Image processing for capturing motions of crowd and its application to pedestrian-induced lateral vibration of a footbridge

Junji Yoshida^a, Yozo Fujino^b and Toshiyuki Sugiyama^c

^a*Department of Civil & Environmental Engineering, University of Yamanashi, Takeda 4-3-11, Kofu, Yamanashi, Japan*

E-mail: jyoshida@ccn.yamanashi.ac.jp

^b*Department of Civil Engineering, the University of Tokyo, Bunkyo 7-3-1, Tokyo, Japan*

E-mail: fujino@civil.t.u-tokyo.ac.jp

^c*Department of Civil & Environmental Engineering, University of Yamanashi, Takeda 4-3-11, Kofu, Yamanashi, Japan*

E-mail: t-sugiyama@yamanashi.ac.jp

Abstract. An image processing technique to capture motions of crowds is proposed and it is applied to understanding pedestrian-induced lateral vibration in a footbridge. Firstly, an outline of recording sequential images of vibration in the bridge is described and, then an image processing for human-head recognition from a single image of crowd is developed. In this method, conventional template matching techniques with human-head templates are extended by employing some selected templates, an updated search-algorithm and a classifier for clustering. Consequently, more than 50% of human-heads could be identified by the proposed method. Then, motions of detected human-heads, together with the bridge response, are tracked. Finally, interaction between the motions of pedestrians and the vibration of the bridge is discussed, with the emphasis on synchronization between the responses of the pedestrians and the bridge.

Keywords: Pedestrians, footbridge, lateral vibration, synchronization, image processing, template matching, pattern classification, human walking

1. Introduction

The vertical dynamic force by pedestrians is known to be around 2 Hz [2], and it sometimes causes resonance on footbridges, when a vertical vibration mode of footbridges possesses its natural frequency of 2 Hz. Hence, attention is paid to designing footbridges so that the natural frequencies of vertical vibration modes are not close to 2 Hz [13].

On the other hand, the horizontal dynamic force by pedestrians is relatively small [2,22] and, thus, it had been believed that significant vibrations of footbridges could not be induced by the horizontal force of pedestrians. However, it was reported that a large number of pedestrians induced lateral vibrations of footbridges with large amplitudes [3,10]. In the study by Fujino et al. [10], they recorded the lateral vibration of a footbridge by an analog video camera and tracked the motions of some pedestrians. Their results indicate that motions of the pedestrians are synchronized with the response of the footbridge. They also found from the experiments that humans on a shaking floor gradually changed their way of walking according to the motion of the floor [17].

Fig. 1. Outside appearance of T-Park Bridge.

Recently, the same type of pedestrian-induced lateral vibration in footbridges occurred in the Solferino-Bridge in Paris [7] and the Millennium-Bridge in London, and both bridges were closed for the safety of pedestrians. In the London Millennium-Bridge, the amplitude of the lateral vibration reached around 70 mm [26] and, after investigations for more than one year, passive dampers were installed and, consequently, the vibration was significantly reduced [5, 6,9]. It was also reported in a newspaper that the lateral vibration occurred in Brooklyn-Bridge in 2003, when a blackout led to a large number of people to rush up to the bridge. However, the interaction of pedestrians and vibrations of those bridges are so complex that the mechanisms of the human-induced lateral vibration in footbridges have not been well understood [26].

Digital image processing has been remarkably developed and, due to the development of digital hardware, it could be applied to practical engineering problems, such as measuring flows of fluids (e.g. [16,20]) or deformation of solids (e.g. [1,25]). Other typical applications are recognizing objects (e.g. [8]) and capturing motions of hands, feet and eyes (e.g. [11,12]). Detecting humans is one of the important applications, but its purpose has been limited to recognizing several humans [4,19].

In this study, an image processing technique to capture motions of a large crowd is developed. Then, it is applied to sequential images, which record motions of a large number of pedestrians on an actual footbridge in Japan during a case of the lateral vibration. Finally, the relationship between the motions of pedestrians and the vibration of the footbridge is evaluated with the aid of the responses computed by the image processing.

2. Measurement of pedestrians with CCD cameras in T-Park Bridge

T-Park Bridge is a two-span cable-stayed bridge with a steel box-type girder (Fig. 1), and is located at the suburbs of Tokyo in Japan. It has 180 m length and 5.25 m width. It connects a boat race stadium and a bus terminal. When a big boat race is finished, a large number of people pass on the bridge deck; at peak, about two thousand pedestrians walk on the bridge deck and this evokes the lateral vibration.

In order to capture the motions of the crowd and the vibration, CCD cameras of 1.3 million pixels and of 30 frame/sec were set up at the roof of the stadium as shown in Fig. 2. It is noted that images taken by the camera mainly include the backs of the heads of pedestrians due to the restriction of the location of the camera.

Figure 3 is an example of the recorded sequential images, which are 8 bit depth of gray scale image. The figure also indicates two regions; one includes pedestrians and is going to be used in detecting human-heads, and another is a part of the bridge deck and is going to be used in tracking bridge motion in the image processing.

Fig. 2. Measurement of sequential images with a CCD camera in T-Park Bridge.

Fig. 3. An example of sequential images recorded in T-Park Bridge and two regions for image processing.

The significant lateral vibration of the bridge usually continues for around 15 minutes and the total number of recorded images during the vibration is about 27,000 ($= 15 \times 60 \text{ sec} \times 30 \text{ frame/sec}$). These image data are directly transferred to a computer, which possesses RAID hard disk system for real-time recording of all images without any compression.

3. Image processing to detect human-heads from images of crowd

3.1. Application of the template matching with simple GA

In the image processing, template matching methods have been applied to problems to detect objects from an image. In these methods, the similarity of brightness patterns between a template and a part of the image is evaluated by the correlation coefficient between them. In practical applications, the template itself is not included in the image, but its similar patterns are involved with changing its position, angle or size. Since parameters specifying these changes are unknown in advance, an optimization algorithm is employed in order to find them [15]. Figure 4 shows an example of the template matching to detect human faces with one template, where parameters to be identified are

(a) Template

(b) Target image

(c) Result of detection

Fig. 4. A typical template matching for detecting human faces from an image.

positions in 2D, a rotational angle and an expansion (or shrinkage) ratio of detected objects relative to the template. Nagao et al. [14] proposed a template matching method with the aid of GA. In this method, human faces were detected from an image with several templates and good results were obtained.

In the direct application of the method by Nagao to our study, ten images of typical human-heads are selected from the recorded images, and they are employed as templates of the matching. However, the number of detections is less than one-third of all pedestrians among the image (see the top bar in Fig. 5).

3.2. Improvement of template matching techniques

3.2.1. Pre-selection of templates

In the application of the template matching technique, the selection of templates is one of the most important factors in obtaining a high success rate in detection procedures. In this study, fifty typical human-heads are randomly extracted from the recorded images, as the candidates of templates. Then, pre-template-matching with each candidate and a few recorded images is executed, and the performance of detection by all the candidates is evaluated by success rates: $\lambda \equiv N_C/N_T$, where N_T and N_C are the number of all detections and the number of detected human-heads by a candidate, respectively. Finally, ten candidates with higher success rates are employed as templates.

It is noted that the use of more than ten templates needs more computational time, but gives almost the same success rate.

Fig. 5. Success rates of detecting human heads by the template matching methods with random templates & simple GA, the selected templates & simple GA and the selected templates & read-coded GA.

Fig. 6. Symmetricity of brightness pattern around vertical line of a human head.

3.2.2. Application of real-coded GA

Real-coded GAs are genetic algorithms using real values as genes and they have been proved to possess a high performance in complex parameter identification problems, compared with GAs using binary code [18]. In this study, a real-coded GA with simplex crossover [24], together with minimal generation gap model [21], is employed for parameter identification in the template matching.

Figure 4 shows comparisons of success rates in detecting human heads from an image by the template matching methods using random templates with simple GA, the selected templates with simple GA and the selected templates with the real-coded GA. From the figure, it is found that the use of the selected templates and employment of the real-coded GA increase the success rate by twice and by around 10%, respectively.

3.3. Removal of miss-detections by classifier

3.3.1. Maximum likelihood classifier

The improvements of the template matching increase not only the success rate, but also the number of miss-detections; objects such as backs of humans and parts of the bridge. To remove these miss-detections, a classification with maximum likelihood classifier is employed after the detection. The classification has been widely used in object recognition of image processing and known to be efficient [8,23]. Accuracy of the classification largely depends on characteristics of features, and employment of an even smaller number of features enables accurate classification, when they possess properties to separate classes well.

Fig. 7. A result of applying the frame difference method to a image of a human head.

(a) An example of classification (human heads: squares with net, others: white squares)

(b) Accuracy of classification with different number of features

Fig. 8. Accuracy of classification by the maximum likelihood classifier.

In this study, two classes are defined; one is a class of human heads, and the other is a class of all the miss-detections, and the following five features are employed.

Feature A: The correlation coefficient in the detection by the template matching method is chosen as a feature. The feature is expected to be a measure of similarity between the detected object and the class of human heads.

Feature B: The detected human heads are almost symmetric on the vertical center line as shown in Fig. 6. Hence, a feature is defined as the sum of the difference of brightness between the two pixels, which are located at symmetric to the line.

Feature C: Human heads possess high variations of brightness-patterns in comparison with that of the neck,

Fig. 9. Time-history of the number of detected human heads.

Fig. 10. An example of human walking paths obtained by tracking human-heads.

shoulder and background. Hence, the difference of the maximum and minimum brightness in the region is chosen as a feature.

Feature *D*: Humans-heads change their positions in the adjacent images, and they could be recognized as dynamic objects, when the frame difference method with three sequential images [23] is applied. Hence, a ratio between the number of pixels of the dynamic object and the total number of all the pixels is employed as a feature. The purpose of the feature is to exclude static objects. As an example, a human-head after applying the frame difference method is presented in Fig. 7, where dynamic objects are drawn in gray levels, while static objects are in white.

Feature *E*: This feature is defined as the sum of the difference of brightness between the two pixels, which are located symmetrically to the vertical line (same as in feature *B*), after applying the frame difference method. This is employed for picking up symmetric and dynamic objects.

(a) Time-history of average amplitude

(b) An example of displacement response during large vibration

Fig. 11. Lateral vibration of the bridge deck obtained by the image processing.

3.3.2. Results of classification

The maximum likelihood classifier with the five features is applied to removing the miss-detections by the template matching methods. In the classification, one hundred training data of the two classes are selected randomly from the recorded images. Figure 8 shows results of the classification by employing the features A , B , the features A , B , C , D and all the features. The figure shows that the results of the classification become accurate as the increase of the number of features, and employment of all the features gives more than 90% accuracy in classification. After the classification, the success rate of detecting human-heads is around 60%.

4. Motions of pedestrians and vibration of T-park Bridge

4.1. Tracking by the template matching method

Firstly, the recorded sequential images are divided into segments so that each segment includes 300 images (images for 10 sec), where 150 images are overlapped with the adjacent segments. Then, Motions of pedestrians are tracked by applying the same template matching method to the sequential images in each segment.

In the tracking process, human-heads are detected in the first image of the segment, and they are used as templates. Then, their similar brightness patterns are tracked from the following sequential images. The centroids of the detected regions by the tracking are considered to be the positions of pedestrians in each image.

Fig. 12. Time-history of the synchronized ratio of pedestrians during the large lateral vibration in T-Park Bridge.

It is noted that the size of detected human-heads becomes smaller in images as the time passes, since almost all the pedestrians move away from the stadium. Therefore, continuous tracking of the same human-heads becomes more difficult and, in this case, computation with 10 seconds images is the limit of correct tracking.

The time-history of the number of detected human-heads by the tracking method is shown in Fig. 9.

Figure 10 presents an example of walking paths of pedestrians for 10 sec and it is found from the figure that heads of pedestrians vibrate harmonically in the horizontal direction with a constant frequency.

The motion of the bridge is also tracked by the same template matching method. In this case, a part of the bridge deck at the basement of an illumination pole (see Fig. 3) is selected as the template. Figures 11(a) and (b) show the time-history of the average amplitude and the lateral displacement response of the bridge deck in a period subjected to large vibration, respectively. From the figures, it is found that the displacement of the bridge deck sometimes exceeds 10 mm in a period of 200–500 sec and it has gradually becomes small after 500 sec.

4.2. Relation between the motions of pedestrians and the bridge

The dominant frequency of horizontal motion of a pedestrian in each segment is identified from the peak of Fourier amplitude spectra of its response. The dominant frequency of the bridge response is also identified by the same process in the corresponding segment. In these identifications, 0.1 Hz is the step size in the frequency domain at this case. Then, pedestrians whose dominant frequencies are within ± 0.1 Hz in comparison with that of the bridge are judged as *synchronized*, and a ratio for synchronization of pedestrians is defined as $\alpha = n_L/n_T$, where n_L is the number of synchronized pedestrians and n_T is the total number of pedestrians detected by the image processing.

The time-history of the synchronized ratio is shown in Fig. 12. The figure indicates that the synchronized ratio is an average of 60% in a period of 100–600 sec in the figure, where the bridge deck is subjected to large lateral vibration as shown in Fig. 11.

5. Conclusions

This study proposes an image processing technique to detect human-heads from a crowd and describes its application to the analysis of pedestrian-induced lateral vibration of a footbridge. The proposed image processing technique is a direct extension of a template matching method and it can detect more than 50% of human-heads from images of a crowd. Then, motions of pedestrians and vibration of the footbridge are tracked and they are compared in frequency domain. The results indicate that around 60% of pedestrians on the bridge deck are synchronized to the response of the bridge.

Acknowledgements

The authors are thankful to the Institute of T-Motorboat Race Track for allowing us to measure the vibration in T-Park Bridge. This research was partially supported by the Maeda Engineering Foundation, and Television Yamanashi Co., Ltd. Foundation for Science Promotion.

References

- [1] K.A. Alshibli and S. Sture, Sand shear band thickness measurement by digital imaging techniques, *Journal of Computing in Civil Engineering* **13**(2) (1999), 103–109.
- [2] H. Bachmann and W. Ammann, Vibrations in Structures Induced by Man and Machines, *IABSE-AIIPC-IVBH*, Structural Engineering Documents, 3e, 1987.
- [3] H. Bachmann, Case studies of structures with man-induced vibrations, *Journal of Structural Engineering, ASCE* **118**(3) (1992), 631–647.
- [4] A.C. Davies, J.H. Yin and S.A. Velastin, Crowd monitoring using image processing, *Electronics & Communication Engineering Journal* **14**(7) (1995), 37–47.
- [5] P. Dallard, A.J. Fitzpatrick, A. Flint, S.L. Bourva, A. Low, R.M.R. Smith and M. Willford, The london millennium footbridge, *The Structural Engineer* **79**(22) (2001), 17–33.
- [6] P. Dallard, A.J. Fitzpatrick, A. Flint, A. Low, R.R. Smith, M. Willford and M. Roche, London millennium bridge: pedestrian-induced lateral vibration, *Journal of Bridge Engineering, ASCE* **6**(6) (2001), 412–417.
- [7] F. Danbon and G. Grillaud, *Characterisation and Modelling of the Dynamic Loading Induced by a Moving Crowd on the Solferino Footbridge in Paris*, Proceedings of the 2nd International Conference on the Design and Dynamic Behaviour on Footbridges, Venice, Italy, 2005.
- [8] R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2002.
- [9] A.J. Fitzpatrick, P. Dallard, S.L. Bourva, A. Low, R.M.R. Smith and M. Willford, *Linking London – The Millennium Bridge*, The Royal Academy of Engineering, 2002.
- [10] Y. Fujino, B.M. Pacheco, S. Nakamura and P. Warnitchai, Synchronization of human walking observed during lateral vibration of a congested pedestrian bridge, *Earthquake Engineering and Structural Dynamics* **22** (1993), 741–785.
- [11] Y. Guo, G. Xu and S. Tsuji, Tracking human body motion based on a stick figure model, *Journal of Visual Communication and Image Representation* **7**(1) (1994), 1–9.
- [12] K.M. Lam and H. Yan, Locating and extracting the eye in human face images, *Pattern Recognition* **5** (1996), 771–779.
- [13] Y. Matsumoto, T. Nishioka, H. Shiojiri and K. Matsuzaki, Dynamic design of footbridges, *IABSE Proceedings*, No. P-17/78, 1978, 1–15.
- [14] T. Nagao, T. Agui and H. Nagahashi, *Extraction of Two-Dimensional Arbitrary Shapes using a Genetic Algorithm*, SPIE Visual Communications and Image Processing '93 SPIE, 1993, No. 2094.
- [15] T. Nagao, *Evolutionary Image Processing*, Shoukou-do, 2002. (in Japanese).
- [16] K. Nishino, N. Kasagi and M. Hirata, Three-dimensional particle tracking velocimetry based on automated digital image processing, *Journal of Fluids Engineering, ASME* **111** (1989), 384–391.
- [17] S. Okamoto, M. Abe, Y. Fujino and Y. Nakano, Characteristics of human walking on a laterally shaking floor, *Journal of Earthquake Engineering and Structural Mechanics, JSCE*, No.441/I-18 (1991), 177–184, (in Japanese).
- [18] I. Ono and S. Kobayashi, *A Real-coded Genetic Algorithm for Function Optimization Using Unimodal Normal Distribution Crossover*, Proceedings of 7th International Conference on Genetic Algorithms, 1997, 246–253.
- [19] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna and T. Poggio, Pedestrian detection using wavelet templates, *Proceedings of 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1997), 193–199.
- [20] M. Raffel, C. Willert and M. Kompenhans, *Particle Image Velocimetry*, Springer verlag, Berlin, 1998.
- [21] H. Satoh, M. Yamamura and S. Kobayashi, *Minimal Generation Gap Model for GAs Considering Both Exploration and Exploitation*, Proceedings of IIZUKA'96, 1996, 494–497.
- [22] M. Satoh and Y. Mataka, Experimental investigation of forces by walk of usual humans with a large floor-type load cell, *Journal of JSME (C)* **55**(518) (1989), 2672–2675, (in Japanese).
- [23] M. Sonka, V. Hlavac and R. Boyle, *Image Processing, Analysis and Machine Vision*, (2nd ed.), ITP, California, 1998.
- [24] S. Tsutsui, M. Yamamura and T. Higuchi, *Multi-Parent Recombination with Simplex Crossover in Real-Coded Genetic Algorithms*, (Vol. 1), Proceedings of Genetic and Evolutionary Computation Conference, 1999, 657–664.
- [25] J. Yoshida, M. Abe, L.C. Sujeewa and Y. Fujino, Measurement method for continua by image processing, *Journal of Structural Engineering, ASCE* **130**(8) (2004), 1145–1156.
- [26] S. Zivanovic, A. Pavic and P. Reynolds, Vibration serviceability of footbridges under human-induced excitation: A literature review, *Journal of Sound and Vibration* **279**(1–2) (2005), 1–74.