Grand Challenges of Machine-Vision Technology in Civil Structural Health Monitoring

Yunchao Tang[1*], Yunfan Lin[1], Xueyu Huang[1], Minghui Yao[1], Zhaofeng Huang[2], Xiangjun Zou[2]
[1]College of Urban and Rural Construction, Zhongkai University of Agriculture and Engineering, Guangzhou, China
[2]College of Engineering, South China Agricultural University, Guangzhou, China
E-mail: ryan.twain@gmail.com

Abstract: Machine-vision technology has progressed remarkably in both accuracy and speed owing to advances in computer technology and artificial intelligence. In this paper, state-of-the-art research on vision-based techniques is reviewed for civil infrastructure condition assessment. The major challenges of machine vision technique in civil structural health monitoring are presented.
Keywords: artificial intelligence, smart sensors, vision technology, deep learning, structural health monitoring

1. Introduction

Civil infrastructures, including bridges, highways, dams, and reservoirs, are critical engineering facilities related to the national economy and people's livelihood[1-3]. Most of them were erected many decades ago and still in service[4-6]. For instance, China has over 98,000 reservoirs, with 36% of them in dangerous condition due to inadequate surveys, poor construction quality, and improper protection[7]. The United States has 607,380 bridges over 50 years old, 25% of them being defective or abandoned, and over 56,000 of bridges structurally defective[8]. Evaluation of the condition of civil infrastructures have been examined via the use of monitoring information[9].

Currently, engineers use smart sensors and multifunctional sensors to evaluate infrastructure status, combined with relevant decision-making national regulations or standards[10]. However, this technique is laborious, environmentally susceptible, costly, and time-consuming[11-15]. Machine-vision technology has been deemed a core and promising component of structural health monitoring in the field of civil engineering[11, 15]. Also, it is the most closely connected and extensive frontier in civil engineering that has been incorporated with artificial intelligence (AI)[16-18]. At present, AI applications in civil engineering typically involve intelligent city planning, three-dimensional (3D) printing, building information modelling, and structural health monitoring[4, 6, 16, 19-22].

Machine-vision systems are devices that automatically collect and intelligently process images of a real object or scene through photoelectric devices or sensors, fully useful information, or equipment used to control machine motion[23]. The three major functional modules of a machine-vision system are image acquisition, image processing, and output and communication[24-26]. The advantages of machine-vision systems are noncontact, real-time detection, long-term stability, high accuracy, remote monitoring capability, high environmental adaptability, controllable image-taking speed, and data archiving applications[27-29]. Stereo vision technology is the core of machine vision technology. The true 3D position of the target is obtained through the principle of optical geometry, such that the target's geometric shape can be directly sensed, key geometric parameters measured, and the damage degree evaluated[30-35]. Therefore, stereo vision technology plays an important role in structural health monitoring.

Given all these advantages, what does the future hold for machine vision-based structural health monitoring? Recently, we conducted a webinar on major unsolved challenges in vision-based structural health monitoring (Fig. 1). These included the major challenges that might require significant breakthroughs or research in the coming decade. Then, we discussed the current state-of-the-art of stereo-vision, target area extraction, and 3D point-cloud processing technology.

DOI: https://doi.org/10.37256/aie.112020250

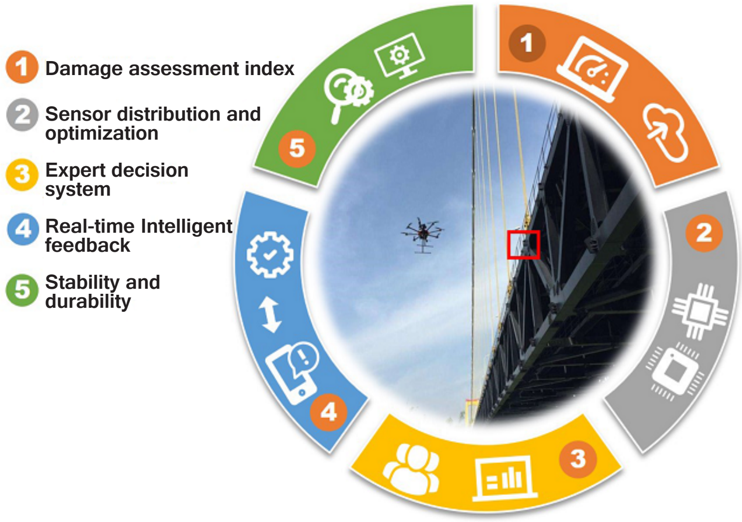

Figure 1. Grand challenges of machine vision systems in civil structural health monitoring

2. Current state-of-the-art of stereo-vision technology

Stereoscopic-vision technology provides 3D stereoscopic information of the target using a projection of the target's two-dimensional (2D) image information, thus providing strong technical support for structural health monitoring. It can be divided into active and passive vision, depending on different perception methods. The 3D reconstruction of active vision needs to obtain depth information with the help of projection equipment, primarily including structured light and time-of-flight methods. The 3D reconstruction of passive vision requires 3D information of the target through the reflection of external light sources. This can be divided into monocular vision, binocular stereovision, and multi-eye stereovision depending on the number of cameras. The imaging principle of the binocular stereoscopic-vision method is similar to that of the human vision system. With stable reconstruction effects and wide application scope, reconstruction of stereoscopic information based on triangulation measurement is a highlight in research of 3D reconstruction technology[36]. Dong *et al*. have detected reinforced concrete beams using multiple stereoscopic-vision systems, correlated the vision system coordinates through a data-fusion algorithm, and finally reconstructed their 3D coordinates. This global measurement method is more efficient and more accurate than previous methods[37]. Malowany *et al*. have also proposed the use of multiple stereoscopic-vision systems for achieving the measurement of large buildings and the use of laser trackers for establishing the relationship between the vision systems, thus achieving the measurement of global deformation. The method is stable and reliable and has application prospects for 3D deformation measurement of large and complex engineering objects from multiple directions[38]. Tang *et al*. have used binocular stereoscopic vision to measure the noncontact deformation of concrete-filled steel tubular columns, used image processing technology and stereoscopic matching to obtain a 3D point cloud, and then extracted the point-cloud section and obtained the maximum deformation by fitting calculations. This method can reflect the deformation process of concrete-filled steel tube columns\ and the measurement results meet the accuracy requirements of engineering measurement (ISO IT7)[39]. Ferrer *et al*. have measured deformation using textural information of the concrete surface, first dividing the surface into small rectangular areas, determining the size of local areas by local entropy, and then determining deformation using image correlations. This method overcomes the limitations of previous methods, which require spraying signs and is more suitable for practical applications[40]. Tang *et al*. have constructed a four-ocular vision system using two sets of binocular vision, which perform real-time detection of concrete-filled steel tube columns for 3D deformation detection under dynamic loading, effectively reproducing the target's 3D deformation process. The mathematical model combining the four-ocular vision coordinate system and point cloud matching proposed by this method expands the visual field of vision measurement and the method has real-time detection capabilities and good application prospects[41].

In the above-mentioned stereoscopic-vision technology, based on optical geometry, the key content is the post-processing and reconstruction of the 3D point cloud of the structural target. A point cloud, as an unordered set of 3D points in space, generally only contains coordinate information. This leads to the lack of support from key features when performing high-level operations on a point cloud. To this end, scholars have proposed a series of algorithms for estimating the features of structural targets. Huang *et al*. were the first to generate a uniform particle set on the original point cloud using the weighted, local, optimal projection operator, estimated the local normal of particle sets, and then introduced an iterative framework to determine the normal propagation scheme for flipping the normal of the combined particle set and original point cloud. This method has good applicability in the integration of original point clouds[42].

In recent years, with the progress of deep-learning research, robust and efficient deep-neural networks have made remarkable achievements in the field of stereoscopic vision. Saurabh Gupta *et al*. have used RGB-D depth cameras to obtain pictures and used a convolutional neural network (CNN) to extract features of RGB color information and depth information to achieve instance segmentation[43]. Due to the introduction of deep information to the network, compared with the contemporary RCNN (Region-CNN), research further elevated the success rate of target detection[44]. Žbontar has proposed that the use of an MC-CNN network, with left and right image segments and standard disparity maps as the input training networks, in combination with a variety of post-processing methods for achieving high-quality disparity maps. Compared with existing traditional stereoscopic matching algorithms, this performs better in both accuracy and speed[45]. Qi *et al*. have proposed the PointNet network to achieve the semantic segmentation of 3D objects[46]. Unlike existing research, this network directly takes the point cloud as input, fully retains the detailed information of the point cloud, and remains constant for different input orders. Qi *et al*. have further developed PointNet++, improved its ability to obtain the underlying local structure near the point and achieved the best test performance in the 3D point-cloud analysis benchmark test at that time [47]. Zhou *et al*. have proposed the VoxelNet network, which combines the advantages of voxel and point-cloud networks[48]. Its performance in the automotive detection benchmark (Karlsruhe Institute of Technology) exceeds the most advanced 3D detection method based on laser radar. Kim *et al*. have combined a mask and Mask RCNN to detect cracks and then used morphological processing to quantify cracks, exhibiting better detection effects for small cracks[49]. Mundt *et al*. have established a public-concrete damage image data set to classify common types of concrete damage with multiple targets[50]. Kim *et al*. have used CNN to process detection images of concrete cracks to further distinguish between noisy point patterns of concrete cracks and similar cracks that are difficult to identify with existing image-processing algorithms[51]. Islam *et al*. have built a full CNN with an encoder and decoder framework for semantic segmentation of cracks, with both the F1 value and the recall rate up to 92%[52].

In general, the traditional stereo vision method based on optical geometry theory has gradually matured, which guarantees the effective application of the stereo vision method in the field of structural health monitoring. On the other hand, deep learning theory yields more possibilities in the development of the stereo vision field. Statistical methods based on big data complement the previous methods based on optical geometry. However, unlike 2D image processing, current deep networks have not yet become mainstream 3D vision methods. The main reason for this is that the application-level 3D visual deep network has deeper layers and more complex logical structures. Compared with 2D tasks, the accuracy of networks in 3D tasks is often low, especially in the most difficult 3D matching links, and deep networks have still not become the dominant method. In addition, the difficulty of acquiring and processing the 3D data set is much greater than that of the 2D image data set. As the requirements of target monitoring tasks gradually increase, we need to find a model that enables the optical geometry method to cooperate with deep networks efficiently, so as to obtain a more intelligent, stable, and accurate stereo vision system.

3. Research status of target area extraction

The extraction of target areas is a key step in visual monitoring. By detecting and segmenting the target position in the image, removing uninteresting targets, and improving the calculation accuracy and efficiency, has been the focus of research in the field of visual monitoring. Traditional target region-extraction algorithms have been proposed since the 1970s, mainly including threshold, regional, and edge detection methods[53-56]. Traditional algorithms rely on information, such as the color, shape, and texture of the target, for detection and are considerably subject to light and background. As a result, its extraction effectiveness in complex scenes is not ideal. Constrained by the limitations of traditional algorithms, scholars have begun to seek new theoretical breakthroughs. After 2000, graphic extraction and clustering-based extraction and segmentation algorithms have emerged. The stability and accuracy of such segmentation methods have been greatly improved, compared with that of traditional algorithms. Methods based on graph theory map the image into an undirected graph with weights and convert the problem of segmenting the target area into a problem of graph partition. Typical methods include the normalized cuts[57], graph cuts[58], and superpixel lattice algorithms[59]. Cluster-based methods use mutual information between pixels to aggregate neighboring pixels with similar features into a superpixel block to complete segmentation. Typical algorithms include the Medoidshift[60], Turbopixels[61], and SLIC algorithms[62].

The above target extraction methods all belong to unsupervised learning algorithms, which require neither labelling of the pictures nor consumption of time and effort for training. However, the accuracy of target extraction is expected to be further improved. In contrast, the CNN target extraction algorithm, which belongs to the category of supervised learning, requires mathematical models built in advance, via training of manually labelled data sets to correctly extract the target but which greatly surpasses other methods in accuracy.

At the end of the last century, LeNet proposed one of the earliest CNNs to achieve accurate recognition of handwritten characters with a recognition rate of up to 98%, laying the foundation for the rapid development of CNN[63]. Affected by computer computing capabilities and other factors, CNN development has remained slow for more than ten years. Krizhevsky *et al*. have proposed AlexNet, which won the ImageNet competition for its great advantages and demonstrated the effectiveness of large-scale CNN[64]. The proposal of the model was of great significance for CNN development, which has made CNN one of the most popular research topics at present. With the birth of more classification CNN models, scholars have also proposed many excellent network models for the task of image target-region extraction.

In terms of target detection, Girshick *et al*. have proposed the first deep-learning model RCNN for target detection[44]. This model first extracts a large number of candidate regional frames from images through a selective search algorithm, then scales all regional frames, and extracts the features one by one via AlexNet. Next, it uses the support vector machine to classify the features present and performs border regression to yield the position of the predicted frame. The average detection accuracy of RCNN in the VOC data set greatly surpasses that of other algorithms. However, RCNN has the problems of requiring a high amount of repeated calculations, susceptibility to loss of feature information, and complex training steps, such that it is not an end-to-end network model. He *et al*. have proposed SPP-Net, which solves the problems of large calculation volume and susceptibility to loss of feature information in RCNN through candidate region frame mapping and spatial pyramid pooling[65]. Ren *et al*. have proposed Faster RCNN, in which constructed a candidate region network is constructed and the candidate region frame-extraction link incorporated into a neural network model, thus achieving end-to-end training and prediction for the first time[66]. The above models belong to the two-stage method; that is, the detection process can be divided into candidate frame extraction and convolution classification. This method has high positioning and detection accuracies but has much slower detection speeds when compared to the one-stage method without the link of candidate frame extraction, which directly converts the problem of frame positioning into the regression problem. The most representative one-stage method is the YOLO model proposed by Redmon *et al*. This model directly completes tasks, such as feature extraction, frame regression, and classification, in a network. The detection speed can reach 45 frames per second, which meets the requirements for real-time performance[67]. Since then, in response to the shortcomings of YOLO, scholars have proposed SSD[68], YOLOv2[69], and YOLOv3 models[70] for continuous optimization, thus materially improving target detection technology.

In terms of object segmentation, Long *et al*. have proposed a fully convolutional network (FCN). This model uses VGG16[71] as the backbone skeleton and uses convolutional layers instead of fully connected layers, restores the size of the feature map by means of deconvolution operations, and also defines a jump layer to improve segmentation accuracy. The segmentation accuracy in the VOC dataset reaches 62.2%[72]. Badrinarayanan *et al*. have proposed the SegNet model and designed an encoder-decoder symmetric structure, which can extract high-dimensional features through convolution, using up-sampling to recover the feature map size, yielding excellent segmentation performance[73]. Chen *et al*. have proposed an excellent series of semantic segmentation models, called DeepLab, developed from DeepLabv1, which initially used cavity convolution and conditional random fields, to produce DeepLabv3, which has an encoder-decoder structure. The segmentation accuracy in the VOC dataset also improved from 71.6 to 89.0%[74, 75].

The proposal of excellent algorithms and models has promoted the research and application of target region-extraction technology in various industries. Silva *et al*. used CNN to build a model suitable for detection of concrete cracks under different conditions, solving the problem of limitations of traditional image-processing methods[76]. Liang *et al*. proposed a dual CNN consisting of a CNN and FCN to identify cracks on concrete bridges, where CNN was used to exclude interference information, such as speckles, and FCN was used to extract crack features. This model has high recognition accuracy[77].

With the rapid development of CNN theory, 2D detection and segmentation framework with general significance has basically matured and offered many choices, which makes the target area monitoring task simple and reliable. Thus far, the structure of interest can be effectively detected and segmented by training images of specific targets, providing reliable input for subsequent point cloud processing.

4. Research status of 3D point-cloud processing

The original 3D point cloud usually has many defects, which are not conducive to the direct computer processing and reduces the accuracy of monitoring. Generally speaking, point cloud processing is a necessary part of the monitoring task and common defects in point clouds include noise, cavities, and unevenness. In addition, the 3D point cloud is discrete data and there is no topological correlation information between points. Usually, a topological surface needs to be generated for further processing and the generation of the surface involves feature information, such as the normal of the 3D point cloud. Thus, 3D point-cloud processing includes several steps, such as filtering, smoothing, feature estimation, and surface reconstruction. Scholars have conducted in-depth research for each processing link, resulting in a large number of excellent algorithms.

Point-cloud smoothing is designed to solve problems, such as uneven point-cloud density, small cavities, large surface fluctuations, and overlapping. Moving least squares, as one of the most widely used smoothing algorithms, was first proposed by Lancaster *et al*.[78]. The moving-least squares method uses the coefficient vector and basis function for fitting and then introduces the probability of compact support. It is considered that any point is only affected by other points in its neighborhood and is not associated with the outside of the neighborhood. This determines that this method has smoothing effects incomparable to any other algorithms. After years of development, scholars have continuously improved the moving-least squares method.

As an unordered set of 3D points in space, a point cloud generally only contains coordinate information, which results in a lack of support for key features when performing high-level operations on a point cloud. To this end, scholars have proposed a series of feature estimation algorithms. Rusu *et al*. have proposed a point-feature histogram as a high-dimensional feature for describing the geometric attributes of the point neighborhood[79]. This feature has a constant density and attitude and provides good initial conditions for operations, such as registration, and improves the evaluation quality of external geometric features of the structure. After presenting the point-feature histogram, Rusu *et al*. have modified the mathematical expression and greatly reduced the calculation time through an optimization method, with the optimized feature called the quick-point feature histogram[80]. Normals are another important feature of a point cloud and have received wide attention from scholars. Currently, the normal estimation methods are roughly divided into three types: the local neighborhood plane fitting method[81], Voronoi/Dalaunay method[82], and a method based on robust statistics[83]. These three algorithms are robust enough to calculate normal information from the point cloud but cannot guarantee the consistency of the normal directions.

Point cloud-surface reconstruction methods mainly include the triangular mesh and implicit function methods[84]. Points are projected onto the plane using the triangular mesh method, with the points on the plane triangulated and a certain triangular patch expanded as the initial surface, which reconstructs as jagged[85]. The main representative algorithm of the implicit function method is Poisson’s surface-reconstruction algorithm. This algorithm provides a point cloud with normal information and solves the surface-indication function, thus distinguishing the interior and exterior of the surface model[86]. The typical algorithm of the implicit function method is the Poisson’s surface reconstruction algorithm. This algorithm is able to generate a smooth watertight surface, automatically filter noisy points, and has high tolerance for an uneven point cloud[87]. V. Estellers *et al*. have proposed a more robust Poisson's surface reconstruction, which transforms the surface reconstruction problem into a convex function minimization problem of the surface internal indicator function and used the Huber penalty, instead of least-square fidelity term, to obtain better results, even under the environment of nonuniform sampling and noisy points[88].

In fact, point-cloud processing is not an established model. Common point-cloud filtering, local feature extraction, and 3D reconstruction operations are optional rather than necessary. The specific choice depends on the requirements of accuracy and speed. To date, point-cloud processing is also developing in a data-driven direction. Deep networks are expected to be better applied in traditional frameworks to achieve robust point-cloud processing.

5. Conclusion

This study summarizes the latest research in civil engineering health monitoring based on stereoscopic vision and presents the available machine-vision technologies for civil health detection by category. The study particularly expounds the developmental status and problems of stereoscopic-vision technology and offers suggestions for development.

The application of machine-vision in civil structural health monitoring is becoming more and more widespread. As scholars continue to study monitoring in terms of stereo vision, feature recognition of civil structures, and post-processing of point clouds, the application of machine-vision in civil structural health monitoring is progressing toward improved developments in efficiency, accuracy, robustness, precision, and degree of automation. In addition, the monitoring field of view is expanding and real-time performance continuously improving, which is a good trend. At the same time, there are challenges. For example, most current research results still require manual input and the degree of automation in the acquisition of machine vision elements, such as images or videos, needs to be improved. Most studies use flat surface samples and the influence of the surface quality or material of the samples on monitoring results remains to be explored. When the weather and other environmental factors change, the stability of a machine vision system remains a challenge. The above challenges still require more research effort.

Based on the above discussion, this article draws the following conclusions:

(1) The acquisition of the geometric features on an object's surface has always been the ultimate goal of visual 3D reconstruction technology and is also the focus of practical applications.

(2) The rapid development of deep learning has greatly promoted the development of visual perception technology and is gradually influencing the field of 3D vision, thus providing remarkable support for traditional 3D reconstruction technology. However, there is an urgent need to find a model for efficient collaboration between deep learning and traditional, optical, geometry methods.

(3) The development of CNN has greatly promoted the development of target area extraction technology, making accurate, real-time, and stable target detection and segmentation possible, and significantly improving the accuracy of damage detection in civil infrastructure.

(4) There has been relatively little research on the basic theory of 3D visual-reconstruction technology, which is particularly critical for tracking 3D deformation in a dynamic or seismic environment.

(5) The integration of intelligent 3D visual-reconstruction technology and deep learning will become a key, frontier research direction.

Acknowledgment

This work was supported by the Key Generic Technology System of the Agricultural Industry in Guangdong Province (No.2019KJ139), the Key-area Research and Development Program of Guangdong Province (No.2019B020223003), Natural Science Foundation of Guangdong Province (No.2015A030310465).

References

[1] Young-Jin Cha, Choi Wooram, Suh Gahyun, *et al*. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Computer-Aided Civil and Infrastructure Engineering*. 2018; 33(9): 731-747.

[2] Christian Koch, Georgieva Kristina, Kasireddy Varun, *et al*. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics*. 2015; 29(2): 196-210.

[3] Xiao Liang. Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization. *Computer-Aided Civil and Infrastructure Engineering*. 2019; 34(5): 415-430.

[4] Osama Abdeljaber, Avci Onur, Kiranyaz Serkan, *et al*. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*. 2017; 388: 154-170.

[5] Gustavo-H Beckman, Polyzois Dimos, Cha Young-Jin. Deep learning-based automatic volumetric damage quantification using depth camera. *Automation in Construction*. 2019; 99: 114-124.

[6] Yang Xu, Wei Shiyin, Bao Yuequan, *et al*. Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network. *Structural Control and Health Monitoring*. 2019; 26(3): e2313.

[7] Chong-xun, Gui-yan, Qing, *et al*. A quantitative model for danger degree evaluation of staged operation of earth dam reservoir in flood season and its application. *Water Science & Engineering*. 2018; 01(11): 85-91.

[8] Billie-F Spencer Jr, Hoskere Vedhus, Narazaki Yasutaka. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*. 2019; 2(5): 199-222.

[9] Philipp Huthwohl, Brilakis Ioannis. Detecting healthy concrete surfaces. *Advanced Engineering Informatics*. 2018; 37(AUG.): 150-162.

[10] Zhenhua Zhu, Stephanie German, Ioannis Brilakis. Visual retrieval of concrete crack properties for automated post-earthquake structural safety evaluation. *Automation in construction*. 2011; 20(7): 874-883.

[11] Thikra Dawood, Zhu Zhenhua, Zayed Tarek. Machine vision-based model for spalling detection and quantification in subway networks. *Automation in Construction*. 2017; 81: 149-160.

[12] Zhenqing Liu, Cao Yiwen, Wang Yize, *et al*. Computer vision-based concrete crack detection using U-net fully convolutional networks. *Automation in Construction*. 2019; 104: 129-139.

[13] Liang Sun, Jianchun Xing, Xun Zhang. An Extraction and Classification Algorithm for Concrete Cracks Based on Machine Vision. *IEEE Access*. 2018; 1.
[14] Baohua, Shan, Shijie, *et al*. A stereovision-based crack width detection approach for concrete surface assessment. *KSCE Journal of Civil Engineering*. 2016; 2(20): 803-812.
[15] Xiangjun Zou, Haixin Zou, Jun Lu. Virtual manipulator-based binocular stereo vision positioning system and errors modelling. *Machine Vision & Applications*. 2012; 23(1): 43-63.
[16] Ali Behnood, Golafshani Emadaldin-Mohammadi. Predicting the compressive strength of silica fume concrete using hybrid artificial neural network with multi-objective grey wolves. *Journal of Cleaner Production*. 2018; 202: 54-64.
[17] Cao-Vu Dung. Autonomous concrete crack detection using deep fully convolutional neural network. *Automation in Construction*. 2019; 99: 52-58.
[18] Cao-Vu Dung, Sekiya Hidehiko, Hirano Suichi, *et al*. A vision-based method for crack detection in gusset plate welded joints of steel bridges using deep convolutional neural networks. *Automation in Construction*. 2019; 102: 217-229.
[19] Sattar Dorafshan, Thomas Robert-J, Maguire Marc. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Construction and Building Materials*. 2018; 186: 1031-1045.
[20] Fu-Chen Chen, Jahanshahi Reza-Mohammad-Reza. NB-CNN. Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion. *IEEE Transactions on Industrial Electronics*. 2017; PP(99): 1.
[21] Dong Liang, Zhou Xue-Feng, Wang Song, *et al*. Research on concrete cracks recognition based on dual convolutional neural network. *KSCE Journal of Civil Engineering*. 2019; 23(7): 3066-3074.
[22] Shiyao Cai, Ma Zhiliang, Skibniewski Miroslaw-J, *et al*. Construction automation and robotics for high-rise buildings over the past decades: A comprehensive review. *Advanced Engineering Informatics*. 2019; 42: 100989.
[23] Baohua Shan, Pan Yunfeng, Huo Xiaoyang, *et al*. Detection of slip for CFRP-concrete interface using stereovision method corrected by epipolar constraint. *Structural Control and Health Monitoring*. 2018; 25(10): e2212.
[24] Chenglin Wang, Luo Tianhong, Zhao Lijun, *et al*. Window zooming-based localization algorithm of fruit and vegetable for harvesting robot. *IEEE Access*. 2019; 7: 103639-103649.
[25] Guichao Lin, Tang Yunchao, Zou Xiangjun, *et al*. Color-, depth-, and shape-based 3D fruit detection. *Precision Agriculture*. 2019; 1-17.
[26] Guichao Lin, Tang Yunchao, Zou Xiangjun, *et al*. In-field citrus detection and localisation based on RGB-D image analysis. *Biosystems Engineering*. 2019; 186: 34-44.
[27] Niannian Wang, Zhao Xuefeng, Zhao Peng, *et al*. Automatic damage detection of historic masonry buildings based on mobile deep learning. *Automation in Construction*. 2019; 103(JUL.): 53-66.
[28] Kim, Hyunjun, Sim, *et al*. Comparative analysis of image binarization methods for crack identification in concrete structures. *Cement & Concrete Research*. 2017.
[29] Shengyuan Li, Zhao Xuefeng, Zhou Guangyi. Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. *Computer-Aided Civil and Infrastructure Engineering*. 2019; 34(7): 616-634.
[30] Nicolas Candau, Pradille Christophe, Bouvard Jean-Luc, *et al*. On the use of a four-cameras stereovision system to characterize large 3D deformation in elastomers. *Polymer Testing*. 2016; 56: 314-320.
[31] Lin Huang, Da Feipeng, Gai Shaoyan. Research on multi-camera calibration and point cloud correction method based on three-dimensional calibration object. *Optics and Lasers in Engineering*. 2019; 115: 32-41.
[32] Marcin Malesa, Kujawińska Ma-Gorzata, Malowany Krzysztof, *et al*. Application of Multi-camera DIC System for Measurements of Industrial Structures. *Procedia Engineering*. 2015; 114: 453-460.
[33] Qing Mei, Gao Jian, Lin Hui, *et al*. Structure light telecentric stereoscopic vision 3D measurement system based on Scheimpflug condition. *Optics & Lasers in Engineering*. 2016; 86: 83-91.
[34] Pei Zhou, Zhu Jiangping, Su Xianyu, *et al*. Experimental study of temporal-spatial binary pattern projection for 3D shape acquisition. *Applied Optics*. 2017; 56(11): 2995.
[35] Mingyou Chen, Tang Yunchao, Zou Xiangjun, *et al*. High-accuracy multi-camera reconstruction enhanced by adaptive point cloud correction algorithm. *Optics and Lasers in Engineering*. 2019; 122: 170-183.
[36] Tomohiro Yokoyama, Matsumoto Takashi. Development of stereo image analysis for measuring small deformation. *Procedia Engineering*. 2017; 171(Complete): 1256-1262.
[37] Dong Shuai, Yu Shanshan, Zheng Huang, *et al*. Target-based calibration method for multifields of view measurement using multiple stereo digital image correlation systems. *Optical Engineering*. 2017; 56(12): 1.
[38] Krzysztof Malowany, Malesa Marcin, Kowaluk Tomasz, *et al*. Multi-camera digital image correlation method with distributed fields of view. *Optics & Lasers in Engineering*. 2017; 98: 198-204.
[39] Yun-Chao Tang, Li Li-Juan, Feng Wen-Xian, *et al*. Binocular vision measurement and its application in full-field convex deformation of concrete-filled steel tubular columns. *Measurement*. 2018; 130: 372-383.

[40] Belen Ferrer, Espinosa Julian, Mas David. A method to measure small local strains in concrete surfaces using its natural texture and image cross-correlation. *Structural Control and Health Monitoring*. 2019; 26(9): e2410.
[41] Yunchao Tang, Li Lijuan, Wang Chenglin, *et al*. Real-time detection of surface deformation and strain in recycled aggregate concrete-filled steel tubular columns via four-ocular vision. *Robotics and Computer-Integrated Manufacturing*. 2019; 59: 36-46.
[42] Huang Hui, Dan Li, Hao Zhang, *et al*. Consolidation of unorganized point clouds for surface reconstruction. *Acm Transactions on Graphics*. 2009; 28(5): 1-7.
[43] Saurabh Gupta, Girshick Ross, Arbeláez Pablo, *et al*. Learning rich features from RGB-D images for object detection and segmentation. Springer; 2014. p.345-360.
[44] Ross-B Girshick, Donahue Jeff, Darrell Trevor, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014. p.580-587.
[45] Jure Žbontar, Lecun Yann. Stereo matching by training a convolutional neural network to compare image patches. *Computer Science*. 2015.
[46] Charles-R Qi, Su Hao, Mo Kaichun, *et al*. Pointnet: Deep learning on point sets for 3D classification and segmentation. *IEEE Computer Society*. 2016.
[47] Charles-Ruizhongtai Qi, Yi Li, Su Hao, *et al*. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*. 2017. p.5099-5108.
[48] Yin Zhou, Tuzel Oncel. Voxelnet: End-to-end learning for point cloud based 3D object detection. *Computer Vision and Pattern Recognition*. 2017.
[49] Hyunjun Kim, Ahn Eunjong, Shin Myoungsu, *et al*. Crack and noncrack classification from concrete surface images using machine learning. *Structural Health Monitoring*. 2019.
[50] Martin Mundt, Majumder Sagnik, Murali Sreenivas, *et al*. Meta-learning convolutional neural architectures for multi-target concrete defect classification with the concrete defect bridge image dataset. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2019. p.11196-11205.
[51] Kim Byunghyun, Soojin Cho. Image-based concrete crack assessment using mask and region-based convolutional neural network. *Structural Control and Health Monitoring*. 2019; 26(8): e2381.
[52] M-M Islam, Kim Jong-Myon. Vision-based autonomous crack detection of concrete structures using a fully convolutional encoder–decoder network. *Sensors*. 2019; 19(19): 4251.
[53] Mike Heath, Sudeep Sarkar, Thomas Sanocki, *et al*. Comparison of edge detectors: A methodology and initial study. *Computer Vision & Image Understanding*. 1998; 69(1): 38-54.
[54] R Adams, Bishof L. Seeded region growing. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2002; 16(6): 641-647.
[55] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 1986; PAMI-8(6): 679-698.
[56] Alain Tremeau, Borel Nathalie. A region growing and merging algorithm to color segmentation. *Pattern Recognition*. 1997; 30(7): 1191-1203.
[57] Jianbo Shi, Malik Jitendra. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; 22(8): 888-905.
[58] Yuri Boykov, Funka-Lea Gareth. Graph cuts and efficient ND image segmentation. *International Journal of Computer Vision*. 2006; 70(2): 109-131.
[59] Alastair-P Moore, Prince Simon-JD, Warrell Jonathan, *et al*. Superpixel lattices. *IEEE*. 2008; 1-8.
[60] Yaser-Ajmal Sheikh, Khan Erum-Arif, Kanade Takeo. Mode-seeking by medoidshifts. *IEEE*. 2007; 1-8.
[61] Alex Levinshtein, Stere Adrian, Kutulakos Kiriakos-N, *et al*. Turbopixels: Fast superpixels using geometric flows. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2009; 31(12): 2290-2297.
[62] R Achanta, Shaji A, Smith K, *et al*. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2012; 34(11): 2274-2282.
[63] Yann LeCun, Bottou Léon, Bengio Yoshua, *et al*. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 1998; 86(11): 2278-2324.
[64] Alex Krizhevsky, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*. 2012; 25(2).
[65] Kaiming He, Zhang Xiangyu, Ren Shaoqing, *et al*. Spatial pyramid pooling in deep convolutional networks for Visual recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2014; 37(9): 1904-1916.
[66] Shaoqing Ren, He Kaiming, Girshick Ross, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2015; 39(6): 1137-1149.
[67] Joseph Redmon, Divvala Santosh, Girshick Ross, *et al*. You only look once: Unified, real-time object detection.

Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p.779-788.
[68] Wei Liu, Anguelov Dragomir, Erhan Dumitru, *et al*. Ssd: Single shot multibox detector. *European conference on computer vision.* Springer; 2016. p.21-37.
[69] Joseph Redmon, Farhadi Ali. YOLO9000: Better, faster, stronger. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2017. p. 7263-7271.
[70] Joseph Redmon, Farhadi Ali. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. 2018.
[71] Karen Simonyan, Zisserman Andrew. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
[72] Jonathan Long, Shelhamer Evan, Darrell Trevor. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015. p.3431-3440.
[73] Vijay Badrinarayanan, Handa Ankur, Cipolla Roberto. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. arXiv preprint arXiv:1505.07293. 2015.
[74] Liang-Chieh Chen, Papandreou George, Kokkinos Iasonas, *et al*. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2016; 40(4): 834-848.
[75] Liang-Chieh Chen, Zhu Yukun, Papandreou George, *et al*. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)*. 2018. p.801-818.
[76] Silva, Wilson, Ricardo, *et al*. Concrete cracks detection based on deep learning image classification. *Proceedings*. 2019.
[77] Liang Yang, Li Bing, Li Wei, *et al*. Deep concrete inspection using unmanned aerial vehicle towards cssc database. 2017: 24-28.
[78] P., Lancaster, K., *et al*. Surfaces generated by moving least squares methods. *Math Comp*. 1981.
[79] Radu-Bogdan Rusu, Blodow Nico, Marton Zoltan-Csaba, *et al*. Aligning point cloud views using persistent feature histograms. *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems.* IEEE; 2008.
[80] Radu-Bogdan Rusu, Blodow Nico, Beetz Michael. Fast point feature histograms (FPFH) for 3D registration. *2009 IEEE international conference on robotics and automation.* IEEE; 2009. p.3212-3217.
[81] Niloy-J Mitra, Nguyen A-N, Guibas Leonidas. Estimating surface normals in noisy point cloud data. *Proceedings of the nineteenth annual symposium on computational geometry.* 2004. p.261-276.
[82] N Amenta, Bern M. Surface reconstruction by voronoi filtering. *Discrete & Computational Geometry*. 1999; 22(4): 481-504.
[83] Li Bao, Schnabel Ruwen, Klein Reinhard, *et al*. Robust normal estimation for point clouds with sharp features. *Computers & Graphics*. 2010; 34(2): 94-106.
[84] Matthew Berger, Levine Joshua-A, Nonato Luis-Gustavo, *et al*. A benchmark for surface reconstruction. *Acm Transactions on Graphics*. 2011; 32(2): 1-17.
[85] Ravikrishna Kolluri, Shewchuk Jonathan-Richard, O'Brien James-F. Spectral surface reconstruction from noisy point clouds. *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing.* 2004. p.11-21.
[86] Michael Kazhdan, Bolitho Matthew, Hoppe Hugues. Poisson surface reconstruction. *Proceedings of the fourth Eurographics symposium on Geometry processing.* 2006.
[87] Jules Morel, Bac Alexandra, Véga Cédric. Surface reconstruction of incomplete datasets: A novel Poisson surface approach based on CSRBF. *Computers & Graphics*. 2018; 74(AUG.): 44-55.
[88] Virginia Estellers, Scott Michael, Tew Kevin, *et al*. Robust poisson surface reconstruction. *International Conference on Scale Space and Variational Methods in Computer Vision.* Springer; 2015. p.525-537.