
Proceedings Paper:
Wang, P., Lin, Y., Ree, M. et al. (2 more authors) (2019) Computer vision methods for automating high temperature steel section sizing in thermal images. In: Proceedings of the IEEE Sensor Data Fusion Workshop. IEEE Sensor Data Fusion Workshop, 15-17 Oct 2019, Bonn, Germany. Institute of Electrical and Electronics Engineers (IEEE) . ISBN 9781728150864

https://doi.org/10.1109/SDF.2019.8916635

Computer Vision Methods for Automating High Temperature Steel Section Sizing in Thermal Images

Peng Wang†,Yueda Lin†, Ree Muroiwa‡, Simon Pike‡, and Lyudmila Mihaylova†
† *Department of Automatic Control and Systems Engineering*
The University of Sheffield
{peng.wang, ylin42, l.s.mihaylova}@Sheffield.ac.uk
‡ *Liberty Speciality Steels*
{ree.muroiwa, simon.pike}@specialityuk.com
Sheffield, United Kingdom

***Abstract*—This paper proposes a solution to autonomously measuring steel sections with images captured by a monocular, uncalibrated thermal camera. A fast structural random forest algorithm extracts the edges of the steel sections from sequentially coming image data. Two approaches are proposed that recognize the edges and remotely evaluate the size of the manufacturing objects of interest, which will facilitate automating the steel manufacturing process. Four sets of experiments are conducted, and the results show that our method achieves accurate dimension measuring results, with a root mean square error less than $2.5\ mm$, which is the maximum tolerance bound of the manufacturing process.**

***Index Terms*—Thermal measurement, Steel manufacturing, Monocular vision, Edge detection, Hot-state sizing**

I. Introduction

Computer vision based [1] and Light Detection And Ranging (LiDAR) based [2] dimension measuring are among the most extensively researched non-contact measurement methods. Without on the scene calibration of the cameras, computer vision based methods are less accurate than LiDAR based methods. However, compared to LiDAR equipment, cameras are usually smaller, portable and cheaper, which make computer vision based methods still popular even in scenarios such as high temperature forging [3] and welding [4]. Structural light based dimension measuring methods [5] combine the advantages of both the LiDAR and computer vision systems. Its application in hot and large forgings has shown the method outperforms the contact measurements obtained with callipers.

The computer vision systems are classified as monocular and binocular. Binocular systems have been widely applied to manufacturing processes, from measuring [6] to inspection [7] and fatigue crack detection [8]. Binocular systems achieve high measuring accuracy thanks to the complementary information that they collect and accurate depth estimation. However, we still find applications of monocular system in high accuracy manufacturing. In [1], they develop a real-time vision-based method to monitor the diameter of a workpiece during the turning process. The results are compared with the manual measurement by using a digital calliper. The vision-based method is found to be effective in measuring the diameter of the workpiece in real-time within an accuracy of 0.6%. Wu et al. [4] propose a monocular-vision-based method for online measuring the pose parameters of a weld stud. They derive an accurate mathematical model constrained by the measuring principle. Based on the model, they further calibrate and optimise the projective transformation parameters. The method is shown to be flexible, fast, and achieves high-precision measurement of the weld stud. Nevertheless, these two applications are not in high temperature environments. Bi et al. [9] propose to use only one Charge Coupled Device (CCD) camera to measure the dimension of forgings with temperatures of $800 \sim 1200$ °C. By designing and using both a digital and a physical filter, they manage to extract the edges of the forging very accurate. They do not report any physical dimension measuring results though.

Fig. 1. Thermal image of a steel section

Whether it is monocular or binocular, to achieve high accurate measurements, the camera calibration is essential. However, in scenarios where the camera calibration is not doable, we need to explore other methods.

In this paper, we aim to provide a high-accuracy monocular-

vision based solution to the high temperature steel section dimension measuring tasks. Particularly, the solution provided deals with the data captured by an thermal camera (figure 1) that is not calibrated due to: i) the camera is installed mainly for surveillance, therefore calibration was not done at the beginning; ii) the severe working conditions make it hard to calibrate the camera on the scene, unless the plant pauses which could cause massive economical loss. We exploit the geometrical information extracted from the thermal images to derive a set of parameters to map pixel numbers from the image space to physical size in the physical world, with the purpose to facilitate automating the steel section manufacturing process. The main contributions of this work are as follows.

• A framework for automating the manufacturing process of steel section sizing is proposed based on vision information processing;

• Steel section edges are extracted with high accuracy and in real time by adapting the structured random forests [10] method;

• The local area selectors are defined and used to enable the dimension measuring locally, which helps filter out the noise caused by distortions.

• Dimension sizing results from a thermal camera are provided and compared, and together, we analyse the advantages and disadvantages of different approaches.

The remaining part of this paper is the following. Section II describes how the steel section edges are extracted and segmented from the background; The mapping function between the image space and the physical world is given in Section III; Section IV is the experiment and analysis, and the paper is concluded in Section V.

II. Steel Section Detection

To measure the steel sections, we need to detect them from the images first. In this paper, the steel sections are detected in two steps: First, all the edges in the image are extracted; The background is then subtracted according to the intensity difference of the section and the background. After these two steps, we are able to detect the steel section along with the extracted edges.

A. Steel Section Edges Extraction

The edge extraction method used in this paper is based on the structured random forest algorithm developed by Pdollar et al. [10], which introduces a "structure" to the traditional random decision forest. The main idea of the random decision forest is to produce decision trees and train the split function

$$h(x, \theta_j) \rightarrow \{0, 1\} \tag{1}$$

where x is the input, θ_j is the trained parameter at node j of the tree, and $\{0, 1\}$ indicates the input x is split left or right to the subsequent nodes. When the inputs reach to the leaves of the trees, they are labelled as $y \in \mathcal{Y}$. The training process

Fig. 2. Edges extracted from one steel section image

of the decision trees is to maximize the information gain I_j [10] of the given node j

$$I_j = I\left(S_j, S_j^L, S_j^R\right) \tag{2}$$

where $S_j \in \mathcal{X} \times \mathcal{Y}$ is the training dataset with $\mathcal{X}$ the sample space, $S_j^L = \{(x, y) \in S_j | h(x, \theta_j) = 0\}$, $S_j^R = S_j \backslash S_j^L$, $x \in \mathcal{X}$ is a sample, and L and R indicate the left and right branches.

After the decision trees are trained, the structured random forest framework further maps the labels $y \in \mathcal{Y}$ into a discrete label set $c \in \mathcal{C}\{1, ..., k\}$. The similarity of the labels are measured by a intermediate mapping

$$\Pi : \mathcal{Y} \rightarrow \mathcal{Z} \tag{3}$$

These labels y with the similar $z \in \mathcal{Z}$ are mapped into the same discrete label c. With the hierarchical label mappings, the structured random forest manages to label each pixel and determine whether the pixel is part of an edge. By assembling a large number of trees with each response to a different feature channel (colour and image gradient, etc.), the final edge detection results are generated by considering the votes among all the trees in the structured forest. More details can be found in [10].

The structured random forest used in this paper is trained with Berkeley Segmentation Dataset and Benchmark (BSDS500), which produces fast and high accuracy edge detection results not only on the test sets but also on our thermal steel section images. Figure 2 shows the edges extracted from one of the steel section images by the structured random forest algorithm. We can see that the edges of the steel section are clearly extracted.

B. Background Subtraction

To facilitate dimension measurement, we need to subtract the background and enhance the steel sections. In our specific case study, as the intensities of the hot steel sections are distinguishable from the dark background, we first convert the RGB images to grey scale images, then binarise the grey scale

Algorithm 1 Edge Extraction and Background Subtraction

Input: I_{rgb}
Output: Detected section with edges I_{eb}
1: Edge Extraction
Structured Forests based Edge Detection
Non-Maximum Suppression $\rightarrow I_{edge}$ //Sharpen edges to one pixel
2: Background Subtraction
Binarise I_{rgb} according to Otsu's method $\rightarrow I_{bw}$ [11]
Morphological denoising $I_{bw} \rightarrow I_{mor}$
3: $I_{eb} = I_{mor} \odot I_{edge}$

images with a threshold $thrs$ obtained from the histogram information according to [11].

Pixels with intensities less than $thrs$ are set to 0 and 1 otherwise. As is well-known, the binary regions produced by thresholding could be distorted by noise and texture. Therefore, the morphological methods such as erosion and dilation are further used to process the image to remove these imperfections.

Algorithm 1 shows the pseudo-code for **Edge Extraction** and **Background Substraction**. I_{rgb} is the input image. It is processed by the structured random forest algorithm and I_{edge} contains all the edges extracted. To improve the dimension measurement accuracy, the Non-Maximum Suppression (NMS) method is applied to sharpen the edges extracted to one pixel. In parallel, I_{rgb} is first binarised as I_{bw} and then the morphological method is applied on I_{bw} to remove the imperfections caused by thresholding, which results with I_{mor}. The final detected steel section with edges extracted is denoted as I_{eb}, which is generated from

$$I_{eb} = I_{mor} \odot I_{edge} \tag{4}$$

where $\odot$ indicates the element-wise production of two matrices.

III. Mapping from Image Space to Physical Space

A. Spacial Resolution Information

The usual routine of vision measuring usually includes a camera calibration procedure, which is lacking here due to the reasons mentioned in the **Introduction**. This brings about the first challenge. Another problem we confronted is that the camera is mounted statically so we can only film single-view videos, which further stops us from automatically calibrating the camera.

To cope with the problem, we extensively explored the data. We found the following two attributes of the videos useful:

- As shown in figure 3, the physical distance w between the conveyor barriers (the width of the conveyor along X) is known, which helps to find the physical correspondence of one pixel.
- There is only one vanishing point (the intersection of the two green segments) in figure 3, and the objects captured on the conveyor have the foreshortening effect.

Algorithm 2 Conveyor Boundaries Extraction

Input: I_{eb}
Output: $f_1(x), f_2(x)$
1: Mask to Select Region of Interest
The mask is created by selecting the points around boundaries $\rightarrow I_{mask}$
The Mask is applied to the edge image $\rightarrow I_{ROIedge} = I_{edge} \cdot I_{mask}$
2: Straight Line Fitting
Line fitting in $I_{ROIedge} \rightarrow f_1(x), f_2(x)$

Fig. 3. Line fitting for conveyor boundaries

According to the space perspective projection as shown in figure 4, we can see that the objects of the same physical size seem to be smaller when the distances between the objects and the camera increase. Therefore, though the physical width of the conveyor remains constant, the width of the conveyor in pixels decreases as y increases. The physical size represented by one pixel increases as well. According to the photography triangulation (see figure 5), we have the ratio r_i between the physical size w and the pixel number w_{pi} as in

$$r_i = w / w_{pi} \tag{5}$$

with w the physical width of the conveyor as shown in figure 5, which corresponds to the digital width w_{p0} in the image space with $y = 0$. It can also be regarded as the physical length represented by one pixel at $y = 0$.

The physical width of the conveyor is known. In order to calculate r_i, we need to extract the conveyor boundaries correspond to the barriers, as shown in figure 3 to get w_{pi}. The two boundaries extracted are represented by $f_1(x)$ and $f_2(x)$, respectively. For a given $y = i$, the width w_{pi} of the conveyor in pixels can be determined.

Then we use equation (6) to calculate the physical size p_s of any steel sections.

$$p_s = r_i \cdot w_{si} \tag{6}$$

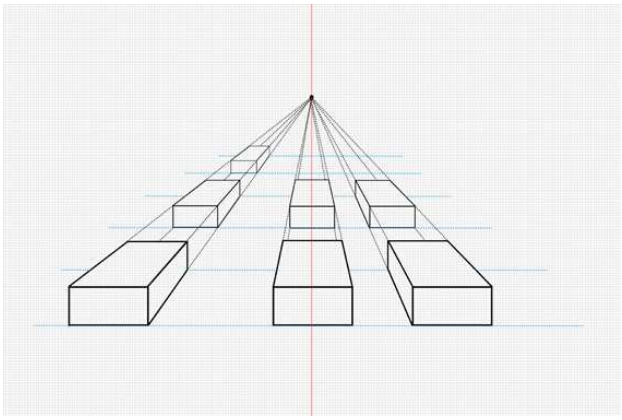


Fig. 4. Space perspective projection

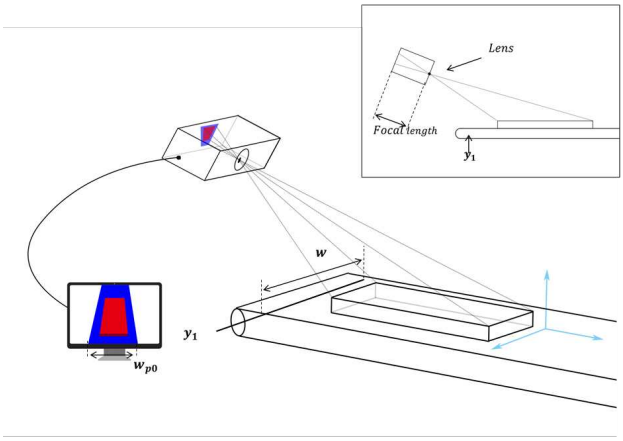


Fig. 5. Schematic diagram of the visual sizing

where w_{si} is the width of the section in pixel at $y = i$.

B. Dimension Measuring Algorithm

To convert pixel numbers in the image space to the physical dimensions, we need to recognise the section edges and count the pixel numbers between the two edges of interest. In this paper, with the purpose of comparison, we use two algorithms as given in algorithm 3 to recognise the edges and then convert them into physical sizes.

Algorithm 3 Boundaries Recognition and Sizing

Input: I_{eb}

Output: Number of pixels between two boundaries w_{pi}

1: Moore-Neighbor Tracing Algorithm
   Boundaries extraction $\rightarrow B_1, B_2, \cdots, B_n$
   Boundaries selection $\rightarrow B_{max}$
   Calculate the number w_{pi} of pixels between two boundaries
2: Boundaries Extraction with Line Fitting
   Initialize local area selector $I_{w\times h}$, step s
   Moore-Neighbor tracing algorithm to extract local boundaries $\rightarrow B_l$
   Line fitting in $B_l \rightarrow L_1$ and L_2
   Calculate the number w_{pi} of pixels between L_1 and L_2

Sub-algorithm 1 uses the Moore-neighbor tracing algorithm directly to get the edge information from the binarised results produced by [10]. The edges of the section are recognised and the diameter of the section in the image space is calculated by counting the number of the pixels between the two edges, as shown in figure 6. Sub-algorithm 2 introduces a local area selector $I_{w\times h}$ to constrain the boundary extraction area. $I_{w\times h}$ moves in the image matrix I_{eb} resulted from algorithm 1, with a stride of s both vertically and horizontally to get a local area $L_{w\times h}$. The boundaries in $L_{w\times h}$ are then extracted by Moore-neighbor tracing algorithm as B_l and further fit into line segments L_1 and L_2 by the first order polynomial regression. The diameter of the section in the image space is calculated by averaging the number of pixels between the two line segments.

With algorithm 3, we now have the number w_{pi} of pixels in the image space that corresponds to the steel section diameter, which is converted to physical dimensions by equation (6).

C. Homographic Extension

The above two subsections provide the solution to measure the steel sections when there is no camera calibration performed. However, the accuracy of the second method does not reach the expected error tolerance interval. Also, the Root Mean Square Error (RMSE) of both methods are quite significant. Thus, we involve homography in the solution to further improve the accuracy and narrow down the RMSE.

To derive the homography matrix, several sets (each set with four points) of points that locate along the conveyor barriers are selected. Figure 7 shows one set of the points A, B, C and D, the coordinates of which are denoted as

$$\begin{bmatrix} x_A & x_B & x_C & x_D \\ y_A & y_B & y_C & y_D \end{bmatrix}^T \tag{7}$$

The corresponding coordinates after applying homographic transformation are denoted as

$$\begin{bmatrix} \tilde{x}_A & \tilde{x}_B & \tilde{x}_C & \tilde{x}_D \\ \tilde{y}_A & \tilde{y}_B & \tilde{y}_C & \tilde{y}_D \end{bmatrix}^T \tag{8}$$

where x_A, x_B, x_C, x_D and y_A, y_B, y_C, y_D are the x and y coordinates of points A, B, C and D, respectively. And those with tildes in equation (8) are the corresponding coordinates after homographic transformation. These coordinates satisfy $\tilde{x}_A = x_A$, $\tilde{x}_B = x_B$, $\tilde{x}_C = x_A$, $\tilde{x}_D = x_B$, $\tilde{y}_A = y_A$, $\tilde{y}_B = y_B$, $\tilde{y}_C = y_C$, and $\tilde{y}_D = y_C$.

To eliminate the errors caused by the points selection procedure, five sets of points are chosen to calculate the homography matrix and a final one given in equation (9) is obtained by averaging the five matrices resulted from each set of points.

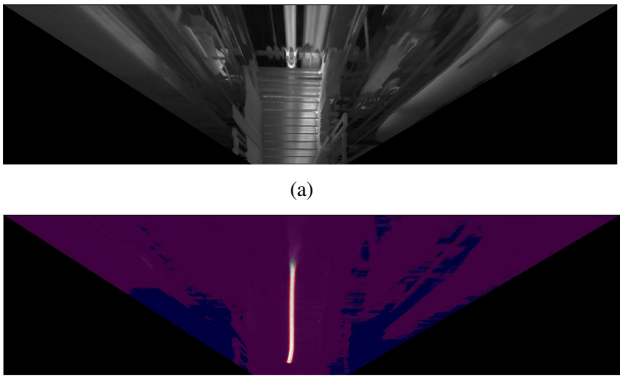
$$H = \begin{bmatrix} -0.0025277 & 0 & 0 \\ -0.0021371 & -0.0032857 & -0.0000043 \\ 0.9424616 & 0.3342812 & -0.0006065 \end{bmatrix} \tag{9}$$

(a) (b)

Fig. 6. Section recognition and edge extraction: (a) The original image; (b) Section extracted with edges

Fig. 7. Points for calculating homography matrix

By applying the homographic tranformation to the original images, we can get the top-view images as in figure 8(a). Then, the top-view images are further processed by algorithm 1 and 3 to get the physical dimensions of the steel sections.



(a) (b)

Fig. 8. Homography transformation: (a) The image of conveyor after homography transformation; (b) Section recognition on transformed image

IV. Experiments ans Analyses

Four sets of experiments are conducted to demonstrate the effectiveness of the methods proposed. The configuration of each set are as follows.

- **Experiment 1:**
  - **Data:** Original images;
  - **Pixel Counting:** Algorithm 3, sub-algorithm 1.
- **Experiment 2:**
  - **Data:** Homographic images;
  - **Pixel Counting:** Algorithm 3, sub-algorithm 1.
- **Experiment 3:**
  - **Data:** Original images;
  - **Pixel Counting:** Algorithm 3, sub-algorithm 2.
- **Experiment 4:**
  - **Data:** Homographic images;
  - **Pixel Counting:** Algorithm 3, sub-algorithm 2.

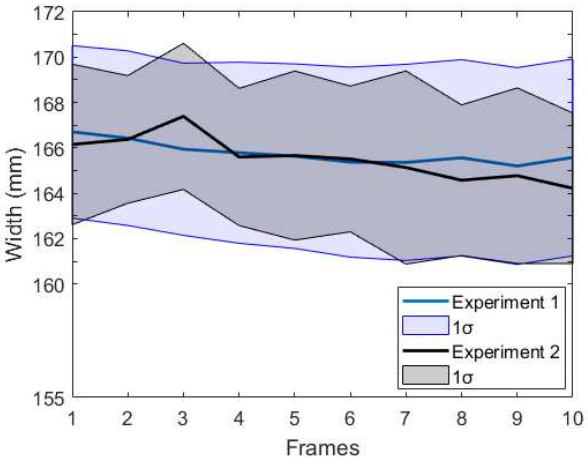
In each set of the experiments, 10 frames of a video filmed by a statically-mounted, uncalibrated thermal camera are processed. The steel section to be measured is a cylindrical one with ground truth diameter $165mm$. The diameter and the corresponding RMSE are calculated respectively as follows

$$l = \sum_{i=1}^{M} \bar{l}_i / M, \quad \text{with} \quad \bar{l}_i = \sum_{j=1}^{M_{ij}} l_{ij} / M_{ij} \tag{10}$$

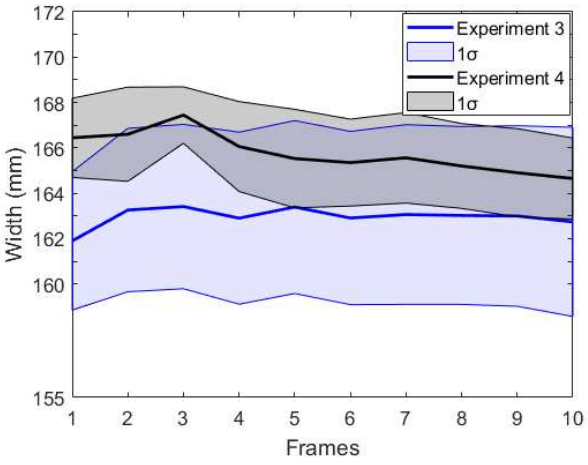
$$RMSE = \sqrt{\sum_{i=1}^{M} \left(\bar{l}_i - l\right)^2 / M} \tag{11}$$

where $i = 1, \cdots M$ is the index of the frames, l_{ij} with $j = 1, \cdots, M_{ij}$ indicates the section diameter corresponds to the j-th y coordinate, $\bar{l}_i$ is the averaged physical diameter from frame i, and l is the mean from the M frames. As we processed 10 frames, so M is set to 10 in the paper.

Figure 9(a) shows the results of **Experiment 1** and **Experiment 2**. We can see that by processing the original images directly, we can get fairly accurate results. However, after homographic transformation, both the diameter estimation



(a)



(b)

Fig. 9. Experiment results: (a) Experiment 1-2; (b) Experiment 3-4

and the RMSE are improved. The reason lies in that by the the homographic transformation, the distortion caused by pixels further away from the camera are corrected to some extent. Figure 9(b) shows the results of **Experiment 3** and **Experiment 4**. We can see that both the diameter estimation and the RMSE are large while processing the original images. However, after the homographic transformation, the diameter estimation accuracy is significantly improved with quite small RMSE. The reason why sub-algorithm 2 given in algorithm 3 reports poor results while processing the original images is the accuracy of the slope and intersection of the extracted lines in the selected area could be affected by the distortion of the steel sections easily. After the homographic transformation, the two edges are almost parallel, which makes the extracted line parameters more stable.

By comparing all the four results, we can see that **Experiment 4** shows the best results, both with the smallest Mean and RMSE. The reasons are twofold: 1) the homographic transformation helps in correcting the distortions, which makes sub-algorithm 2 more robust; 2) The line fitting method implicitly filters out some system noises contained in the selected area ($w \times h = 20 \times 85$ with stride $s = 5$ in our experiments), which helps to improve the steel section measuring results further.

V. Conclusions

In this paper, we propose a thermal image based steel section dimension-measuring method. The edges of the sections are extracted from the background first. They are further recognized, and the number of pixels between two edges of interest is calculated by two different methods. By exploiting the only physical distance between two conveyor boundaries, our approach can form a mapping relation between the image space and the physical world. Fairly accurate results are achieved by processing the original images directly. To make our method more robust and accurate, we applied a homographic transformation to eliminate distortions in the images, and even better results are achieved.

Future work will focus on deep learning methods with adaptable area selectors and uncertainty quantification.

Acknowledgements. We are grateful to the Liberty Speciality Steels (LSS), UK for providing the case studies, data and to the Knowledge Exchange grant (Internet of Things for Overcome Barriers in the Steel Rolling Measurement Technology) the LSS and the University of Sheffield. We also thank NSFC(61703387).

References

[1] J. Che and M. Ratnam, "Real-time monitoring of workpiece diameter during turning by vision method," *Measurement*, vol. 126, pp. 369–377, 2018.

[2] J. Zhu, M. Li, Y. Jiang, T. Xie, F. Li, C. Jiang, R. Liu, and Z. Meng, "Research on online 3d laser scanner dimensional measurement system for heavy high-temperature forgings," in *AOPC 2017: 3D Measurement Technology for Intelligent Manufacturing*, vol. 10458. International Society for Optics and Photonics, 2017, pp. 104 581Q–1–104 581Q–9.

[3] X. Fu, B. Liu, and Y. Zhang, "An optical non-contact measurement method for hot-state size of cylindrical shell forging," *Measurement*, vol. 45, no. 6, pp. 1343–1349, 2012.

[4] B. Wu, F. Zhang, and T. Xue, "Monocular-vision-based method for online measurement of pose parameters of weld stud," *Measurement*, vol. 61, pp. 263–269, 2015.

[5] Z. Tian, F. Gao, Z. Jin, and X. Zhao, "Dimension measurement of hot large forgings with a novel time-of-flight system," *The International Journal of Advanced Manufacturing Technology*, vol. 44, no. 1-2, pp. 125–132, 2009.

[6] Z. Wang, R. Liu, T. Sparks, H. Liu, and F. Liou, "Stereo vision based hybrid manufacturing process for precision metal parts," *Precision Engineering*, vol. 42, pp. 1–5, 2015.

[7] X. Sun, J. Gu, S. Tang, and J. Li, "Research progress of visual inspection technology of steel productsa review," *Applied Sciences*, vol. 8, no. 11, pp. 2195–2220, 2018.

[8] X. Kong and J. Li, "Vision-based fatigue crack detection of steel structures using video feature tracking," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 9, pp. 783–799, 2018.

[9] C. Bi, J. Fang, D. Li, and X. Qu, "Study on application of color filters in vision system of hot forgings," in *Optical Measurement Technology and Instrumentation*, vol. 10155. International Society for Optics and Photonics, 2016, pp. 1 015 522–1–1 015 522–9.

[10] P. Dollár and C. L. Zitnick, "Fast edge detection using structured forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 8, pp. 1558–1570, 2014.

[11] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.