deliver high-resolution tracking of an oscillating structure — an important requirement for practical WVSN applications.
- We demonstrate, through controlled lab experiments,[24, 25] the use of sub-pixel VVS to acquire modal shapes and frequencies of structures, and their use in a wavelet-based structural damage detection algorithm. Experiments using oscillating steel cantilever beams have been conducted to evaluate performance.

2. METHOD

Our method starts with an oscillating structure, upon which markers have been placed, and applies computer- vision algorithms to video data to extract and link the markers from successive frames. These markers are the virtual visual sensors.
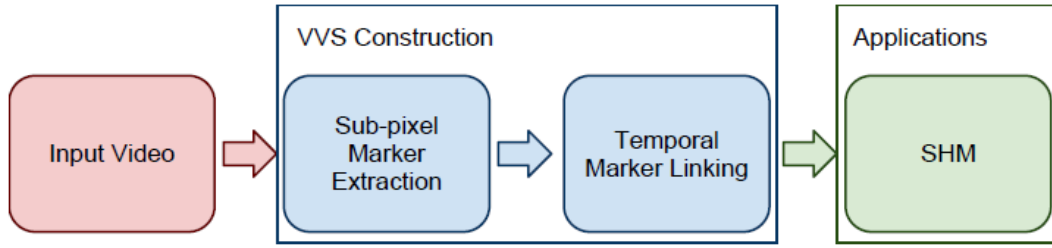
Figure 1. A system diagram of the proposed SHM scheme

Camera motion estimation or camera tracking, i.e., working out relative locations of the camera at a given time, has been a relatively well studied problem in computer vision, where real-time solutions[27] and off-the-shelf commercial products □ already exist. In essence, the problem asks for feature points extracted from the scene to be tracked in real-time, through which relative position of the camera can be estimated.

Here, we assume that camera motion is already decoupled through a pre-processing step. This could be done by extracting feature points, such as SIFT,[28] from the video background (i.e., on regions outside of the structure under monitoring) at each frame. Camera motion can then be estimated using any of the techniques mentioned above, and recorded for later decoupling from motion of the structure.

For now, we do not consider the matter of camera motion further, but focus on the movement of the virtual visual sensors, which is processed to obtain modal shapes for subsequent wavelet analysis, resulting in damage detection and localization. The stages of the method are summarized in Figure 1. We first extract markers from each video frame, using a modified version of the Hough Transform[29] (HT) that yields marker locations with sub-pixel accuracy. We then formulate marker tracking as a point set registration/correspondence problem where markers are treated as a set of points in space which are aligned with another existing set from a subsequent video frame. This is different from traditional tracking by searching,[22] which seeks new positions of markers in frame $t + 1$ by searching for those already detected in frame t. Tracking by correspondence is better in this particular context since sub-pixel accuracy is maintained, contributing overall to system accuracy and thereby allowing it to reflect small changes in position that would otherwise be masked. We now describe marker extraction and point set matching in more detail.

2.1 Marker Extraction

There are two key critical properties for marker extraction: accuracy and efficiency, both of which are achieved by an enhanced version of the Hough Transform (HT), that we describe here. Traditional HT[29] and its variants generally locate features with pure geometric shapes, i.e, lines, circles and squares. Because of the geometric nature of Hough features, HT is more tolerant of discontinuities and more robust to noise compared to marker-less feature extraction techniques. Although modern computer vision algorithms are able to detect points of interest that are useful in applications, such as 3D reconstruction and object recognition, they are sensitive to outliers and post-processing is commonly required. However, combining markers and HT leads to much more robust detection. Furthermore, the Hough Transform can also be computed very efficiently, even in real-time.[30]

For application to SHM, we use the Circular Hough Transform (CHT)[29] to detect markers on an oscillating structure. CHT determines the radius, R, of a circle by locating points that lie on its perimeter. The parametric form of a circle is often used:
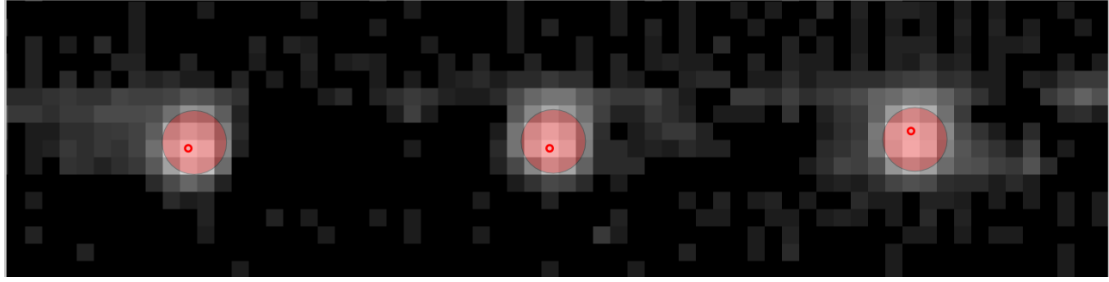
$$x = a + R\cos(\theta) \tag{1}$$

$$y = b + R\sin(\theta) \tag{2}$$

where R is the unknown radius and (a, b) is the centre. θ determines the location of a point, (x, y), on the perimeter of the circle. The triplet (a, b, R) constitutes the Hough parameter space of an image and a CHT algorithm searches such a space to locate circles. A voting procedure is then used to build an accumulation array,
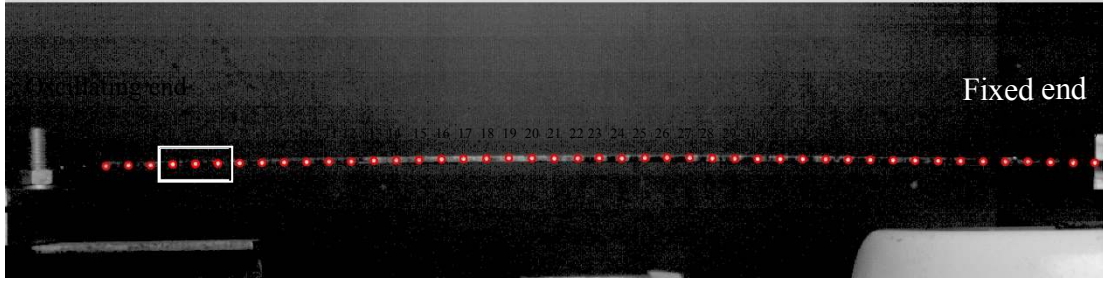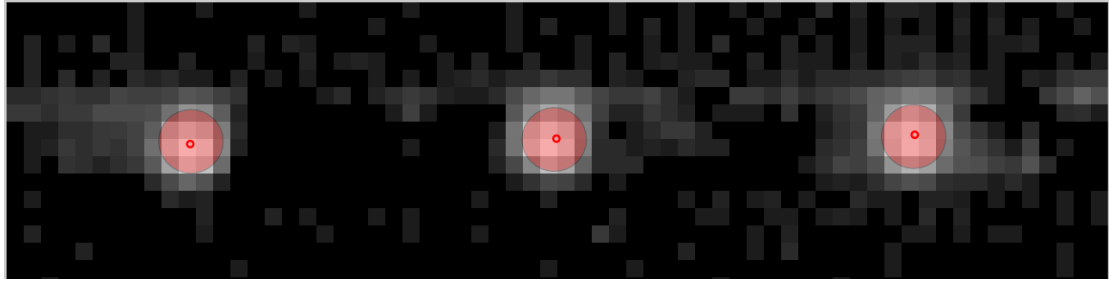
(a) marker result for the whole beam

(b) traditional result for the highlighted region in (a)

(c) sub-pixel accurate result for the highlighted region in (a)

Figure 2. Marker Extraction for a Steel Cantilever Beam (Beam Length: 463mm, Marker Spacing 10mm)

A, where local maxima correspond to circle centres in the image space. A is a two-dimensional accumulator array that has the same size as the input image and each dimension corresponds to quantised values for a and b.

Figure 2(a) is a frame from a video of an oscillating cantilever beam, with 45 markers and fixed at the right- hand end. To demonstrate the difference resulting from the sub-pixel VVS, we zoom into the region identified by the box in Figure 2(a). Circle centres detected using traditional CHT may be offset from the true centre, as can be seen in Figure 2(b). This is largely due to the limit on the spatial resolution of images; an inevitable result of the discrete sampling nature of digitised images. However, using the sub-pixel VVS, more accurate positioning is achieved, as Figure 2(c) shows. We put forward a quantitative assessment of the benefits of the sub-pixel VVS in section 3.

In enhancing the accuracy of the HT, we follow a similar approach to that developed by Lowe,[31] but modified to find sub-pixel locations of circle centres. The method works by using a Difference of Gaussians (DoG) scale-space:

$$D(x, y, \sigma) = G(x, y, k_i\sigma) * I(x, y)$$
$$- G(x, y, k_j\sigma) * I(x, y)$$

where I is a greyscale image (two-dimensional array of pixel intensities), $G(x, y, k_i\sigma)$ is a Gaussian blur kernel centred around a point (x, y) of scale k_i ($i < j$ are kernel scale indices and σ is the 'octave' scale) and '□' stands for the convolution operation. Here, we use the previously calculated Hough accumulation array, A, as the input image, $I = A$. The true maxima of A are then estimated by interpolating neighbouring pixels of each provisional one and fitting a 3D quadratic function to them. A Taylor expansion (up to the quadratic term) of the DoG scale-space function $D(x, y, \sigma)$ is used for this interpolation, which is given by:

$$D(x) = D + \frac{\partial D^T}{\partial x}x + \frac{1}{2}x^T\frac{\partial^2 D}{\partial x^2}x, \tag{3}$$

where D, $\partial D/\partial \mathbf{x}$ and $\partial^2 D/\partial \mathbf{x}^2$ are evaluated at the provisional centre $\bar{c}$, obtained via the standard HT. The algorithm then seeks to find a suitable offset $\mathbf{x} = (x, y, \sigma)$ that moves $\bar{c}$ to its true centre $\hat{c}$. It follows that its extremum, $\hat{\mathbf{x}}$, can be found by taking the derivative of $D(\mathbf{x})$ with respect to x and setting it to zero, giving

$$\hat{x} = -\frac{\partial^2 D(x)^{-1}}{\partial x^2} \frac{\partial D(x)}{\partial x}.$$

(4)

If $\hat{\mathbf{x}}$ is larger than a threshold in any of its three dimensions, the candidate centre is updated to be the new offset point where the same interpolation procedure is performed again; otherwise, the offset is desirable, and hence used to find the true maximum using $\hat{c} = \bar{c} + \hat{\mathbf{x}}$.

2.2 Temporal Linking of Markers

The second step of the process is to link the markers temporally to enable the analysis of their dynamic characteristics, which provides the reference for the movement of the beam. There are two approaches to tracking:
(i) searching, in which the target objects are identified in frame one and the tracking process then locates in- stances of the same objects in subsequent frames; (ii) correspondence, in which targets are first found in each frame and the task is then to find the correspondences between consecutive frames. We have chosen the second approach, since it preserves the previously acquired sub-pixel accuracy marker positions and because the conventional search approach results in integer coordinates that require a separate refinement pass to deliver sub-pixel results.

The process treats the markers from each video frame as a point set, where $\mathbf{c}^t$ is a column vector for a single frame at time t, where N is the total number of markers. Consequently, the tracking problem becomes that of finding one-to-one correspondences – called registration – between consecutive point sets, $\mathbf{c}^t$ and $\mathbf{c}^{t+1}$ (a forward mapping in this context). This mapping process not only retains marker accuracy but also reduces incorrect tracking, since markers are pre-extracted using sub-pixel Hough Transform and registration is established globally across marker sets. We use the Coherent Point Drift (CPD) algorithm developed by Myronenko and Song[26] to establish these correspondences, due to its robustness in outlier rejection and computational efficiency via Fast Guass Transform.

CPD finds a mapping P between two point sets, $\mathbf{c}^t$ and $\mathbf{c}^{t+1}$, that is:

$$M : c^t \xrightarrow{\quad P \quad} c^{t+1}.$$

(5)

which captures point set registration as a probabilistic density estimation problem. Each point set is represented as a single Gaussian mixture model (GMM), where each point represents the centroid of an individual Gaussian component. The algorithm then seeks to align two consecutive point sets by finding the maximum of the GMM posterior probability for a given point set. The key concept of the registration process lies with the preservation of point set topology, achieved by constraining GMM centroids to move as a group.

Thus, the point set of one frame, $\mathbf{c}^t$, are treated as centroids of a GMM model and those of the next, $\mathbf{c}^{t+1}$, as data points to be fitted, assuming equal cardinality, since the number of points/sensors is constant. Let m = 1, 2, ..., N denote all GMM components (a single Gaussian) and $0 \leq w \leq 1$ be a weighting factor (often determined experimentally), we can then write the GMM probability density function as:
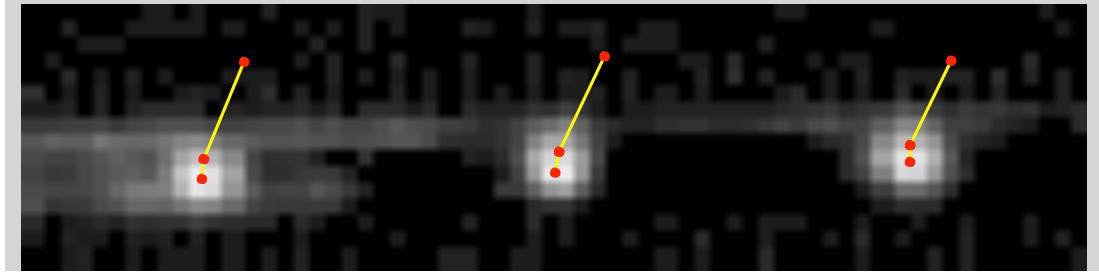
$$p(c^{t+1}) = \omega \frac{1}{N} + (1 - \omega) \sum_{n=1}^{N} P(m) p(c^{t+1} \mid m),$$

(6)
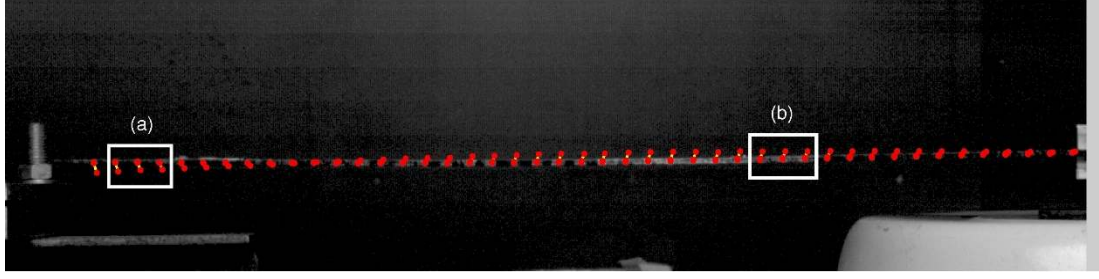
(a) Tracking markers across frames 1–25; boxes identify enlarged regions in Figures 3(b) and 3(c)
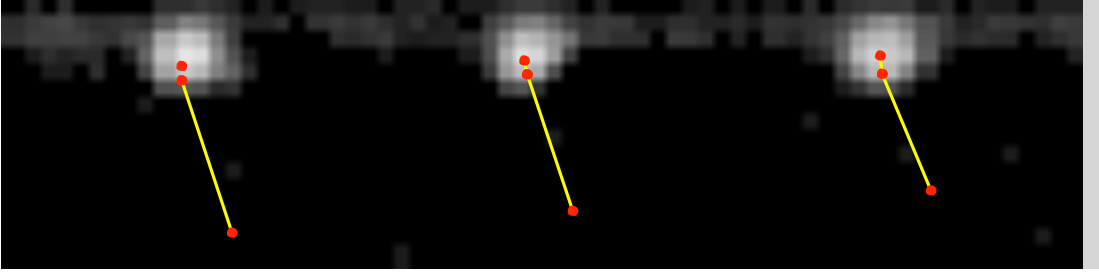
(b) Enlargement of box (a); bottom, middle and top row markers correspond to Frame 25, 5 and 1

(c) Enlargement of box (b); bottom, middle and top row markers correspond to Frame 1, 5 and 25

Figure 3. Marker tracking an oscillating steel cantilever beam (frequency of oscillation 66.5Hz, beam length 463mm): VVS are red dots with trajectories in yellow. Marker spacing: 10mm, beam fixed at right-hand end.

Where $P(m) = \dfrac{1}{N}$ and

$$p(c^{t+1} \mid m) = \frac{1}{(2\pi\sigma^2)} \exp^{-\frac{\|c^{t+1}-c_m^t\|^2}{2\sigma^2}}. \tag{7}$$

The following negative log-likelihood function is then minimised using the Expectation-Maximization algorithm[32] to find θ and σ^2:

$$E(\theta,\sigma^2) = -\sum_{n=1}^{N} \log \sum_{m=1}^{N+1} P(m)\, p(c_n^{t+1} \mid m), \tag{8}$$

where θ is the re-parameterisation of GMM centroids and σ are the corresponding variances.

Figure 3 shows markers tracked across the first 25 video frames of a beam oscillating at 66.5Hz. Red dots identify markers tracked from frame 1 to frame 5 then to frame 25, where yellow lines show the trajectories between these frames. We zoom in on two regions of the beam in Figures 3(b) and 3(c) to reveal how individual markers give rise to trajectories over a sequence of frames. This demonstrates that the tracking algorithm can successfully track markers even when shape deformation is relatively large (frame 1 to frame 25).

Now that the VVS construction process has been discussed, the following section describes the experimental methods.
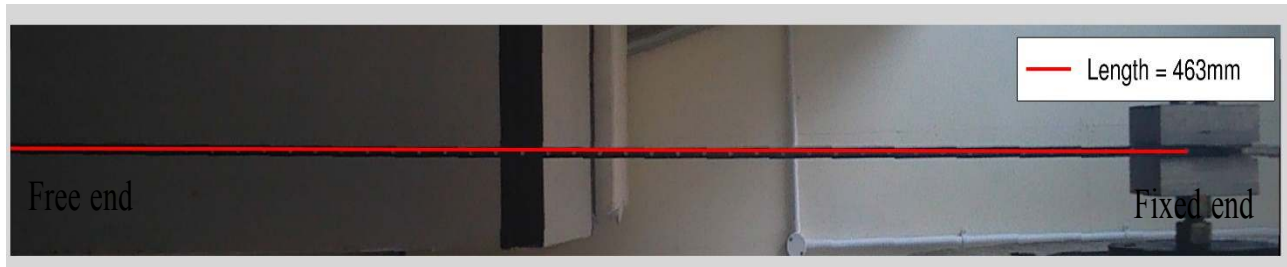
(a) Picture of the experiment setup
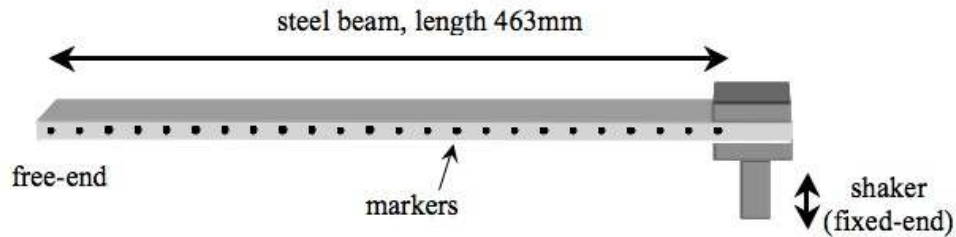
(b) Schematic of the experiment setup

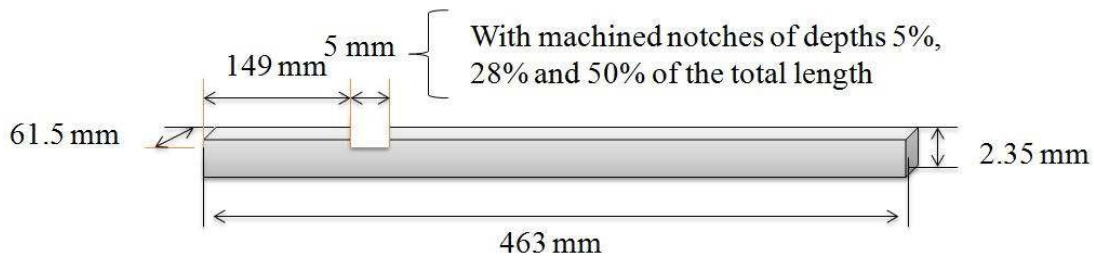Figure 4. Picture and schematic of the experimental setup

Figure 5. Schematic diagram of the three steel beams

3. EXPERIMENTS AND RESULTS

We evaluate the performance of the VVS within a SHM setting and demonstrate how it can generate modal shapes and modal frequencies of a steel cantilevered beam. We validate the VVS technique by showing that it can provide sufficiently accurate data to identify and localise varying levels of damage.

3.1 Experimental Setup

We use one steel beam of dimension 2.35mm x 61.5mm x 463mm, with one end attached to a shaker and the other end free to oscillate. A slow-motion camera (FASTCAM SA3 model 120K-M2) with a spatial resolution of 1024 x 256 and frame rate of 2000 fps is used to record the motion of the beam. A picture of such a set up can be found in Figure 4.
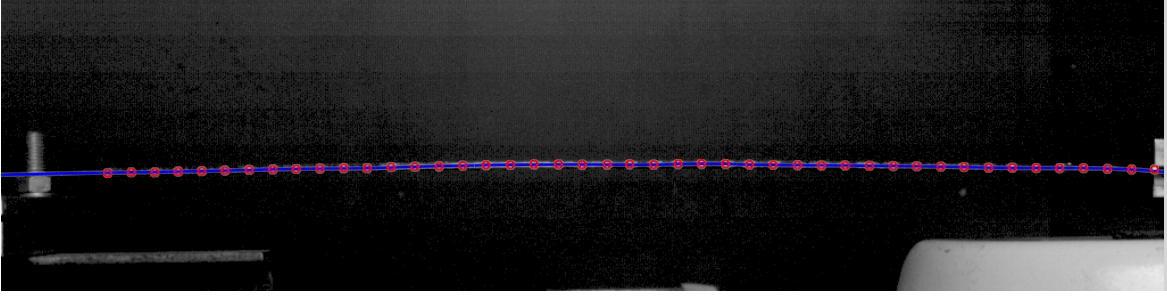
To introduce damage into the structure, a 5mm wide notch of increasing depth (5%, 28% and 50% respectively) was machined at a distance of 149mm (1/3 of the total length) from the free end (see Figure 5 for a schematic). Prior to the introduction of damage the beam was tested as a reference. The beam is always clamped at a fixed position to ensure invariance of boundary condition.

For each of the four beams, we put 45 white circular markers along its edge and record the motion of each beam resonating under its second modal frequency. Experiments were conducted on second modal shapes, with the downstream damage detection application in mind – because they exhibit more local bending than mode one – which in turn amplifies the damage response 35.
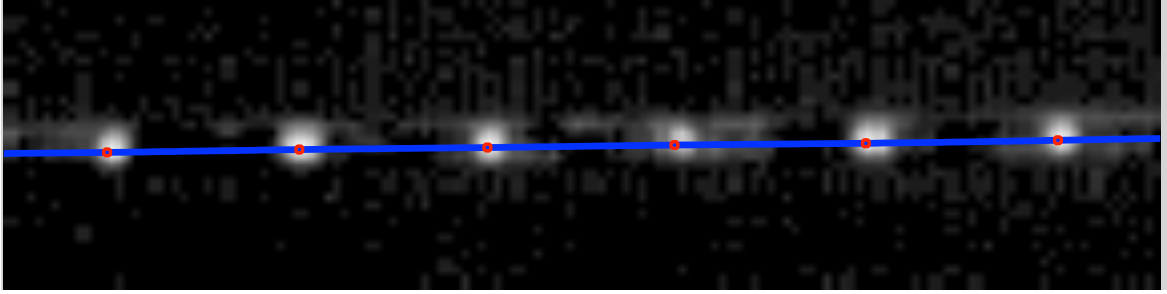
Table 1. Overall Mean Curvatures of All Six Motion Clips

Clip ID	Overall mean curvature of sub-pixel VVS, $K_{sub-pixel}$	Overall mean curvature of standard VVS, $K_{standard}$	Improvement over standard VVS, $(K_{standard} - K_{sub-pixel})/K_{standard}(\%)$
1	20.85	63.38	67.10%
2	25.20	64.82	61.12%
3	18.27	55.66	67.17%
4	22.34	65.48	65.88%
5	19.66	57.90	66.05%
6	17.36	63.52	72.66%
Average	**20.61**	**61.80**	**66.66%**

(a) modal shape via spline fitting

(b) a zoomed-in section of the modal shape in (a)

Figure 6. Modal Shape (Mode 2 at 66.5Hz) Reconstruction Result

3.2 Modal Shape Acquisition

With the aid of VVS, we now show that modal shapes can be reliably and accurately reconstructed. Given a set of markers extracted from frame t, $\mathbf{c}^t$, the modal shape is built by fitting a cubic interpolating spline $\mathbf{s}$ to $\mathbf{c}^t$, that is

$$s = cubic\_spline(c^t).$$

Figure 6(a) shows a modal shape (Mode 2 at 66.5Hz) reconstructed using the above spline interpolation procedure (the modal shape is drawn in blue with markers in red). A close-up of part of the same modal shape is provided in Figure 6(b). As can be seen, the spline reconstructed using VVS is smooth and rests near the true median axis of the beam.

From a theoretical perspective, modal shapes are perfectly smooth splines. However, their experimental realisations do not always conform to such a smoothness constraint. Local discontinuities often exist in experimentally reconstructed modal shapes, due to noise introduced by the measurement devices. These local discontinuities make analysis difficult and are known to generate false positives in damage detection algorithms.[33]
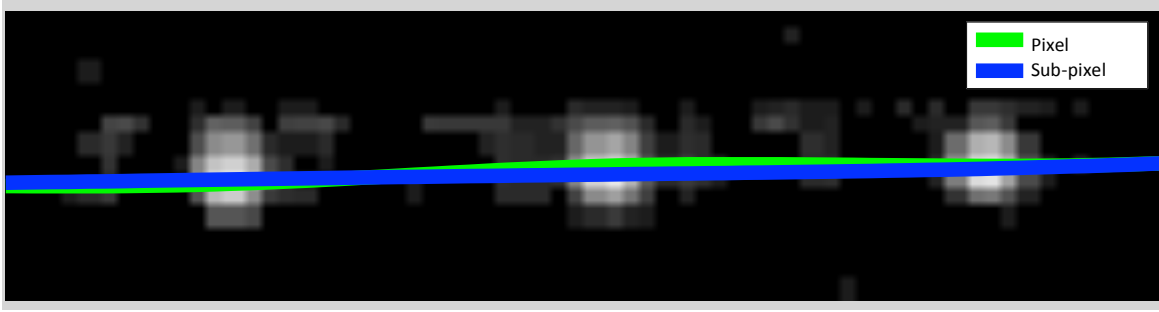
Figure 7. Modal Shapes (Mode 2 at 66.5Hz): sub-pixel VVS (blue) vs. standard VVS (green), marker spacing 10mm

Table 2. Modal Frequency Comparisons

| Beam ID	Frequency with sub-pixel VVS (Error %)	Frequency with standard VVS (Error %)	Actual frequency (Ground truth)
1(5%)	65.6Hz (1.35%)	64.8Hz (2.55%)	66.5Hz
2(28%)	64.5Hz (2.12%)	63.7Hz (3.34%)	65.9Hz
3(50%)	62.5Hz (1.11%)	61.5Hz (2.69%)	63.2Hz
undamaged	66.4Hz (0.74%)	65.2Hz (2.54%)	66.9Hz

Figure 7 offers a qualitative comparison between modal shapes (Mode 2 at 66.5Hz) acquired using the sub-pixel VVS approach and using standard VVS. It is observed that standard VVS (green) yields modal shapes that tend to vary locally, whereas that reconstructed using our approach (blue) appears to contain less local variation.

To quantify the benefits of our method, we conduct an experiment using the *mean curvature* of a spline to measure the 'smoothness' of a modal shape. We compare modal shapes computed using our sub-pixel VVS approach against those obtained via standard VVS. For each of the six motion clips (three beams oscillating at their first and second modal frequencies), we first calculate the mean curvature for each frame, we then sum curvature values over all frames and divide by the number of frames to calculate the *overall mean curvature* for that clip. The results are summarised in Table 1, where sub-pixel VVS leads to an average of *67%* improvement in smoothness over modal shapes obtained via standard VVS. We note that our current Matlab implementation of the sub-pixel marker extraction algorithm takes 0.3978 sec on average for one video frame, whereas pixel accuracy extraction yields an average of 0.2154 sec, i.e. 46% less computational cost (on a typical Core2Duo 2.4GHz machine with 4GB of RAM). Both techniques delivered speeds some way from real-time (0.0167 sec per frame, 60 frames per second). However, an implementation in C could improve performance significantly. Indeed, a recent paper[34] has demonstrated that it is possible to perform standard circular Hough transform in real-time.

3.3 Modal Frequency Acquisition

Experimental modal frequencies are normally calculated by identifying peaks of Frequency Response Functions (FRF), obtained by applying a Fourier Transform to data from sensors such as accelerometers. Because of its non-invasive nature, VVS naturally overcomes the spatial constraints of physical sensors and here we show that Power Spectral Density (PSD) computed using VVS can identify modal frequencies. Again, to demonstrate the benefit of our approach we compare modal frequencies obtained from both pixel and sub-pixel accuracy PSDs.

To determine modal frequencies, VVS is first constructed for a given video clip. Then, a time vs. displacement signal is computed for each VVS marker, from which a PSD is generated. Figure 8 shows this signal at sub-pixel (red) and pixel (blue) accuracy from three selected VVS on an undamaged beam resonating under mode 2. Three VVS (markers 6, 16 and 26 in Figure 2(a)) with decreasing degrees of displacement were chosen. Figure 8(a) is the displacement function of the marker with the largest displacement, while Figure 8(c) has the least. Figures 8(d)–8(f) show PSDs constructed using the sub-pixel displacement functions given in Figures 8(a)–8(c), respectively. As can be seen, peaks corresponding to modal frequencies are clearly identifiable in all three cases. Moreover, the magnitude of the frequency response tends to positively correlate with the decreasing degrees of marker displacements. Similarly, Figure 9 illustrates displacement functions and corresponding PSDs for a damaged beam with a 5mm wide notch with a depth of 5%.
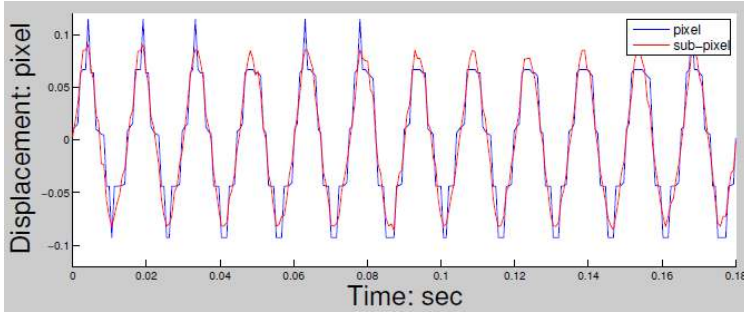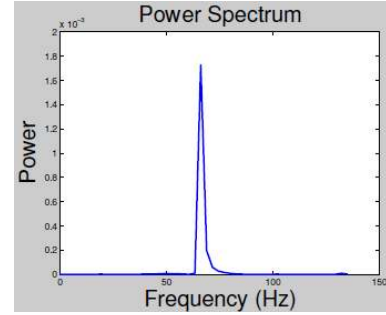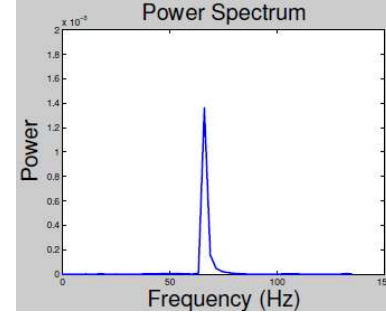
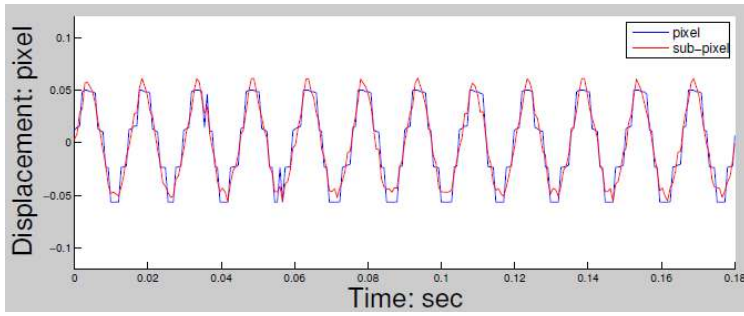(a) Displacement function of marker 6 in Figure 2(a)   (d) PSD from marker 6 (Figure 8(a))
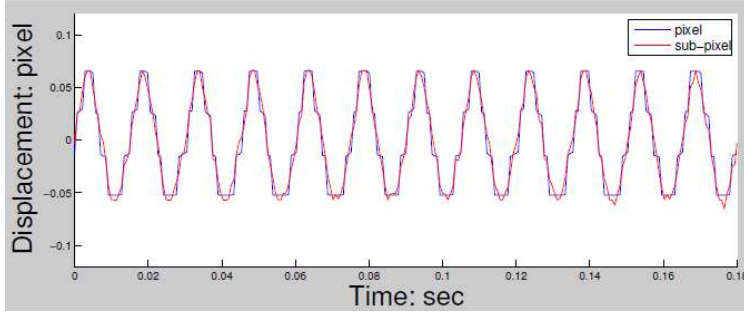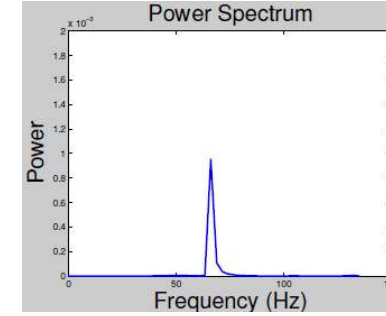
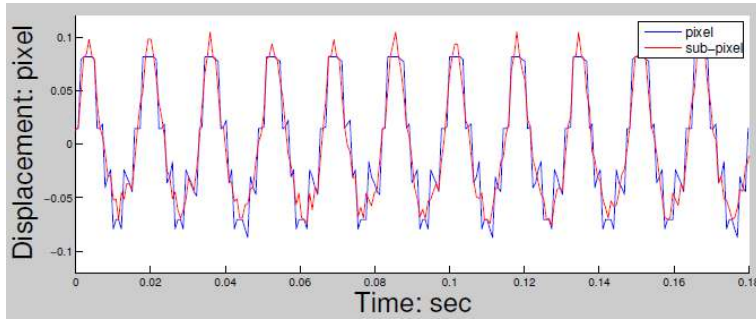(b) Displacement function of marker 16 in Figure 2(a)   (e) PSD from marker 16 (Figure 8(b))

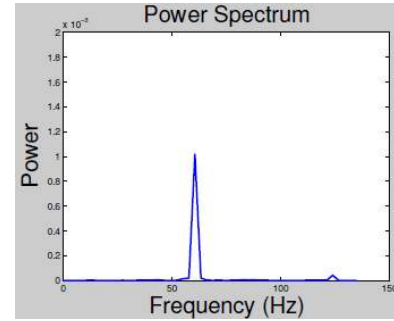(c) Displacement function of marker 26 in Figure 2(a)   (f) PSD from marker 26 (Figure 8(c))

Figure 8. Undamaged beam, mode 2 at 66.9Hz: displacement functions for three selected markers (Figures 8(a)–8(c)); x-axis: 0 to 0.18 secs in steps of 0.02 secs, y-axis: -0.1 to 0.1 in fractions of 0.05 of a pixel and the corresponding PSDs (Figures 8(d)–8(f)) generated from those displacement functions; x-axis: Frequency (Hz), range 0 to 150Hz in steps of 50Hz, y-axis: Power ($(abs(fftx)/fs)^2$), where $fftx$ is the FFT response and fs is the sampling rate, range 0 to 2×10^{-2} in steps of 0.2×10^{-2}

We see that the displacement functions from sub-pixel markers (red lines in Figure 8 and 9) offer a smoother interpolation of displacement signals. Furthermore, they also exhibit better consistency both locally and globally, while pixel accuracy markers yield signals with abrupt changes especially at peaks. This is particularly apparent in Figure 8(c) and Figure 9(a). Similarly, peaks of displacement signals (red) are better preserved using sub-pixel markers, whereas displacement functions of pixel accuracy markers tend to be flat at the extremes.

(a) Displacement function of marker 6 in Figure 2(a)

(d) PSD from marker 6 (Figure 9(a))
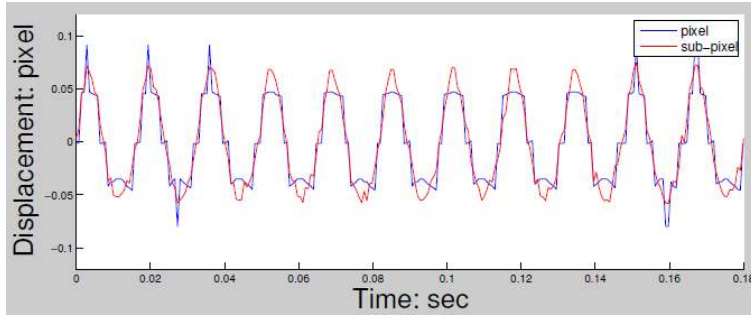
(b) Displacement function of marker 16 in Figure 2(a)
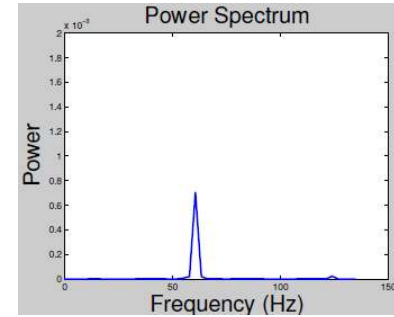
(e) PSD from marker 16 (Figure 9(b))
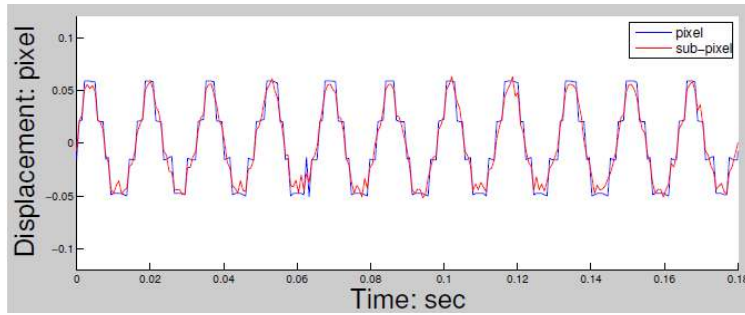
(c) Displacement function of marker 26 in Figure 2(a)

(f) PSD from marker 26 (Figure 9(c))

Figure 9. Beam with 5% damage, mode 2 at 66.9Hz: displacement functions for three selected markers (Figures 8(a)–8(c)); x-axis: 0 to 0.18 secs in steps of 0.02 secs, y-axis: -0.1 to 0.1 in steps of 0.05 pixel and the corresponding PSDs (Figures 8(d)–8(f)) generated from those displacement functions; x-axis: Frequency (Hz), range 0 to 150Hz in steps of 50Hz, y-axis: Power $((abs(fftx)/fs)^2)$, where $fftx$ is the FFT response and fs is the sampling rate, range 0 to 2×10^{-2} in steps of 0.2×10^{-2}
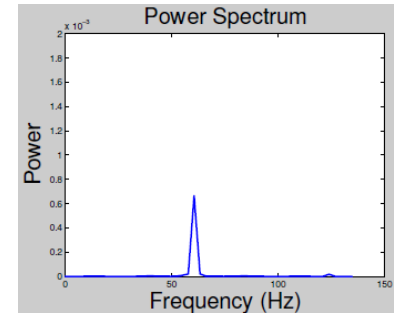
Table 2 compares modal frequencies extracted using standard and sub-pixel VVS, against actual frequencies used in the experiments. The frequencies are the averages of all markers along all three test beams vibrating under mode 2. It shows that modal frequencies from sub-pixel VVS (average error of 1.63%) offer a closer match the experimental data compared to standard VVS (average error of 3.07%). Moreover, it can also be seen from Table 2 that frequency shifts exist when comparing the undamaged beam to the damaged beams. Such shifts are caused by the different degrees of manufactured damage, which can change mass and local stiffness in the structure. It can be seen that frequency dropped 3.6Hz – a difference of about 5% – from 66.1Hz (undamaged) to 62.5Hz (50% damage). For the steel samples tested here is the change in mass was small ranging from 0.05%

for 5% damage up to a maximum of 0.5% for the 50% damage case. This suggests that change in stiffness is the main reason for the frequency change, as reported previously[4, 5] This also suggests that VVS would be applicable to a range of damage scenarios where the presence of damage changes the local stiffness. This has been demonstrated in the case of delamination damages in our parallel studies of composite laminate structures.[35] Using the acquired modal frequencies, we are able to detect the presence of damage in the structure, but further analysis is needed to locate them. In the next section, we will demonstrate how a wavelet analysis of the VVS data can be used to detect and locate damage of increasing degrees of severity.

Using the same specification computing platform as given at the end of section 3.2, the CPD algorithm tracks 45 markers at an average of 0.0559 sec per video frame, which is just below real-time.

4. WAVELET-BASED DAMAGE DETECTION AND LOCALISATION

We employ a wavelet-transform (WT) based algorithm to perform damage detection and localisation, using modal shapes acquired using sub-pixel VVS markers as an input. In contrast to Fourier analysis where time (space) information is lost, wavelets are located in time (space) and have "windows" of variable size. These distinct characteristics allow WT to reveal important information buried within signals that other techniques fail to detect.

We utilize both the discrete and continuous variants of WT: Continuous WT (CWT) offers accurate localisation of damage; whilst Discrete WT (DWT) delivers noise filtering, which is essential in experimental modal shape acquisition. A CWT of a function u(t) is defined as:

$$CWT(\tau, S) = \frac{1}{\sqrt{|S|}} \int u(t) \psi^* (\frac{t - \tau}{S}) \partial t \tag{9}$$

where τ is the translation parameter, S is the scale parameter, u(t) is the signal to be analysed (t being time/space parameter) and ψ^* is called the 'mother wavelet' — a family of source functions that satisfy the admissibility criterion and the daughter wavelets (those after convolving) are simply the translated and scaled versions of it.

The CWT provides a measure of similarity between the mother wavelet and the signal. The calculated wavelet coefficients reflect the strength of correlation between the signal and the mother wavelet at a given scale. Given a modal shape, **s**, reconstructed from the extracted markers, damage can be detected by treating **s** as a time signal and performing a CWT as stated in Equation 9. CWT is naturally sensitive to local discontinuities and smoothness of the input data and minimum noise are critical for the successful application of CWT, in order that the only discontinuity on the modal shape should correspond to the actual damage location. In experimental modal shape acquisition, measurement noise is inevitable, which can contaminate wavelet responses and makes damage assessment inconclusive.

To this end, we also evaluate modal shapes obtained from sub-pixel markers using DWT[6] prior to applying CWT. An important characteristic of DWT is that of multi-resolution analysis, where signals are broken down into detail and approximation components at each level of the dyadic space. Information that is often not apparent can be reviewed on different levels of a multi-resolution decomposition. DWT differs from CWT by digitizing the scale parameter, S and the translation parameter, τ from Equation 9, giving:

$$DWT(\tau, S) = \frac{1}{\sqrt{S}} \Psi(\frac{t - p\tau S}{S}), \tag{10}$$

where the following dyadic sampling is often performed:[36]

$$S = 2^o, \tau = 2^o p$$

in which $o, p \in Z^+$ with Z^+ being the set of all positive integers.

We used "db4" from the Debauchies family of wavelets to perform both DWT and CWT, a family of wavelets that was proven to work well on damage detection of cantilevered beams by Rucka et. al.[33] Specifically, we set $x(t) = \mathbf{s}$ and $S = 1 - 3$ in Equation 9 and damage is located where the CWT coefficient is at its global maximum.
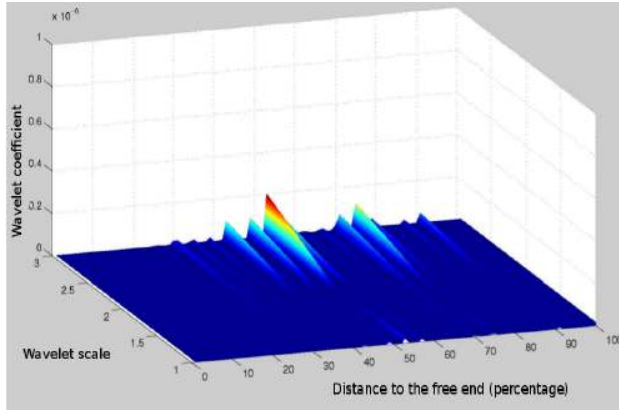
(a) VVS (pixel accuracy): 5% damage

(d) VVS (sub-pixel accuracy): 5% damage
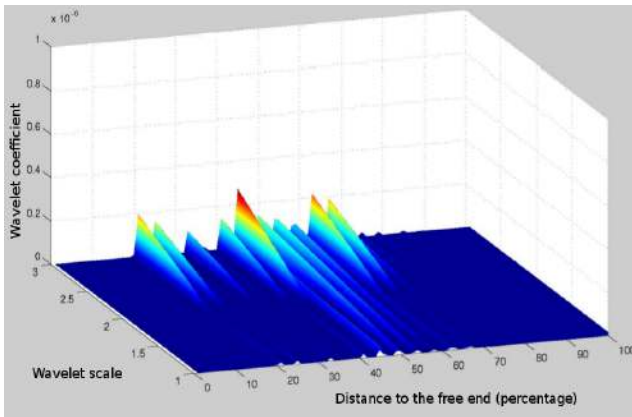
(b) VVS (pixel accuracy): 28% damage

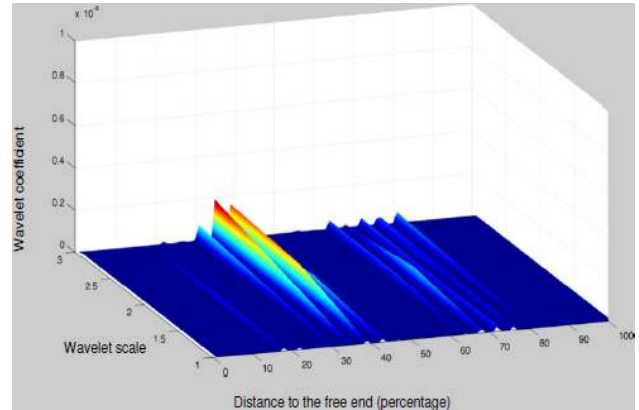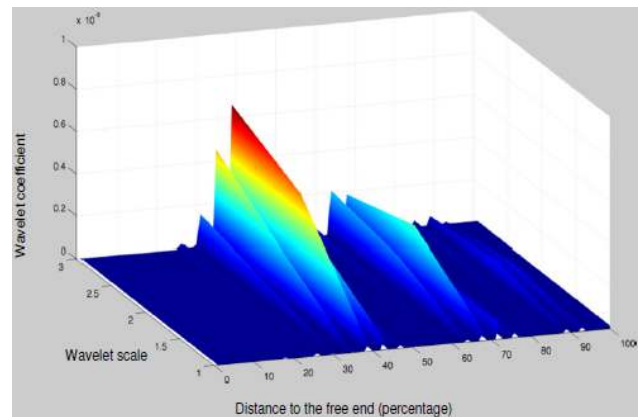(e) VVS (sub-pixel accuracy): 28% damage
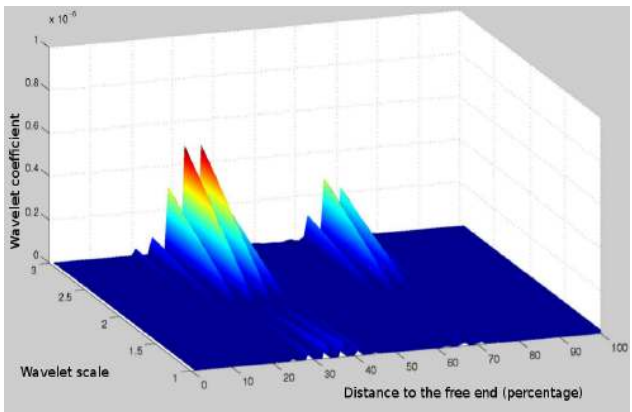
(c) VVS (pixel accuracy): 50% damage
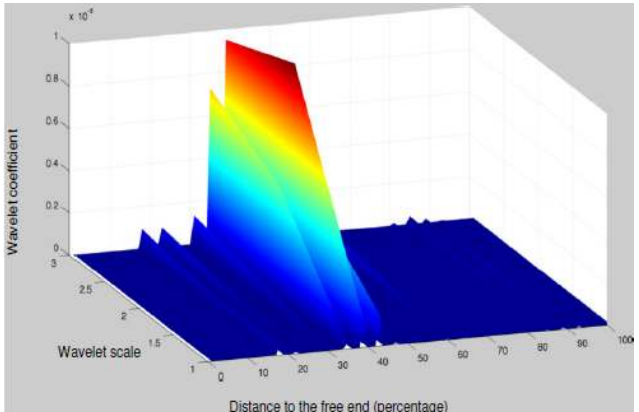
(f) VVS (sub-pixel accuracy): 50% damage

Figure 10. Comparison of damage detection results on three steel beams (Beam Length: 463mm, Mode Number: 2). Figures 10(a)–10(c): pixel accuracy VVS; Figures 10(d)–10(f): sub-pixel accuracy VVS; (top to bottom) damage levels of 5%, 28% and 50%. Note: in this and all damage detection results, the x-axis denotes the length of the beam (normalized by the total length), y-axis represents the wavelet scale and z-axis are the coefficient values

Using data from the previously described experimental setup (Section 3.1), we examined all three damaged beams using their second modal shapes. CWT coefficient plots of all three beams under mode 2 are shown in Figure 10, with standard VVS on the left and sub-pixel VVS on the right. The coefficients are plotted as height maps where larger values simply correspond to peaks in wavelet coefficient. The x-axis denotes the length of the beam (normalized by the total length), y-axis represents the wavelet scale and z-axis are the wavelet coefficient values. These plots further show the benefits deriving from the sub-pixel technique, because identification and localisation is enhanced at lower damage levels than with the standard technique. For the sub-pixel VVS plots one can clearly observe peaks in the vicinity of the machined notch located 1/3 the way along the beam from the free end for 50% and 28% damage, but the peaks are less significant and noisy for the standard VVS plots. The 5% damaged beam is less conclusive for both cases. However, the standard technique is successful at the highest level of damage, confirming the results presented by Rucka and Wilde.[37]

5. CONCLUSIONS AND FUTURE WORK

Using a new sub-pixel accuracy marker extraction algorithm and a combination of dynamic marker linking within a tracking-by-correspondence we have demonstrated the concept of Virtual Visual Sensor (VVS) as physical sensor substitutes. We show that through dynamic tracking of VVS, (i) PSDs can be computed to identify modal frequencies in the time domain (ii) modal shapes can be reconstructed via spline fitting of VVS in the spatial domain, and (iii) by the application of wavelet-based techniques can locate damage on notch depths of 28% and 50% in steel cantilever beams. Due to its non-invasive and highly portable nature, VVS has significant advantages for sensor installation and management: it does not suffer the spatial and mass constraints such as physical accelerometers and strain gauges. Moreover, with the increasing popularity of Wireless Visual Sensor Networks (WVSN), VVS could play an important part in practical applications.

As part of our plan for moving from the lab-based experiments towards practical applications, we have constructed a larger skeletal structure with the intention of placing it on an available multi-axis shaking table (MAST). This structure offers the opportunity to explore not only the effectiveness of the approach at a larger scale (structure is 2.5m high), but also the introduction of selective weakness, by slacking off the connecting bolts between various sections. The aim is to address various shortcomings of the current work in respect of: (i) investigating the feasibility of real-time VVS for SHM, which will broaden the scope for application in practical WVSN settings, (ii) investigation of the effect on performance of VVS of random and white-noise type vibration including the velocity and acceleration, (iii) handling camera node motion (currently the camera is static during data acquisition) by decoupling the camera from that of the monitored object, (iv) determining the feasibility of utilising lower spatial and temporal resolutions and hence the use of cheaper cameras.

ACKNOWLEDGMENTS

REFERENCES

[1] Carden, E. P. and Fanning, P., "Vibration based condition monitoring: A review," *Structural Health Monitoring* **3**, 355–377 (12 2004).

[2] Radzieński, M. and Krawczuk, M., "Experimental verification and comparison of mode shape-based damage detection methods," *Journal of Physics: Conference Series* **181**(1), 012067 (2009).

[3] Gillich, G.-R., Praisach, Z.-I., and Iavornic, C. M., "Reliable method to detect and assess damages in beams based on frequency changes," in [*Proceedings of the ASME 2012 International Design Engineering Technical Conferences, Computers and Information in Engineering Conference IDETC/CIE 2012*], 109–120 (August 2012).

[4] Doebling, S. W., Farrar, C. R., Prime, M. B., and Shevitz, D. W., "Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review," *Distribution* **213**(10), 1762–70 (1996).

[5] Fan, W. and Qiao, P., "Vibration-based damage identification methods: A review and comparative study," *Structural Health Monitoring* **10**(1), 83–111 (2011).

[6] Okafor, A. C. and Dutta, A., "Structural damage detection in beams by wavelet transforms," *Smart Materials and Structures* **9**(6), 906 (2000).

[7] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E., "Wireless sensor networks: a survey," *Computer Networks* **38**(4), 393 – 422 (2002).

[8] Akyildiz, I. F., Melodia, T., and Chowdhury, K. R., "A survey on wireless multimedia sensor networks," *Comput. Netw.* **51**(4), 921–960 (2007).

[9] Chintalapudi, K., Fu, T., Paek, J., Kothari, N., Rangwala, S., Caffrey, J., Govindan, R., Johnson, E., and Masri, S., "Monitoring civil structures with a wireless sensor network," *Internet Computing, IEEE* **10**, 26 – 34 (March 2006).

[10] Kim, S., Pakzad, S., Culler, D., Demmel, J., Fenves, G., Glaser, S., and Turon, M., "Health monitoring of civil infrastructures using wireless sensor networks," in [*IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks*], 254–263, ACM, New York, NY, USA (2007).

[11] Jang, S., Spencer, Jr., B., Rice, J., and Wang, Z., "Full-scale experimental validation of high-fidelity wireless measurement on a historic truss bridge," *Advances in Structural Engineering* **14**, 93–101 (02 2011).

[12] Sohrabi, K., Gao, J., Ailawadhi, V., and Pottie, G. J., "Protocols for self-organization of a wireless sensor network," *IEEE Personal Communications* **7**, 16–27 (2000).

[13] Anastasi, G., Conti, M., Di Francesco, M., and Passarella, A., "Energy conservation in wireless sensor networks: A survey," *Ad Hoc Netw.* **7**(3), 537–568 (2009).

[14] Porret, A.-S., Melly, T., Enz, C. C., and Vittoz, E. A., "A Low-Power Low-Voltage Transceiver Architecture Suitable for Wireless Distributed Sensors Network," in [*Proc. IEEE Int. Symp. Circuits Syst.*], **1**, 56–59 (2000).

[15] Ferentinos, K. P. and Tsiligiridis, T. A., "Adaptive design optimization of wireless sensor networks using genetic algorithms," *Comput. Netw.* **51**(4), 1031–1051 (2007).

[16] Werner-Allen, G., Lorincz, K., Welsh, M., Marcillo, O., Johnson, J., Ruiz, M., and Lees, J., "Deploying a wireless sensor network on an active volcano," *IEEE Internet Computing* **10**, 18–25 (March 2006).

[17] Milenkovic, A., Otto, C., and Jovanov, E., "Wireless sensor networks for personal health monitoring: Is- sues and an implementation," *Computer Communications* **29**(13-14), 2521–2533 (2006). Wirelsess Sensor Networks and Wired/Wireless Internet Communications.

[18] Kim, Y., Schmid, T., Charbiwala, Z. M., and Srivastava, M. B., "Viridiscope: design and implementation of a fine grained power monitoring system for homes," in [*Proceedings of the 11th international conference on Ubiquitous computing*], Ubicomp '09, 245–254, ACM, New York, NY, USA (2009).

[19] Quek, S. T., Wang, Q., Zhang, L., and Ong, K. H., "Practical issues in the detection of damage in beams using wavelets," *Smart Materials and Structures* **10**(5), 1009 (2001).

[20] Goferman, S., Zelnik-Manor, L., and Tal, A., "Context-aware saliency detection," in [*Computer Vision and Pattern Recognition – CVPR 2010*], 2376–2383, IEEE (2010).

[21] Song, Y.-Z., Arbelaez, P., Hall, P., Li, C., and Balikai, A., "Finding semantic structures in image hierarchies using laplacian graph energy," in [*Computer Vision – ECCV 2010*], Daniilidis, K., Maragos, P., and Paragios, N., eds., *Lecture Notes in Computer Science* **6314**, 694–707, Springer Berlin / Heidelberg (2010).

[22] Ta, D.-N., Chen, W.-C., Gelfand, N., and Pulli, K., "SURFTrac: Efficient tracking and continuous object recognition using local feature descriptors," in [*International Conference on Computer Vision and Pattern Recognition*], 2937–2944 (2009).

[23] Kho, J., Tran-Thanh, L., Rogers, A., and Jennings, N. R., "An agent-based distributed coordination mechanism for wireless visual sensor nodes using dynamic programming," *The Computer Journal* **53**(8), 1277–1290 (2010).

[24] Patsias, S. and Staszewskiy, W. J., "Damage Detection Using Optical Measurements and Wavelets," *Structural Health Monitoring* **1**(1), 5–22 (2002).

[25] Shi, J., Xu, X., Wang, J., and Li, G., "Beam damage detection using computer vision technology," *Nondestructive Testing and Evaluation* **9**(1), 1477–2671 (2009).

[26] Myronenko, A. and Song, X., "Point set registration: Coherent point drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**, 2262–2275 (2010).

[27] Vacchetti, L., Lepetit, V., and Fua, P., "Combining edge and texture information for real-time accurate 3d camera tracking," in [*Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*], ISMAR '04, 48–57, IEEE Computer Society, Washington, DC, USA (2004).

[28] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60**, 91–110 (2004).

[29] Duda, R. O. and Hart, P. E., "Use of the hough transformation to detect lines and curves in pictures," *Commun. ACM* **15**, 11–15 (January 1972).

[30] Fernandes, L. A. F. and Oliveira, M. M., "Real-time line detection through an improved hough transform voting scheme," *Pattern Recogn.* **41**, 299–314 (January 2008).

[31] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60**(2), 91–110 (2004).

[32] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum likelihood from incomplete data via the em algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B* **39**(1), 1–38 (1977).

[33] Rucka, M. and Wilde, K., "Application of continuous wavelet transform in vibration based damage detection method for beams and plates," *Journal of Sound and Vibration* **297**(3-5), 536–550 (2006).

[34] Sun, H., Mao, Y., Yang, N., and Zhu, D., "A real-time and robust multi-circle detection method based on randomized hough transform," in [*Computer Science and Information Processing (CSIP), 2012 International Conference on*], 175–180 (aug. 2012).

[35] Song, Y.-Z., Bowen, C., Kim, H. A., Nassehi, A., Padget, J., Gathercole, N., and Dent, A., "Non-invasive damage detection in composite beams using marker extraction and wavelets," in [*SPIE conference on Smart Structures and Materials, Non-Destructive Evaluation and Health Monitoring*], (2011). http://dx.doi.org/10.1117/12.880055.

[36] Mallat, S., [*A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way]*, Academic Press, 3rd ed. (2008).

[37] Rucka, M. and Wilde, K., "Crack identification using wavelets on experimental static deflection profiles," *Engineering Structures* **28**(2), 279 – 288 (2006).