Computers and Structures 200 (2018) 32–52

Contents lists available at ScienceDirect

Computers and Structures

journal homepage: www.elsevier.com/locate/compstruc

A computer vision approach for the load time history estimation of lively individuals and crowds

Ozan Celik, Chuan-Zhi Dong, F. Necati Catbas *

Dept. of Civil, Environmental and Construction Engineering, University of Central Florida, 12800 Pegasus Drive, Orlando, FL 32816-2450, USA

ARTICLE INFO

Article history:
Received 30 August 2017
Accepted 8 February 2018

Keywords:
Human structure interaction
Monitoring
Vibration serviceability
Jumping
Crowd loading
Load modeling
Stadium
Grandstand
Computer vision

ABSTRACT

A computer vision approach for measuring the load time history due to individuals and crowds jumping and bobbing is investigated. The method comprises of tracking the displacement trajectories of individuals and crowds using optical flow based algorithms followed by generating force time histories. Laboratory experiments, in which individuals and groups perform jumping at regular beats and songs on a force platform and on a grandstand simulator, are conducted. The estimated trajectories are compared directly with conventional sensors as well as indirectly with responses acquired from finite element models. The method is further validated via a field demonstration. Limitations of the method and future work for improvement are discussed. The proposed methods along with their applications on a real structure, and findings from a laboratory grandstand simulator that can accommodate experiments for groups of different sizes and structural configurations show great promise for computer vision based load modeling. In this sense, the study is taking an important step in support of creating a database for crowd loading that is needed as it is pointed out in the literature.

1. Introduction

1.1. Loading for human structure interaction

A widely accepted vibration serviceability criterion is to evaluate the system as a three-step framework namely excitation, path and response [14,17,32,33]. The excitation, which is the focus of this study, requires correct characterization of loads caused by humans. A number of research projects that have been done on human induced load measurement and modeling in civil engineering indicate that the estimation of jumping type loads (jumping/bouncing/bobbing/etc.) of both individuals and crowds are critical since these may create extreme vibration responses [25,50,54]. In case these loads are acted upon flexible structures such as stadiums, concert venues, dance floors, which have low frequency behavior by design requirements, vibration and displacement levels could be exacerbated to the point where catastrophic damages may occur [33]. On the other hand, when the motion of the crowd is in question, the severity of loading may vary significantly and may not be as strong depending on the level of synchronization between individuals.

There are several methods applied in the current practice of single and crowd load modeling. The first approach is adopting conservative mathematical functions which makes use of Fourier series. Early attempts of this kind are now known to be incorrect for they assume the loading being perfectly periodic and ignore the non-stationary variations. The second approach mildly remedies this problem by incorporating the probability of occurrence of jumping pulse variability and amplitude changes of real jumping records for the reconstruction of force records [21,35]. A similar method uses the probability of mean delay and phase scatter between pulses and recreates the jumping history with an auto-regression model [50]. However, measured and generated force recordings still do not closely match along the full frequency band due to the reason that cosine-squared functions that are utilized could only simulate smooth shaped jumping pulses. In the last decade, some promising research has been done to advance this approach with stochastic process based load generators in which sum of Gaussian functions are fitted to the measurements from real subjects to reconstruct accurate mathematical models [42–44,46].

Unfortunately, such applications are still limited to individual subjects and similar measurements or load generation models for crowds almost do not exist except a few [19,20]. Besides, whether the case is utilizing the reconstructed simulations or directly using the measured forcing functions, accurate measurements for both individuals and crowds especially when on flexible structures are

---

* Corresponding author.

E-mail addresses: ozancelik@knights.ucf.edu (O. Celik), ceczdong@knights.ucf.edu (C.-Z. Dong), catbas@ucf.edu (F.N. Catbas).

https://doi.org/10.1016/j.compstruc.2018.02.001
0045-7949/

must. The conventional apparatus for dynamic force measurements of singular subjects are force plates. However, use of force plates brings some concerns such as (1) having small dimensions ($\sim 0.6 \times 0.4$ m) that require controlled jumping which is quite tough when the subject is to jump at higher frequencies, therefore having distorted ground reaction force (GRF) patterns and (2) giving inaccurate results when mounted on a flexibly moving structure such as grandstand for the reason that additional inertial forces contribute to the measurements [41,45]. In search of solutions to such problems, some novel techniques are proposed.

A GRF estimation study, in which the body of a test subject is subdivided into fifteen major body segments and then instrumented with physical markers and tracked with video-based optoelectronic technology, shows [43,44] good alignment when compared with direct or indirect load measurements. The proposed method seems applicable in the field for sparsely distributed groups that remain in sight. As to the field measurements where exists a high density of crowd, this method does not seem to be feasible as the tracking device requires the subject to be visible at all times.

Some other alternatives exploit digital image processing and computer vision techniques. Several applications of computer vision based health monitoring studies may be found therein [36,37,55,56,57,58]. The earliest works of this kind are carried out mostly using a well-known family of algorithms called digital image correlation (DIC). One study includes tracking the prominent body parts of people by segmentation and looking at their correlation in consecutive images to estimate the loads applied to grandstands by large crowds [31]. Inspired by some early works on contour detection [5,6] and Bayesian clustering methods for crowd tracking [7], researchers use an easier method specifying each tracked segment with a rectangle for estimating jumping and bobbing pattern on a laboratory grandstand. Similarly, other trials focus on motion measurement of people and the patterns of their behavior in terms of velocity amplitude and frequency through either simulation in a computer controlled environment [10,12] or utilizing off the shelf regular or thermal imaging cameras on a portion of a real grandstand [9,18]. Particle image velocimetry (PIV) is used in these studies. The application includes tracking the displacement of the most similar regions in the consecutive images and eventually acquiring acceleration time histories [11,12]. Another study estimates the load generated by a crowd at a section of an actively monitored real-life grandstand [34]. The aim is to generate loading functions based on the developed acceleration time histories. Another vision based approach renders measurement of jumping and bobbing loads on a crowd and on the field by making use of DIC algorithms [40]. The efficiency of the proposed method is verified in the laboratory and at the field with a small group of people. The aforementioned methods require improvements for the fact that they assume the motion of the crowd in the same direction and may perform poorly under changing environmental conditions.

Apart from DIC based studies, couple of more advanced methods capture the displacement and acceleration information where a group of people in various sizes demonstrates jumping activities. The novelty of these studies lie within the adaptive nature of the algorithm to non-stationary changes as illumination changes, object deformation. The results are compared with data marker tracking system and wireless accelerometers that can be attached on the human body and on a flexible laboratory slab floor with indirect acceleration responses [59,60].

1.2. Objectives of the study

The objective of the study is to introduce alternative load time history measurement techniques for singular individuals and crowds using advanced computer vision algorithms along with their applications in both laboratory and real life. First, the algorithms are briefly explained and their applications are presented starting from singular individuals on a simple beam going up to groups in different sizes on a modular grandstand simulator that is instrumented with different types of conventional sensors. Then, the results of a sample field study at a stadium hosting a fairly large crowd is provided. The study makes a valuable contribution in terms of providing contactless vision based load measurement techniques that are verified with both direct and indirect comparisons between estimated results and conventional sensor measurements. Additionally, applying the method on a real crowd, building a simulator that can be adjusted to flexible and stiff configurations thereby accommodating experiments for groups of different sizes are other contributions. In this sense, the study is taking an important step in support of creating a database for crowd loading that is strongly needed as it is pointed out by other researchers dealing with the same problem.

2. Force measurement using computer vision

Utilization of computer vision allows the body motion of individuals or crowds to be measured with a camera from a long distance, without any contact and on any structure regardless of their flexibility. In case each body part is tracked individually, according to Newton's second law of motion the ground reaction force is estimated as follows:

$$F_{GR} = \sum_{i=1}^{n} m_i (a_i - g) \tag{1}$$

where m_i and a_i are mass and acceleration of the center of the mass of the i-th body segment, g is the standard gravity and n is the total number of body segments. The first step for force estimation is to find the displacement of the body. Once the displacement time history is known, acceleration record can be determined by taking either the second numerical derivative or the numerical gradient of the record. For the measurement of ground reaction forces, the mass is assumed to be concentrated at the head of a subject since this is a realistic assumption for the head is almost only visible part of the body regarding especially densely populated crowds. Besides, the motion of the head is more stable and it is easier to avoid abrupt mechanical motions when compared to other limbs of the body.

Optical flow is an image registration technique where the surface motion in three-dimensional environment is approximated onto approximate two-dimensional motion field by making use of spatiotemporal patterns of image intensity [2,4]. The optical flow measurement along with its improved different versions have been accepted as a reasonable and fairly accurate estimation of displacement and velocity and has had numerous applications such as recognition, tracking, motion modeling, segmentation, etc. in different engineering fields. According to various application scenarios and estimation strategies, optical flow can be divided into two as sparse flow and dense flow. Sparse flow is always combined with local features to do object tracking in the consecutive image sequence and the object is represented by selected features. Dense flow is used to calculated the flow vector at each pixel of the whole image. In the scope of this paper, two well-established variations proposed by Lucas and Kanade [39] and Farneback [23,24] for sparse and dense motion estimation respectively are applied.

2.1. Sparse flow

Determining optical flow requires the assumption of constant brightness in spatial and temporal space of consecutive frames by

$$f(x, y, t) = f(x + dx, y + dy, t + dt) \tag{2}$$