Journal of Intelligent Manufacturing
https://doi.org/10.1007/s10845-023-02251-9

Towards automated remote sizing and hot steel manufacturing with image registration and fusion

Yueda Lin[1] · Peng Wang[2] · Zichen Wang[1] · Sardar Ali[1] · Lyudmila Mihaylova[1]

Received: 3 February 2023 / Accepted: 13 October 2023

Abstract

Image registration and fusion are challenging tasks needed in manufacturing, including in high-quality steel production for inspection, monitoring and safe operations. To solve some of these challenging tasks, this paper proposes computer vision approaches aiming at monitoring the direction of motion of hot steel sections and remotely measuring their dimensions in real time. Automated recognition of the steel section direction is performed first. Next, a new image registration approach is developed based on extrinsic features, and it is combined with frequency domain image fusion ofoptical images. The fused image provides information about the size of high-quality hot steel sections remotely. While the remote sizing approach keeps operators informed of the section dimensions in real time, the mill stands can be configured to provide quality assurance. The performance of the developed approaches is evaluated over real data and achieves accuracy above 95%. The proposed approaches have the potential to introduce an enhanced level of autonomy in manufacturing and provide advanced digitised solutions in steel manufacturing plants.

Keywords Vision measurement · Steel manufacturing · Camera calibration · Sizing

Introduction

The development of the new generation of Industry 5.0 is based on a digital transformation of Industry 4.0. In the past years, Industry 4.0 has introduced advanced intelligent manufacturing technologies based on artificial intelligence and data analysis to enable an increase in production and the enhancement of operation efficiency (Barari et al., 2021; Qian et al., 2021). Different from the automation of Industry 4.0, Industry 5.0 aims to enhance safety in human-computer cooperation and optimise the overall performance of human-computer (Leng et al., 2022) systems. These objectives are also valid for high-quality steel production.

High-quality steels are essential in industrial sectors such as aerospace, oil and gas production. High-quality steels are produced via a rolling process during which the steel acquires the desired shape, size and desired mechanical property. Introducing autonomy in the monitoring and control of the steel rolling process is essential for improving the efficiency of the whole production. Achieving a high-quality standard for the pure metal pieces called *ingots*, as well as the steel sections manufactured from ingots, is important.

However, despite the technological advances, many steel rolling plants nowadays are still relying on human operators to manually control and monitor the manufacturing rolling process. It has been shown that the long-term exposure to a high-temperature, intense light environment in steel factories could cause injuries, particularly to human eyes (Hoyos and Zimolong, 2014). In order to address such and related challenges, this paper presents approaches that can be used for remote monitoring and sizing of steel sections.

✉ Yueda Lin
yueda.lin1214@gmail.com

Peng Wang
p.wang@mmu.ac.uk

Zichen Wang
zwang294@sheffield.ac.uk

Sardar Ali
smali5@sheffield.ac.uk

Lyudmila Mihaylova
l.s.mihaylova@sheffield.ac.uk

[1] Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, UK

[2] Department of Computing and Mathematics, Manchester Metropolitan University, M15 6BH Manchester, UK

Published online: 15 November 2023

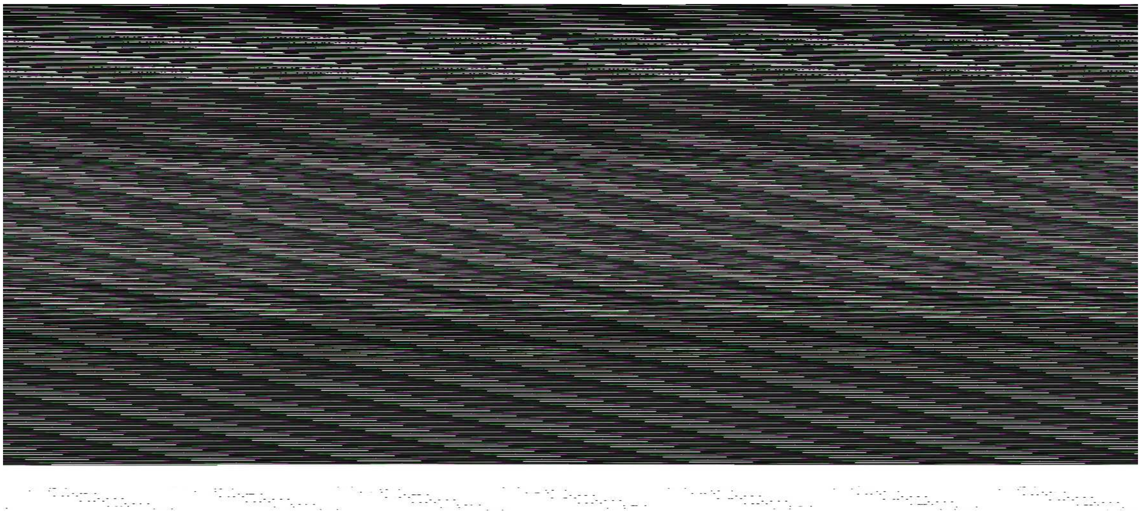


Fig. 1 Overall diagram of the steel production process. This paper focuses on the starting and the final stages of the production, and provides computer vision techniques for autonomous sizing and hot steel section direction recognition. These stages are indicated in the upper left part with the dashed lines, and respectively, in the upper right part with the dashed lines, for the section sizing

A monocular real-time measurement algorithm is proposed in our previous work (Wang et al., 2020, 2019). A fast structural random forest algorithm detects edges of steel sections, and the detected edges are further enhanced by a regression algorithm to suppress edge detection noises and increase the measurement accuracy. The steel section dimensions are then calculated based on the regression results in the image plane and the results are next converted from the image plane to the physical plane to represent real sizes.

However, the monocular camera measurement system depends significantly on the camera calibration. Due to safety concerns and unwillingness to disrupt the rolling process, the camera calibration becomes extremely difficult in remote sizing of hot steel sections when the monitoring cameras are positioned at dozens of metres' distance from the rolling mills (Wang et al., 2019). In addition, in such remote sizing cases there are requirements for a certain estimation accuracy to be achieved which is required to be less than 2.5 mm error for the estimated hot steel sections. In order to achieve accurate remote sizing, instead of using one camera, a two camera measurement system is proposed in this work.

The developed framework and measurement system comprises of two GoPro® cameras. Due to glares emitted from the high-temperature steel sections, the two cameras need to be carefully configured, especially for working at a fast shutter speed. This helps to reduce significantly the edge blur caused by glares in the input image. The overall diagram of the steel rolling system is shown in Fig. 1. The cameras are situated about 2.5 ms apart from each other and are above the steel rolling plant. The cameras are aligned approximately with respect to their image planes.

Ingots are reheated and moved to the rolling line, where steel sections are processed by a few mills to change their size and shape. Dimension measurement during the rolling process plays a key factor for quality assurance and is performed wherever necessary. In the considered industrial case study, sections are measured after the last mill. Laser range finder measurements are provided only at the last mill and these measurements are used as ground truth to assess the performance of the developed computer vision remote sizing measurement system.

In addition, the positioning of the ingots in the mills is an important factor determining the quality of the steel. In particular, the top and bottom ends of the ingots are of different sizes and impurity contents. The hot steel sections have to be directed with their bottom ends toward the blooming mill first. Currently, human operators take part of the monitoring process and make decisions whether the ingots are placed correctly - with the leading side upfront. A computer vision pattern recognition system can replace human operators in such repetitive tasks. It can not only increase the accuracy and reliability of the decision making process, but can also protect the operators from potential eye injuries that the intense glares could cause.

Initial computer vision results for remote sizing with edge detection approaches, preliminary results for image registration and image fusion are reported in Wang et al. (2019, 2020), Lin et al. (2021) where a checkerboard is only used at the beginning of the process, for the camera calibration.

Contributions

This paper presents an innovative approach for remote sizing of objects with optical camera data. An image registration approach based on extrinsic image features is proposed which includes a virtual checkerboard and copes efficiently with measurement errors due to environmental conditions and variations of section dimensions and the different heights at which measurements are taken.

The images provided by two optical cameras are registered first and then fused using several types of Discrete Wavelet Transforms (DWTs) (Sundararajan, 2016). A detailed comparison is made to evaluates the performance of the image fusion algorithms for remote sizing of the steel sections. The paper also considers the ingot direction recognition problem at the very beginning stage of the whole rolling process. A solution aimed at automating the ingots monitoring process and reducing the involvement of human operators is proposed. Although the approach presented is applied to industrial tasks, it can also be applied to other areas such as forestry.

The main contributions of this work can be summarised as follows:

i) A new two-camera-based approach for hot steel section remote sizing is proposed, which incorporates efficient image fusion methods. The approach is robust to environmental changes, including high temperature, evaporation and other sources of noise. It achieves high precision results for non-contact measurements in medium-range distances.

ii) A new image registration approach is proposed which uses extrinsic features from a virtual checkerboard and this approach also improves the system's robustness against environmental changes.

iii) An efficient image recognition approach is developed for ingot direction recognition, offering a new perspective on automating the recognition of steel ingot orientation.

iv) The proposed framework has been validated using real-world data collected from a high-quality steel manufacturing plant, demonstrating the efficiency of the proposed approach and its potential for industrial applications. The achieved remote sizing accuracy is above 95% with a tolerance range of 2 mm, representing a significant technical advance in the process of remotely measuring the steel sections.

Section "Related Work" gives an overview of related works. The ingot direction recognition approach is given in Section "The Proposed Approach for Ingots Direction Recognition". Section "The Proposed Approach for Remote Sizing of Steel Section" elaborates the proposed two-camera sizing system. Section "Performance Validation and Evaluation" analyses the performance of different fusion results. Section "Conclusions" summarises the results and discusses directions for future work.

Related work

Computer vision technologies play a critical role in Industry 4.0 by providing a high level of automation of the production for real-time evaluation and processing, improving productivity and reducing waste. Recent data suggest that organisations report up to 12% increases in manufacturing production, factory utilisation, and labour productivity after investing in smart factory projects (Lu et al., 2016). Technologies introducing autonomy are developing rapidly and are expected to grow in long term, with adopters already reaping the benefits of increased profit margins while non-adopters lag behind (Meindl et al., 2021). Robotics inspection systems can operate faster and with enhanced automation compared with human operators, as faults and exceptions are easily identified. Kuo proposes a deep learning-based method for foreign object detection in the graphic card assembly line, which can effectively detect and mark foreign objects (Kuo and Nursyahid, 2022).

Management structures constructed with computer vision systems enable safe cooperation between robots and human operators, increasing their efficiency. In addition, the Industrial Internet of Things (IIoT) provides connectivity between activities at different levels from the bottom to the top. Vision based technologies can further enhance the functionality and usability of sensors which reduces the IoT bandwidth needs. The IoT combined with vision based technologies continues to play a crucial role in industrial automation and intelligent manufacturing, bringing new opportunities and challenges for sustainable development of enterprises (Javaid et al., 2022).

In the steelmaking process, accurate real-time non-contact steel detection and measurement can guarantee high quality, can help avoiding hazards and financial loss. Vision-based systems have been widely applied for detecting defects on steel surfaces, as demonstrated in the work of Luo and He (2016). Zhou developed an online approach based on feature line reconstruction of stereo vision with high measurement accuracy for estimating the diameter of hot forgings (Zhou et al., 2018). However, this method requires high ambient lighting and accurate camera calibration.

Fig. 2 These four images show a hot steel section moving on the mill. The images **a** and **c** given on the left column are captured with the left camera. The images **b** and **d** shown at the right column are taken by the right camera. Feature points extracted by the SURF method are shown with a green plus sign with a circle on images **a** and **b**. Feature points extracted by the FAST method on images **c** and **d** are visualized with a green plus sign

Similar to industrial applications where remote sizing is necessary, there is a demand in forestry for non-contact tree size estimation. A single camera method was developed in Putra et al. (2021), where data collected in advance are used for camera calibration. In Eliopoulos et al. (2020), remote measuring of the diameter and height of trees with binocular cameras is presented, achieving a measurement accuracy of 1–2 cm error at 1–5 ms from the measured trees.

Methods for image registration

Image matching, also known as image registration, is the process of establishing the correspondence of pixels from two or more images. This involves finding geometric relationships, and the multiple scenes are combined into a single integrated image (Zitova and Flusser, 2003). In order to understand well the changes in a scene or an object over a long period of time, images are captured from various sensors at different times and from multiple viewpoints. The image registration process is mainly divided into four stages: feature detection, feature matching, transform model estimation, and transformation (Zitova and Flusser, 2003). Image registration methods have been an active area of research and a wealth of methods have been developed for medical purposes (Brock et al., 2017; Guan et al., 2018). Some registration algorithms also use deep learning (Boveiri et al., 2020; Chen et al., 2021). However, deep learning methods require large datasets, and the learning process of the networks still needs improvement, especially under challenging environmental conditions.

There is a range of methods for feature extraction from images in an automatic way (Mutlag et al., 2020). These feature points are often corner points or reference points of an object, which can describe the shape and position of the object. Traditional image feature extraction algorithms such as the scale invariant feature transformation (SIFT) focus on extracting key points, image's corner points or edge features(Dalal and Triggs, 2005). Based on SIFT, many other algorithms were developed with highly accurate performance, such as the Speeded Up Robust Features (SURF) method, which reduce the amount of calculation and speed up the feature extraction process so that it can meet the real-time requirements (Bay et al., 2006; Tafti et al., 2018). In addition to traditional feature detection methods, feature extraction algorithms based on convolutional neural networks (Dargan et al., 2019) have also become popular in recent years. Encouraging results in accuracy and computing speed are reported in Zheng et al. (2017). The methods mentioned above extract intrinsic feature points in images. However, if the noises present in images dominate and in combination with low lighting conditions, such algorithms cannot extract enough feature points. In the proposed framework, extrinsic feature points will be used in the image registration (matching) and fusion to cope with these challenges.

After detecting feature points from images, these feature points should be matched to each other. Feature matching can be performed in a number of ways one of which is by calculating Euclidean distances (Brock et al., 2017) of feature points in a pair of images. In addition to the spatial relationship between features, different feature descriptors (Tafti et al., 2018) and similarity metrics (Czolbe et al., 2021; Tong et al., 2019) are used to evaluate the results with respect to the matching accuracy and this is followed by other tasks such as image fusion (Ma et al., 2019). To register the images, the transform model between images need to be estimated. Next, the corresponding features obtained from the previous step are used to calculate the model. The choice of the transform model depends on the prior knowledge of the image acquisition process and the expected image distortion.

In the process of steel rolling, the challenging low-level illumination conditions are challenging for computer vision systems. The sparse feature points generated by conventional approaches lead to inaccurate matching results, as shown in Fig. 2. In this figure, the green points represent the feature points detected by SURF and FAST (Features from Accelerated Segment Test) algorithms (Tafti et al., 2018). Advantages of FAST algorithms consist in their efficiency in feature detection, which makes them suitable for real-time applications, including manufacturing. However, an improvement of the registration accuracy can be achieved based on extrinsic features thanks to a virtual checkerboard which is proposed in this paper.

Methods for image fusion

Image fusion aims to generate a high-quality image, with quality that is better than those of the separate images (Jin et al., 2017; James and Dasarathy, 2014). Different fusion algorithms can be broadly divided into transform domain fusion methods and spatial domain fusion methods in pixel-level image fusion (Hall and Llinas, 2001).

Generally, the spatial domain fusion method directly uses the intensity level of image pixels for image fusion. For example, the simple average, minimum, maximum, max-min and weighted average methods keep the pixels with low or high intensity at the same position of two images. More advanced methods include hue intensity saturation, Brovey transform, principal component analysis and guided filtering methods (Khan et al., 2021). Image registration is also often performed in the frequency domain (Tong et al., 2019).

The proposed approach for ingots direction recognition

The steel mill transports the hot steel sections during the forming process and transfers them onto another production line. The top and bottom ends of the ingot are different. If the top and bottom directions are reversed, the piece of material becomes unusable. Therefore, in order to inspect the sections autonomously, the orientation of the ingot needs to be recognized remotely. The following subsections present the main stages of the proposed approach for recognising the front and end parts of the hot steel sections.

Edge detection

In order to extract the cross-section of the ingot, the edge information is first obtained. Canny-based edge extraction (Canny, 1986; Song et al., 2017; Dollár et al., 2021; Dollár and Zitnick, 2014) is a feature extraction method that can represent the edge features of the original image and significantly reduce the amount of image information that needs to be processed. The specific implementation can be divided into four steps: Gaussian filtering, calculating the gradient strength and direction, non-maximum suppression, and double threshold detection to classify the edges into strong and weak edges. Non-maximum suppression and double threshold detection are used to find local maximum points and these are classified as part of strong edges and weak edges. An example of the original image and the image with the extracted edges is shown in Fig. 3.

Ingot contour segmentation

After the edge detection stage is completed, the next step is to classify these edges and to extract the edges belonging to the steel section. Many edge feature points provided by the edge operators do not belong to the steel section, as shown in Fig. 3b. Since these isolated edge points are due to oxides and sediments on the surface of ingots, the surface texture has a complex shape and the pixel intensity distribution is uneven. A significant portion of the misidentified edges is due to the texture of the steel material and strong lighting conditions (with temperature around and above 1000° C) inside the rolling mill. Therefore, a filter based on the SUZUKI contour algorithm (Suzuki, 1985) is employed to help define the hierarchical relationship between boundaries.

Figure 4a shows with dots the detected edge points. The different colors represent different contour groups based on contour algorithm classification. After that, only the prominent parent edges remain, which represent the contour of ingot as shown in Fig. 4b.

Algorithm 1 Hough gradient method for estimating the centre and radius of the circular steel section

```
Input: The results of Canny $E_k$, the gradient direction $\theta$ of each point in $E_k$.
Output: Circle Centers $C$ and radius $R$.
Estimate the location of the center of the circle
Initialize the centre voting space $V_c$, $a$ and $b$ are the width and height of the region of interest.
for $i = 1, \cdots, a$ do
    for $j = 1, \cdots, b$ do
        Extend forward along the gradient direction $\theta(i, j)$ of the point $E_k(i, j)$
        Every time meet a point :$V_c(i, j) += 1$
    end for
end for
Sort the voting space $V_c(i, j)$ and get $C$
Estimated radius
Initialize Circle radius Voter $V_r$
for $m = 1, \cdots, k$ do
    $r_m = |CE_m|$
    $V_r(r) += 1$
end for
Sort $V_r$, the larger the value, the more likely it is the radius
```

Cross-section Extraction

After the edge points are filtered, the sections are extracted by the Hough gradient method (Nixon and Aguado, 2020) as described in Algorithm 1. Algorithm 1 has two parts: finding the centre of the circle and finding the length of the circle radius. First, the circle centre is initialized by the voting space V_c. Next, the traversal of all the edge points is performed, and this traversal is expanded forward along the gradient direction. This removes all points that pass through the corresponding voting space $V_c(a, b) + 1$, where (a, b) is the position of the corresponding point. Finally, the voting space $V_c(a, b)$ is sorted, and the point with a higher number of votes is more likely to be the centre of the circle.

Let C be the circle centre, and let initialize the radius voting space $V_r(r)$ for the circle centre, where r is the radius. Next, we calculate the distance r from the edge points to the centre of the circle. Finally, the V_r space is sorted in order to get the radius of the cross section of cylinder ingots. The extracted edges are used as inputs for the section extraction with the Hough gradient method. One such result shown in Fig. 5. The range of the green box in the figure is the minimum bounding rectangle of the steel section determined by the previous edge detection.

The cyan box shows the minimum bounding rectangle of the circle detected by the Hough circle detection, while the blue circle and red dot represent respectively the detected circle and its centre. The circular cross-sections extracted from the algorithm are shown in Fig. 6. Due to the presence of sediment and oxides during the placement of ingots, there is a visible difference between the top and bottom of the steel ingots. The top row represents the top of the ingots, which appears darker in colour compared with its surroundings and exhibit a brain-like pattern due to the presence of oxides. In contrast, the bottom of the ingots exhibits a smoother and brighter surface with relatively higher temperatures and fewer oxides compared with the top side.

Image classifier of ingot directions

After the steel sections are segmented, the top and bottom images of each ingot constitute the patterns that need to be recognised. All other end images are compared to these patterns using the structural similarity index measure (SSIM) (Wang et al., 2004), which is a one of the best measures for evaluating the similarity between two images. When the algorithm detects the section, the section end $I_{section}$ is compared to I_{top} and I_{bot} with the SSIM which is calculated as described in Algorithm 2. The direction of a section is determined based on SSIM value.

Algorithm 2 SSIM Direction Recognition

```
Input: $I_{top}, I_{bot}, I_{section}$
Output: The direction of section $I_{section}$
Calculate the SSIMs between $I_{section}$ and $I_{top}, I_{bot}$
$SSIM_{top} = \boldsymbol{SSIM}(I_{top}, I_{section})$
$SSIM_{bot} = \boldsymbol{SSIM}(I_{bot}, I_{section})$
if $SSIM_{top} >= SSIM_{bot}$ then
    The direction of $I_{section}$ is $TOP$
else
    The direction of $I_{section}$ is $BOT$
end if
```

Fig. 3 Figure **a** shows the original image, Figure **b** shows the Features Extracted by Canny Detector

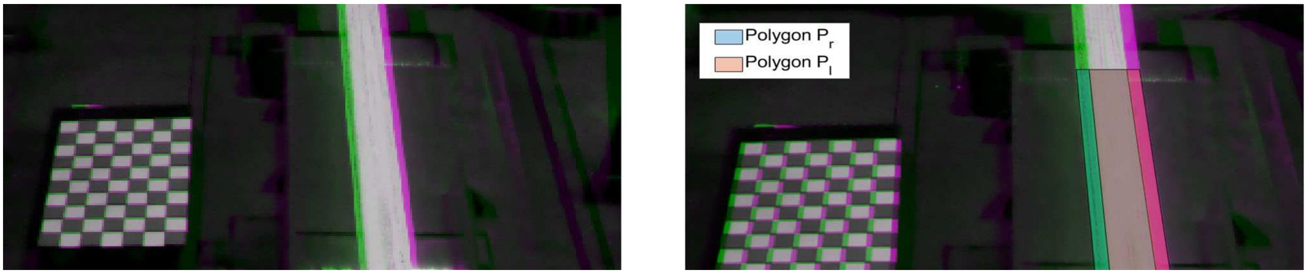


Fig. 4 Figure **a** shows feature points detected by the Canny edge detector on the surface of the ingot. Figure **b** shows results after filtering and the minimum bounding rectangle

The proposed approach for remote sizing of steel section

Before cutting hot steel sections into different shapes and sizes, they are moving on mills situated at different height with respect to the ground. This creates challenges to computer vision systems. Also a monocular measurement system is not able to deal with the plane difference due to the lack of depth information. Although, there are attempts to reconstruct depth, e.g. with deep learning methods and monocular data, the factory environment poses significant challenges to deep learning algorithms which require significant amounts of data for training and testing (Luo et al., 2018).

Achieving accurate vision-based measurements of the hot steel sections at distances between 10 and 30 ms with one camera is difficult. When the object of interest which is sized remotely has a certain thickness or is not in the same plane as the camera calibration plane, the estimated object sizes vary significantly from the actual object sizes. In contrast, traditional binocular cameras can obtain image depth by using the parallax effect and after calibration. However, the camera placement and environmental light conditions have an additional impact on the object sizing accuracy.

Binocular imaging systems are more accurate than monocular ones and hence a binocular camera solution is proposed in this paper. A key advantage of the proposed approach is that it is adaptive to the changes of the measurement plane. After acquiring the images from two cameras, each image pair are registered first. However, due to the changing lighting conditions in the factory and height of the mills, image registration based on intrinsic feature points from the images becomes unreliable. Furthermore, the large baseline of the two cameras leads to a significant difference in the estimated position of the object of interest in the images, which further increases the image registration difficulty. The approach proposed in this paper overcomes these difficulties

Fig. 5 Cross section extraction

by performing the image registration with extrinsic feature points from a checkerboard positioned close to the mills. The checkerboard is part of the camera calibration process and also it enhances the image registration process. The next Section "Image registration" describes this process in detail.

Image registration

The image registration process is given succinctly in Algorithm 3. The transformation matrix T_{RL} is calculated by matching the corner points of the checkerboards that are part of images I_L and I_R and the two images are registered successfully. In Algorithm 3 C_R denotes a corner point from the checkerboard of the right image and C_L represents a corner point from the checker board of the left image. The following Section "Image registration evaluation" describes in detail how the image registration process is evaluated.

Algorithm 3 Image Registration

```
Input: Images $I_L, I_R$
Output: The registered right image $I_{Rr}$
1: Extract the corner points $C_L, C_R$ of checkerboards in $I_L, I_R$
2: Calculate the geometric transformation $T_{RL}$ ,which transform $C_R \rightarrow C_L$
3: Apply $T_{RL}$ to $I_R$ :$T_{RL}(I_R) = I_{Rr}$
```

Image registration evaluation

To evaluate the quality of image registration, first extract the steel section and its edges in the images. After the steel section edges are obtained, the edges in the region close to the checkerboard are selected to be evaluated. The region of interest (ROI) in which the steel section is situated is determined automatically.

One way to automatically select the ROI around the steel section is based of the y coordinate at the highest point of the checkerboard and the lowest y coordinate of the checkerboard. Next, the highest and lowest y coordinate values are expanded with a preset pixels which is set to 10 pixels in the experiment.

Algorithm 4 Registration Evaluation

```
Input: Region of interest from the image $I_{ROI}$
Output: A Quality Index of Registration $Q_R$
1: Select region of interest near the checkerboard
2: Create polygons $P_{r,l}$ with Edges in the $ROI$
3: Calculate $Q_R$ based on $P_r \cap P_l$
```

The polygons $P_{l,r}$ from Algorithm 4 consist of the upper and lower limits of the region of interest and the left and right margins of the steel section. After $P_{l,r}$ polygons are created, the quality of registration index Q_R is calculated by evaluating the overlapped area between two polygons. The smaller the Q_R values are, the better the image registration is.

Figure 7a shows the overlaid registered image and the original left camera image. The left camera image is shaded in green and the registered image is shaded in magenta both on a) and b). The regions with light gray color represent the overlapped areas from the two images. On b) the detected polygons P_r and P_l are also visualised. The overlapped area between two polygons over the polygon for left image P_l form the quality index Q_R of registration and this is described with the equation:

$$Q_R = 1 - \frac{P_r \cap P_l}{P_l}. \tag{1}$$

In order to give a direction to this quality of registration index, we consider Q_R to be positive when P_l is on the left of P_r and vice versa.

The following Section "Embedding the Height Information in the Registration Results" describes the how the prior height information can incorporated to further improve the image registration results.

Embedding the height information in the registration results

The accuracy of the registered image can be further improved by adjusting the height of the checkerboards from the ground. Using real checkerboards with different heights would require collecting many data. Since this is quite a demanding process, a virtual checkerboard is created by interpolation and extrapolation, which only needs to collect the checkerboard data twice at the beginning equipment setting process.

When a camera captures an image, the checkerboard is projected from the three-dimensional to the two-dimensional

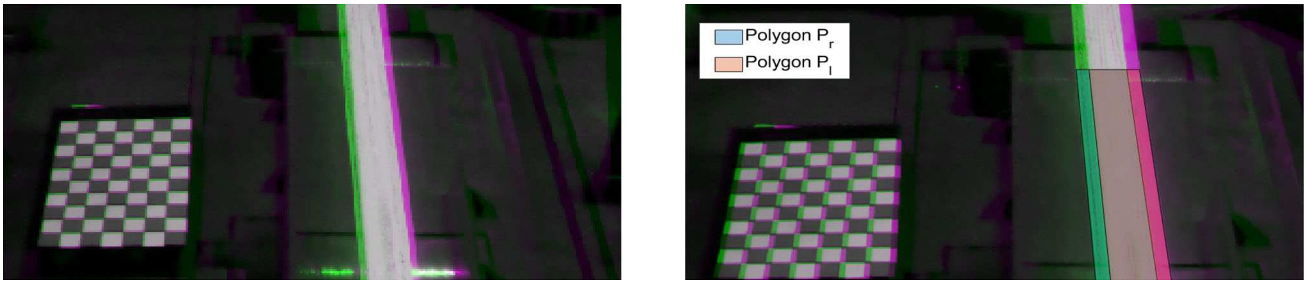
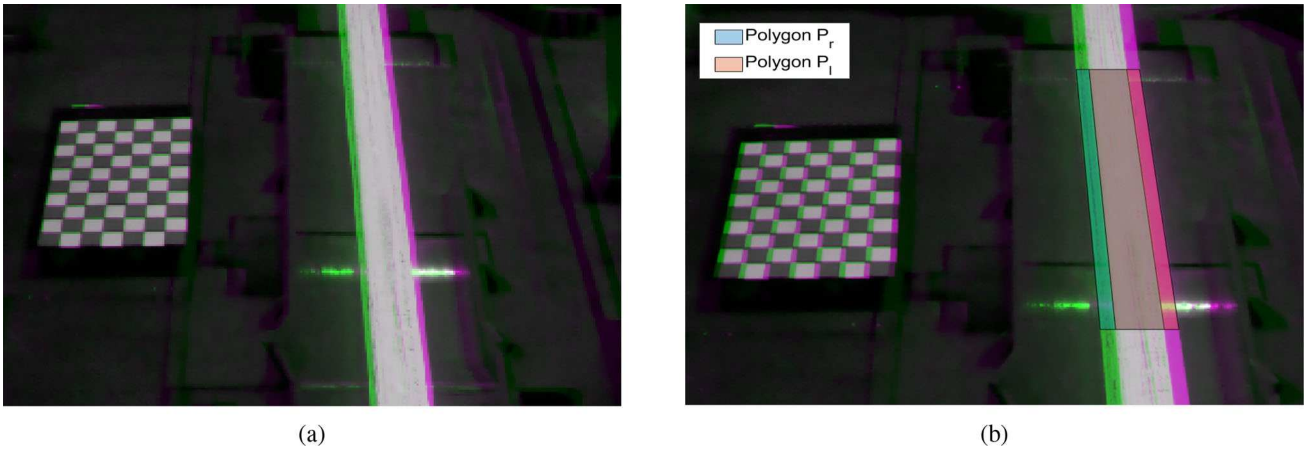


Fig. 6 Sample images for ingot direction recognition. First row shows ingot top samples, and the second row shows ingot bottom samples



(a)

(b)

Fig. 7 Input Steel Section Images and DWT Fusion Results: **a** Registered Images; **b** Polygons P_r and P_l

image plane and the definite ratio point of division between different height checkerboards is also a linear projection in the image plane according to the following formula about the steel section

$$\mathbb{P}\left(\frac{(x_1, y_1) + \lambda\,(x_2, y_2)}{1+\lambda}\right) = \frac{\mathbb{P}\,(x_1, y_1) + \mathbb{P}\lambda\,(x_2, y_2)}{1+\lambda}, \quad (2)$$

where $\mathbb{P}$ is the projection transformation, $\lambda \in \mathbb{N}^+$ is the ratio of division, x_1, y_1 are the coordinates of corner points of lower checkerboard, x_2, y_2 are the coordinates of corner points of higher checkerboard.

Therefore, the interpolation and extrapolation process can be realized by directly inserting and extending data points between the checkerboard's corresponding points at different heights. The interpolation process is applied to improve the registration process results and it is described in Algorithm 5. After the interpolation, combined with the previous registration quality index Q_R, the registration result is updated automatically.

The Q_R index serves as an indicator for the positioning of the virtual checkerboard relative to the measurement plane. Positive values of the Q_R index suggest that the virtual checkerboard is positioned too high, requiring the algorithm to adjust its coordinates according to the height of the steel section. Conversely, negative values imply that the virtual checkerboard is too low in relation to the measurement plane. Ideally, for optimal positioning, the Q_R value should be close to 0.

Algorithm 5 Refining the Image Registration Result with the Virtual Checkerboard

```
Input: $C_{LL}, C_{LH}, C_{RL}, C_{RH}$
Output: The adjust registered image $I_{Rr}$
Interpolate $n$ set of corner points of checkerboards $C_{Li}$,$C_{Ri}$
for $k = 1, \cdots, n$ do
    Calculate $Q_Ri$ for $C_{Li}$,$C_{Ri}$
end for
Find the $I_{Rr}$ with maximum $|Q_R|$
```

The next Section "Image Fusion" describes how the two registered images from the left and right cameras can be fused with different wavelet transform methods.

Image fusion

Thanks to the virtual checkerboard and image registration algorithm, the images from the two cameras are registered with high accuracy. After the image registration, the positions of the steel section parts taken by the left and right cameras in the image can completely coincide. Next, the images taken

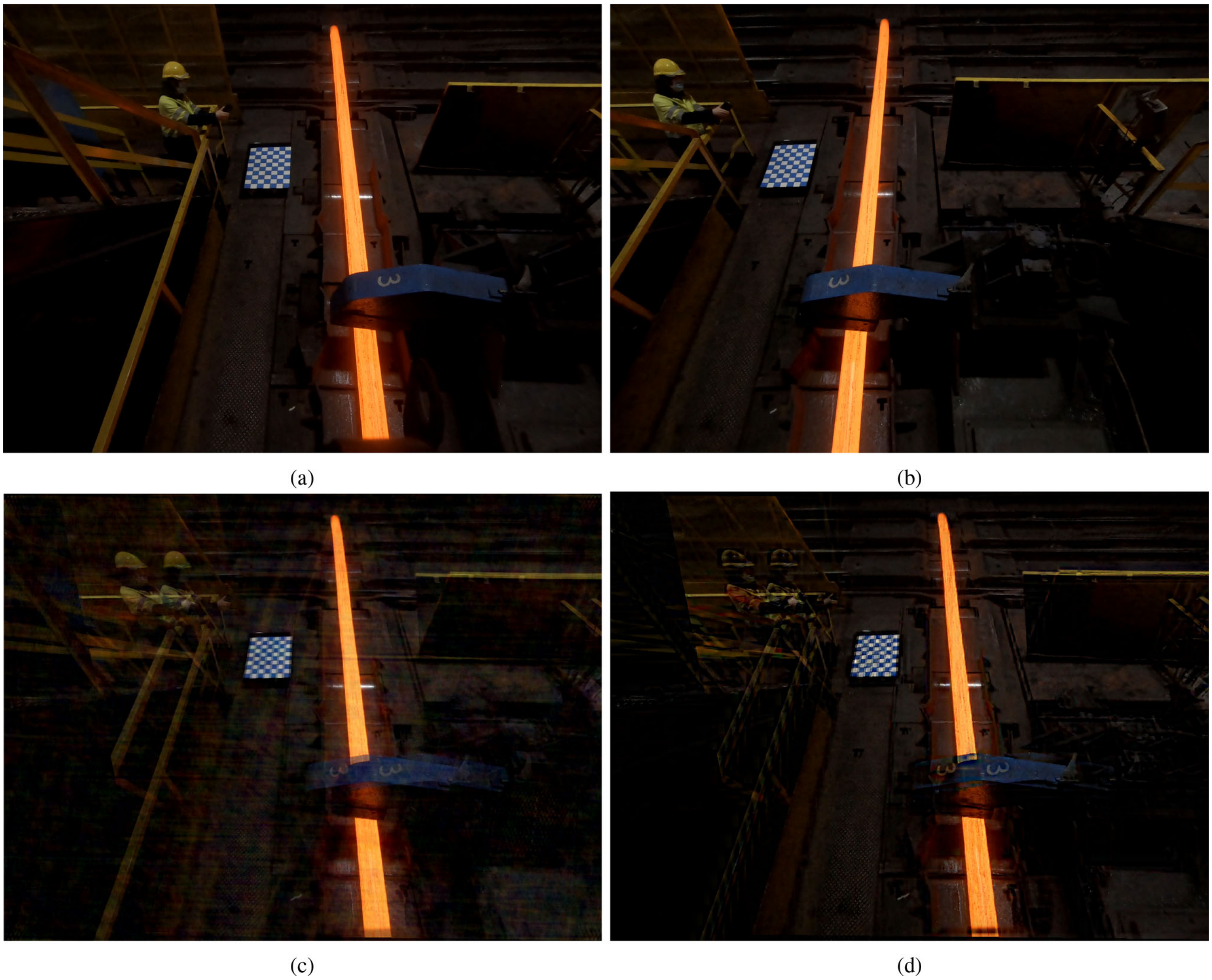


Fig. 8 Input Steel Section Images and DWT Fusion Results: **a** Left Camera Image; **b** Right Camera Image; **c** FFT Fusion Results; **d** DWT Fusion Results

by two cameras are fused in order to bring together the complementary information from the two separate images. From the fused image the steel section size is calculated remotely - first in the image plane and next the result is converted into the physical plane. Image fusion algorithms based on a Fast Fourier Transform (FFT) (Cooley and Tukey, 1965) and discrete wavelet transforms (DWT) (Pajares and de la Cruz, 2004; Sundararajan, 2016) are implemented and validated over video data collected from the Liberty Speciality factory in the UK (Fig. 8).

FFT

The left and right images are first transformed in the frequency domain with the discrete FFT (Gao et al., 2021). Then, the fusion step is completed by operating on the magnitude and phase map of the images in the frequency domain. The two dimensional (2D) FFT $F[k,l]$ of an image $I[m,n]$ of size $m \times n$ is given by:

$$F[k,l] = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I[m,n] e^{-j2\pi\left(\frac{k}{M}m + \frac{l}{N}n\right)}. \tag{3}$$

The result of the Fourier transform $F[k,l]$ can be expressed by the amplitude $\|F[k,l]\|$ and phase $\angle F[k,l]$, M and N characterise the respective image size. The FFT process of fusing two images I_1 and I_2 in the frequency domain is given as Algorithm 6.

DWT-Daubechies Wavelets

Discrete wavelet transforms (DWTs) can provide efficient image fusion results (Pajares and de la Cruz, 2004) and we adopt the one-dimensional DWT based on the selected

Algorithm 6 FFT Image Fusion of Images I_1 and I_2

```
Input: $F_{I1}[k,l]$, $F_{I2}[k,l]$
Output: The fused image $I_F$
Calculate the magnitude $\|F[k,l]\|$ and phase $\angle F[k,l]$
for $k = 1, \cdots, M$ do
    for $l = 1, \cdots, N$ do
        if $\|F_{I1}[k,l]\| > \|F_{I2}[k,l]\|$ then
            $\|F[k,l]\| = \|F_{I1}[k,l]\|$
            $\angle F[k,l] = \angle F_{I1}[k,l]$
        else
            $\|F[k,l]\| = \|F_{I2}[k,l]\|$
            $\angle F[k,l] = \angle F_{I2}[k,l]$
        end if
    end for
end for
Inverse FFT $F[k,l] \rightarrow I_F$
```

wavelet basis in the x and y directions of the image to achieve a two-dimensional DWT. The selection of different wavelet bases will lead to different fusion effects. Daubechis (DB) wavelets (Daubechies, 1996; Hermessi et al., 2021) are first tested with different wavelet coefficients. The DWT method results are presented with 2, 4 and 16 coefficients and this is denoted as DB2, DB4, DB8 and DB16, respectively.

DWT-Fejer-Korovkin wavelets

In addition to DB wavelets, Fejer-Korovkin wavelets (FK) wavelets transform are applied in DWT fusion. The FK wavelets are denoted as FK4, FK6, FK8 and FK18, respectively, according to different filter coefficients. The FK wavelets have shown better high-frequency performance than other waveforms (Olhede and Walden, 2004).

When the image fusion process is completed, the steel section edges from the fused image are identified and extracted. The steel section pixel size can then be measured. Combined with the camera parameters obtained during the camera calibration process, the pixel size of the steel section can be converted to the actual steel width in physical units.

Performance validation and evaluation

The proposed framework has been validated over real images from the Liberty Speciality Steels industrial plant in the UK producing high-quality steels.

Ingot direction recognition

Through the above section extraction technology, a total of 2220 end-section images of ingots were extracted from 9 bottom videos and 9 top videos. The results are presented in Table 1.

Figure 10 shows the confusion matrix for the proposed SSIM classifier that distinguishes the top from the end part of the steel sections. Table 1 shows the precision, recall and accuracy performance measures for the classification results of the SSIM classifier, respectively. For the top side of the steel sections, based on 1329 detected end sections from 20 videos, the evaluated value of the precision is equal to 1, the calculated recall rate is 0.9589, and the overall classification accuracy is 0.9819. Figure 9 shows the results for the SSIM of the top ingot side and that the $SSIM_{top}$ is larger than $SSIM_{bot}$ in most cases. Hence, the classifier detects the top end sides with a high success rate. Out of a total of 1329 images corresponding to the end side of the ingot, 24 are incorrectly classified which corresponds to 1.8% misclassification. The analysis of the videos shows that these 24 misclassifications occur in images in which the end side of the ingot is not well visible. Due to the rising temperature, the brightness of the ingot in the images increases, leading to significant changes in the SSIM values. The results show that the 'Top Side' of the ingot has been successfully recognised, and an early warning can be given to the human operator.

Table 1 SSIM classification results

	Precision	Recall	Accuracy
Top Sides	1	0.9589	0.9819
Bottom Sides	0.9688	1	

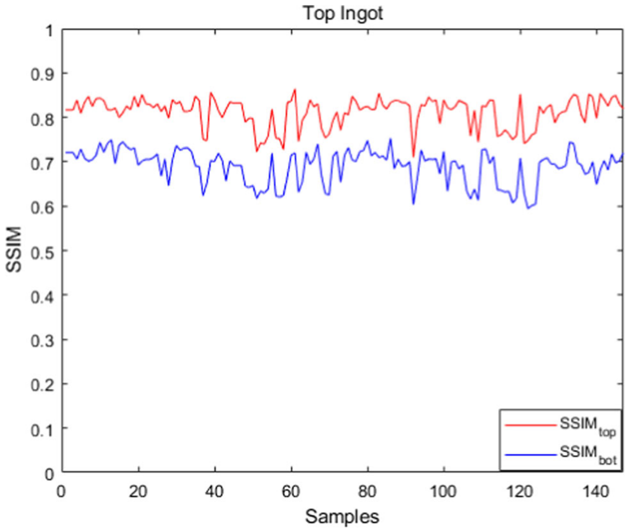


Fig. 9 Top side results for the SSIM

Remote calculation of the size of the steel hot rolling sections

The whole sizing process is shown in the flow chart presented in Fig. 11. At the beginning of the process, we have images of the steel section taken simultaneously by two cameras in the left and right directions. The two cameras

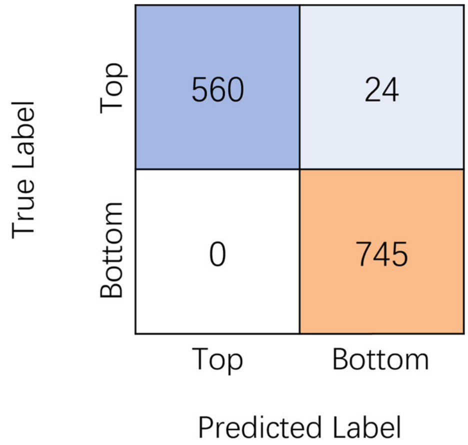


Fig. 10 Confusion matrix of the SSIM classifier distinguishing the top and bottom parts of the steel section

are focused on the area of the movement of the steel section but also incorporate different views of the industrial plant. Next, the two images are registered with the approach described in the previous sections, and the virtual checkerboard corresponding to the measurement plane of the steel section is generated as described in Section "Image registration evaluation". After that, the whole image is transformed by computing the geometric transformation between the virtual and the initial checkerboards. Using this algorithm, we can make the measurement plane coincide with the initial checkerboard's height.

As Fig. 12 shows that when the checkerboard is set on the ground, the cameras use the calibration data on a different plane from the measurement plane. Therefore, the measured results will be larger than the actual results in this case. In order to correct this measurement error, the position of the checkerboard should be raised to the height consistent with the steel radius to make the height of the measurement plane compatible with the steel radius. Through the virtual calibration plate, the calibration plate's height can be freely moved as shown in Fig. 12 (c). Through the measured value via Algorithm 5, the height of the virtual calibration plate can also be evaluated, as shown in Fig. 12 (b).

After the image registration is completed, the images taken by the two cameras are registered together. Knowing the camera internal and external calibration parameters and using the algorithm from Wang et al. (2019), the pixel size of a steel section can be transformed to its actual physical size.

In the considered steel production case study, the two edges of a hot rolling bar (HRB) are scanned by a sliding window $I_{H\times W}$, with height H and width W. This process smooths the edge selection process and improves the edge detection accuracy (Wang et al., 2020). The transformation equation

$$\begin{bmatrix} x_{ij}^{w} \\ y_{ij}^{w} \\ 1 \end{bmatrix} = \mathbf{H}^{-1} \begin{bmatrix} x_{hj}^{I} \\ y_{hj}^{I} \\ 1 \end{bmatrix}, \mathbf{H} = [\mathbf{K}]\,[\mathbf{R}|\mathbf{T}] \tag{4}$$

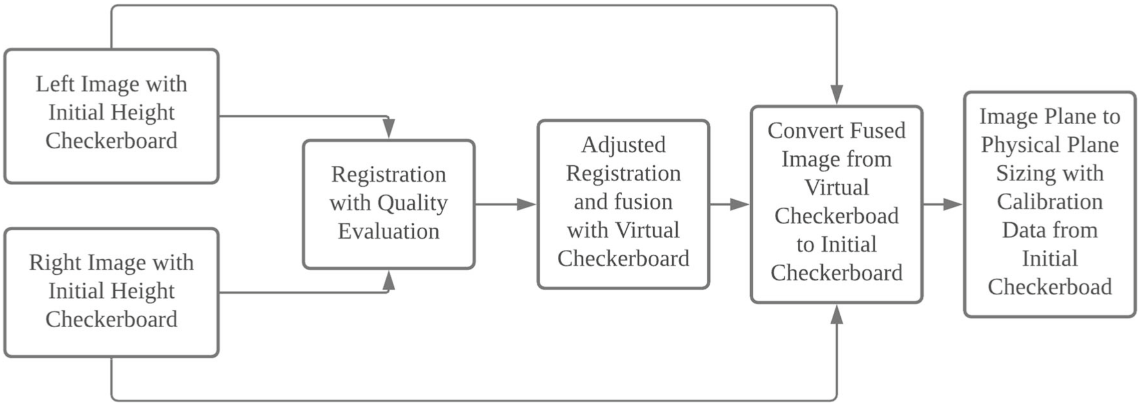

Left Image with Initial Height Checkerboard
Right Image with Initial Height Checkerboard
Registration with Quality Evaluation
Adjusted Registration and fusion with Virtual Checkerboard
Convert Fused Image from Virtual Checkerboard to Initial Checkerboard
Image Plane to Physical Plane Sizing with Calibration Data from Initial Checkerboad

Fig. 11 Flow Chart of Sizing Process

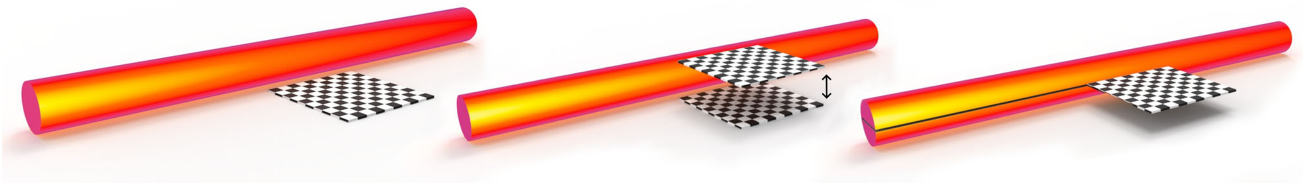


Fig. 12 Virtual Checkerboard for Image Registration: The Leftmost Image Shows Image Captured when the Checkerboard is on the Floor; The Middle Image Shows the Checkerboard on Another Height; The Rightmost Image Shows the Virtual Checkerboard at the Desired Height

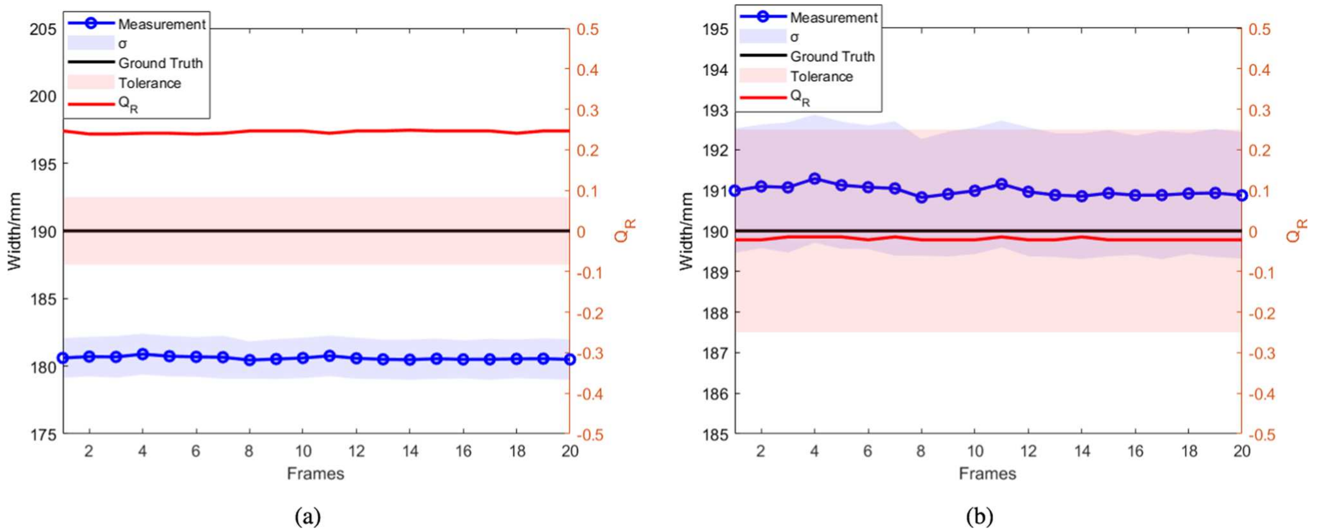


Fig. 13 Sizing Results: **a** Q_R is positive and the absolute value $|Q_R|$ is large; **b** Q_R is negative and $|Q_R|$ is small

Table 2 RMSE of Fig.13a, b

	a	b
RMSE	9.3741	0.9872

converts the x_{ij}^I and y_{ij}^I image plane coordinates into the $[x_{ij}^w, y_{ij}^w, 1]$ physical plane coordinates. The transformation is performed with the help of the rotation matrix **R**, of the translation matrix **T** and the **K** matrix containing the intrinsic camera parameters. The specific values of **R**, **T** and **K** can be calculated through the calibration process of the GoPro® cameras. The calibration was performed using the Camera Calibration Toolbox in MATLAB® (Bouguet, 2004). The cameras used for capturing the video had a resolution of 2704 x 2028 and a frame rate of 30 FPS. They employed a linear Field of View (FOV) mode with a shutter speed of 1/480 s.

Given the vectors $I_{i1} = [x_{i1}^I, y_{i1}^I]^T$ and $I_{i2} = [x_{i2}^I, y_{i2}^I]^T$ on two HRB edges with $x_{i1}^I = x_{i2}^I$, the diameter l of the HRB is then calculated through

$$l = \|P_1 - P_2\|_2\,, \tag{5}$$

where $P_1 = [x_{i1}^w, y_{i1}^w]^T$ and $P_2 = [x_{i2}^w, y_{i2}^w]^T$ are the physical plane correspondences to I_{i1} and I_{i2}. Here $\|.\|_2$ denotes the Euclidean norm.

Figure 14 shows seven measurement results taken every 100 frames in the same video sequence. The actual diameter of the hot steel section is equal to 190 mm. The blue data points represent measurements obtained with the virtual checkerboards and the red data points are direct estimates without using the virtual checkerboard. Figure 13a and 13b show the sizing results with different checkerboard parameters and how $|Q_R|$ evaluates the sizing quality. In Fig. 13a, the sizing results are underestimated. The Q_R values are large and positive, showing that the true size should be much larger than the estimation. In Fig. 13b, a small negative Q_R shows the true size is slightly smaller than the estimated value.

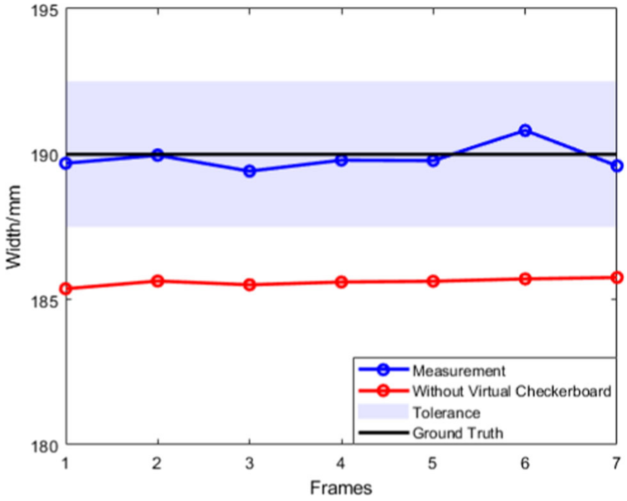


Fig. 14 Sizing Results for Seven Different Frames

Image fusion performance metrics and analysis of the results

Several performance assessment metrics (Hermessi et al., 2021; James and Dasarathy, 2014) that do not rely on reference images are utilised to evaluate the quality of fused images and these are given below.

Information entropy *H* (Singh and Anand, 2018)

$$H\left(I\right)=-\sum P\log_2 P, \tag{6}$$

where P is the normalized histogram and I denotes the considered image. Information entropy (Hermessi et al., 2021) characterises the amount of information contained in an image. The higher the information entropy, the more information the image contains. The unit of entropy when calculated for the images is $bit/pixel$.

Standard Deviation *SD* (Haghighat et al., 2011)

$$SD=\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I\left(i,j\right)-\bar{\mu}\right)^2/MN}, \tag{7}$$

where I is the intensity of pixel and it is within the range $[0, 255]$, $\bar{\mu}$ is the average image intensity over the considered number of pixels, M and N represents the image size with respect to the two coordinate axes. The standard deviation of an image represents the variation of the pixel brightness. The larger the standard deviation is, the more obvious the brightness difference between image pixels is, and the greater the contrast is.

Spatial Frequency *SF* (Eskicioglu and Fisher, 1995)

$$SF=\sqrt{(RF)^2+(CF)^2}, \tag{8}$$

where RF is the row frequency and CF is the column frequency. The spatial frequency SF combines both frequencies calculated in columns and rows.

Average Gradient *AG* (Prewitt, 1970)

$$AG=\frac{1}{\left(M-1\right)\left(N-1\right)}\sum_x\sum_y\frac{G\left(x,y\right)}{\sqrt{2}}, \tag{9}$$

where M and N are size of image, G is the gradient magnitude of image pixels calculated by Prewiit operator. The first-order difference between the intensity of a pixel and its adjacent pixels reflects the brightness changes of the image. Images with a high average gradient will be clearer than images with small gradient values.

Feature Mutual Information *FMI* (Haghighat et al., 2011)

$$FMI=MI_{FA}+MI_{FB}, \tag{10}$$

where MI_{FA} is the mutual information between Image A and the fused image F, MI_{FB} is the mutual information between image B and image F. The FMI evaluates the dependency between the input images and the fused image. A large FMI means that the fused image contains more information from image A and B.

Sum of the Correlation of Differences *SCD* (Aslantas and Bendes, 2015)

$$SCD=r(D_1,I_1)+r(D_2,I_2), \tag{11}$$

where D_i is the difference between the input image $I_i, i=1,2$ and the fused image, $r\left(.\right)$ denotes the correlation function.

Edge-based Structural Similarity *ESSIM* (Chen et al., 2006)

$$ESSIM=function\left(l\left(I_1,I_2\right),c\left(I_1,I_2\right),e\left(I_1,I2\right)\right), \tag{12}$$

$$N_{essim}(I_f)=\left(1-\frac{ESSIM(I_1,I_f)+ESSIM(I_2,I_f)}{2}\right)\times 1000, \tag{13}$$

where $l(I_1,I_2)$ is a function characterising the luminance difference between images I_1 and I_2, $c(I_1,I_2)$ is a function for the contrast comparison and $e(I_1,I_2)$ is the function for the edge comparison between the two considered images. Compared with the original SSIM, $ESSIM$ uses edge comparison to replace the original structural comparison. This makes the metrics more sensitive to the edge information, which is more critical for the steel sizing algorithm. Here, we consider the $ESSIM$ between fused image I_f and original images I_1, I_2 as shown in equation (13). The lower the N_{essim} value, the better the fusion quality is.

Table 3 presents average results over 4 videos, each containing 100 frames. It also includes results for the fused images from both cameras with different methods, evaluated with different performance metrics. It is evident that the FFT and DWT with high-order coefficients give better results than the same approaches with low-order coefficients. The FFT gives best results according to the following criteria: SD, FMI, SCD and $ESSIM$.

According to the fusion performance metrics given in Table 3, the DWT results achieved with a 16-coefficient Daubechies wavelet transform have the highest information entropy, which shows that it contains the most information in all fusion results. At the same time, according to the indicators related to image contrast and edge sharpness (spatial

Table 3 Image Fusion Performance Evaluation Results

Metrics	H	SD	SF	AG	FMI	SCD	ESSIM
Left Image	1.4805	53.9503	6.5669	4.3598	–	–	–
Right Image	1.5052	54.3409	6.3998	4.0942	–	–	–
FFT	1.5056	**56.4264**	6.8742	4.8237	**0.9663**	**1.3364**	**0.1458**
DWT-DB Wavelets 2	1.5334	53.3971	7.0793	5.4118	0.9610	0.7495	0.1792
DWT-DB Wavelets 4	1.5492	53.1009	7.0279	5.3707	0.9622	0.7502	0.1791
DWT-DB Wavelets 8	1.5579	52.9425	6.9745	5.3455	0.9609	0.7370	0.1794
DWT-DB Wavelets 16	**1.5738**	52.6410	6.9257	5.3080	0.9593	0.7375	0.1799
DWT-FK Wavelets 4	1.5185	53.5126	**7.1620**	**5.4216**	0.9612	0.7606	0.1712
DWT-FK Wavelets 6	1.5392	53.2564	6.9780	5.3429	0.9626	0.7485	0.1795
DWT-FK Wavelets 8	1.5455	53.1555	7.0109	5.3871	0.9624	0.7548	0.1803
DWT-FK Wavelets 18	1.5628	52.9686	6.9345	5.3159	0.9606	0.7486	0.1780

frequency and average gradient), DWT fusion using a 4-coefficient Fejer-Korovkin wavelet showed the best results. For the remaining metrics, the FFT image fusion results show the best performance. Overall, the FFT fused image contains more information than the original images. The DWT fusion with an FK4 wavelet gives results with high contrast, which benefits the edge extraction process.

In summary, the proposed approach has three main key elements that are essential for the high-accuracy sizing process.

Firstly, the proposed accurate edge detection algorithm plays a pivotal role for identifying well the edges of the steel sections within the images. This algorithm is specifically designed to be highly sensitive to edges, thereby ensuring that the true boundaries of the steel sections are captured, even amidst significant noise.

Secondly, the virtual checkerboard serves as an innovative tool that provides high-precision scale conversion from the image plane to the physical plane. This allows us to accurately map the dimensions of the image to the corresponding real-world dimensions. This step is critical, as it ensures that the representation in the image reflects accurately the actual sizes of the steel sections. Furthermore, the external features from the virtual checkerboard aid in aligning the images, which results in high registration accuracy. By adopting this approach to create consistent calibration features across different images from cameras, we achieve precise alignment among the images, a critical prerequisite for subsequent image fusion and measurement procedures.

Finally, the image fusion method enhances further the edges and significantly reduces the impact of image noise from the measurements. By combining the images in a manner that maximises the information content and edge clarity, we are able to generate a final image that is both clear and accurate, despite the presence of image noises.

In conclusion, the high accuracy and robustness of our framework can be attributed to the combination of precise edge detection, accurate image registration with the virtual checkerboard, and effective image fusion.

Conclusions

This paper presents a new framework for image registration, fusion, sizing, and object recognition. It includes a two-camera system that collects optical images of moving hot steel sections. The paper also details a proposed recognition algorithm that introduces autonomy, incorporates the structural similarity measure, and ensures the correct placement of the hot steel sections.

The developed image registration approach embeds extrinsic features using a virtual checkerboard, making it adaptive to environmental changes. It also helps in including the height information at which the two images from the left and right cameras are collected. The incorporation of the checkerboard also helps cope with the challenges due to missing features in the steel sections.

Efficient image fusion with Daubechies wavelets, the Fast Fourier Transform, and Fejer-Korovkin wavelets algorithms is achieved with the registered images. A series of performance metrics evaluate the quality of the fused image from the perspectives of information content and edge clarity. The information content is evaluated with metrics such as entropy, standard deviation, image feature mutual information, sum of the correlation of differences, and edge-based structural similarity. The DWT-DB wavelets 16 create fused images with the largest entropy, but fast Fourier transform fusion gives better results among other metrics.

The image edge clarity is evaluated with the spatial frequency and average gradient criteria. The DWT-FK Wavelets 4 fusion gives the best results in spatial frequency and average gradient metrics.

The performance of the system is evaluated on various real data with different metrics. We have shown that high precision sizing results with a tolerance range of less than $2mm$ are achieved, which helps with quality assurance of manufacturing tasks. The achieved remote sizing accuracy is above $95\%(\pm 2mm\ /\ 195mm)$ thanks to the efficient registration approach with extrinsic image features combined with accurate image fusion algorithms.

The evaluation of the proposed approach on real data suggests that it can achieve high precision sizing results. However, its performance may vary depending on the type of steel sections or environmental conditions. Moreover, in the production environment, camera position deviation may be affected by factors such as vibration, which requires regular maintenance and recalibration of the system to ensure accurate measurement results.

The recognition task utilises only two images as reference, while the remaining data is used for validation purposes. Therefore, increasing the size of the dataset does not directly impact the accuracy of the recognition model. However, if we incorporate image fusion techniques to combine features from multiple images as reference, having a larger dataset may potentially lead to further improvement of the results.

For the remote steel section sizing task, the dataset is not involved in the modeling process but rather used for result validation. Therefore, the quantity of the data does not significantly impact the effectiveness of the method.

Future work includes the development of efficient image segmentation deep learning methods for pre-segmentation of regions of interest of target objects to improve the system's robustness against interference. In addition, semantic segmentation would provide another efficient solution for measuring multiple objects in an image or different surfaces of a complex object. Theoretical quantification of the impact of different uncertainties such as measurement noises, occlusions, environmental and other conditions on the final solutions is another area of current research.

Acknowledgements We would like to thank our industrial collaborators from the Liberty Speciality Steels and especially Simon Pike, Ree Muroiwa, David Robinson, Bing Huo, Martyn Doncaster and Craig Grimstone for the collaboration via the joint projects that we had, for the inspiring discussions and data provided. We are grateful to our sponsors through the Knowledge Exchange grant (Internet of Things for Overcome Barriers in the Steel Rolling Measurement Technology) with the Liberty Speciality Steels, the UK EPSRC IAA grant R/164995-11-1. We are grateful to the EPSRC for funding this work through EP/T013265/1 project NSF-EPSRC: ShiRAS Towards Safe and Reliable Autonomy in Sensor Driven Systems which is a joint grant with the USA Grant NSF ECCS 1903466. We also thank China NSFC (grant 61703387). 'This work was partially funded by EPSRC project EP/T024291/1 Autonomous method for detecting cutting tool and machine tool anomalies in machining. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Author contributions Y. L.: conceptualisation, methodology, software, validation, data collection, visualisation, data analysis, writing the first draft of the paper. P. W.: conceptualisation, methodology, software, validation, data collection, analysis, writing and correcting the first paper draft and corrections. Z. W.: conceptualisation, software, results analysis, S. A.: conceptualisation, software, results analysis, L. M.: conceptualisation, methodology, writing and correcting the first paper draft, analysis, supervision, securing funding, project management

Data availability There is no open data available at the moment.

Declarations

Conflict of interests There is no conflict of interests.

References

Aslantas, V., & Bendes, E. (2015). A new image quality metric for image fusion: The sum of the correlations of differences. *AEU-International Journal of Electronics and Communications, 69*(12), 1890–1896.

Barari, A., de Sales Guerra Tsuzuki, M., Cohen, Y., & Macchi, M. (2021). Intelligent manufacturing systems towards industry 4.0 era. *Journal of Intelligent Manufacturing, 32*(7), 1793–1796.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features. In Proceedings of the European Conference on Computer Vision. Springer, pp. 404–417.

Bouguet, J.-Y. (2004). Camera calibration toolbox for MATLAB. http://www.vision.caltech.edu/bouguetj/calib_doc/index.html.

Boveiri, H. R., Khayami, R., Javidan, R., & Mehdizadeh, A. R. (2020). Medical image registration using deep neural networks: A comprehensive review. *Computers and Electrical Engineering, 87*, 106767.

Brock, K. K., Mutic, S., McNutt, T. R., Li, H., & Kessler, M. L. (2017). Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the aapm radiation therapy committee task group no. 132. *Medical Physics,44*(7), e43–e76.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,PAMI–8*(6), 679–698.

Chen, G.-H., Yang, C.-L., Po, L.-M., & Xie, S.-L. (2006). Edge-based structural similarity for image quality assessment. In Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, vol. 2. IEEE, pp. II–II

Chen, X., Diaz-Pinto, A., Ravikumar, N., & Frangi, A. F. (2021). Deep learning in medical image registration. *Progress in Biomedical Engineering, 3*(1), 012003.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation, 19*, 297–301.

Czolbe, S., Krause, O., & Feragen, A. (2021). Semantic similarity metrics for learned image registration. In Medical Imaging with Deep Learning. PMLR, pp. 105–118.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1. IEEE, 886–893

Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2019). A survey of deep learning and its applications: A new paradigm to machine learning. *Archives of Computational Methods in Engineering*, pp. 1–22.

Daubechies, I. (1996). Where do wavelets come from? a personal point of view. *Proceedings of the IEEE, 84*(4), 510–513.

Dollár, P., Singh, M., & Girshick, R. B. (2021). "Fast and accurate model scaling," In Proceedings of CVPR. Computer Vision Foundation / IEEE, pp. 924–932.

Dollár, P., & Zitnick, C. L. (2014). Fast edge detection using structured forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(8), 1558–1570.

Eliopoulos, N. J., Shen, Y., Nguyen, M. L., Arora, V., Zhang, Y., Shao, G., Woeste, K. E., & Lu, Y.-H. (2020). Rapid tree diameter computation with terrestrial stereoscopic photogrammetry. *Journal of Forestry*.

Eskicioglu, A., & Fisher, P. (1995). Image quality measures and their performance. *IEEE Transactions on Communications, 43*, 2959–2965.

Gao, Y., Li, X., Wang, X. V., Wang, L., & Gao, L. (2021). "A review on recent advances in vision-based defect recognition towards industrial intelligence," Journal of Manufacturing Systems.

Guan, S., Wang, T., Meng, C., & Wang, J. (2018). A review of point feature based medical image registration. *Chinese Journal of Mechanical Engineering, 31*, 1–16.

Haghighat, M. B. A., Aghagolzadeh, A., & Seyedarabi, H. (2011). A non-reference image fusion metric based on mutual information of image features. *Computers & Electrical Engineering, 37*(5), 744–756.

Hall, D. L., & Llinas, J. (2001). *Handbook of Multisensor Data Fusion* (1st ed.). USA: CRC Press.

Hermessi, H., Mourali, O., & Zagrouba, E. (2021). Multimodal medical image fusion review: Theoretical background and recent advances. *Signal Process, 183*, 108036.

Hoyos, C. G., & Zimolong, B. (2014). *Occupational safety and accident prevention: behavioral strategies and methods*. Netherlands: Elsevier.

James, A. P., & Dasarathy, B. V. (2014). Medical image fusion: A survey of the state of the art. *Information Fusion, 19*, 4–19.

Javaid, M., Haleem, A., Singh, R. P., Rab, S., & Suman, R. (2022). Exploring impact and features of machine vision for progressive industry 4.0 culture. *Sensors International,3*, 100132.

Jin, X., Jiang, Q., Yao, S., Zhou, D., Nie, R., Hai, J., & He, K. (2017). A survey of infrared and visual image fusion methods. *Infrared Physics & Technology, 85*, 478–501.

Khan, S. S., Khan, M., Alharbi, Y., Haider, U., Ullah, K., & Haider, S. (2021). Hybrid sharpening transformation approach for multifocus image fusion using medical and nonmedical images. *Journal of Healthcare Engineering, 2021*(7000991), 17.

Kuo, R., & Nursyahid, F. F. (2022). Foreign objects detection using deep learning techniques for graphic card assembly line. *Journal of Intelligent Manufacturing*, pp. 1–12.

Leng, J., Sha, W., Wang, B., Zheng, P., Zhuang, C., Liu, Q., Wuest, T., Mourtzis, D., & Wang, L. (2022). Industry 5.0: Prospect and retrospect. *Journal of Manufacturing Systems, 65*, 279–295.

Lin, Y., Wang, P., Muroiwa, R., Pike, S., & Mihaylova, L. (2021). mage fusion for remote sizing of hot high quality steel sections. In Proceedings of the 20th UK Workshop on Computational Intelligence (UKCI), Aberystwyth University, Penglais, Aberystwyth, Wales, UK, Sept. 8-10.

Lu, Y., Morris, K. C., Frechette, S., et al. (2016). Current standards landscape for smart manufacturing systems. National Institute of Standards and Technology, NISTIR, *8107*(3)

Luo, Y., Ren, J., Lin, M., Pang, J., Sun, W., Li, H., & Lin, L. (2018). "Single view stereo matching," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 155–163.

Luo, Q., & He, Y. (2016). A cost-effective and automatic surface defect inspection system for hot-rolled flat steel. *Robotics and Computer-Integrated Manufacturing, 38*, 16–30.

Ma, J., Ma, Y., & Li, C. (2019). Infrared and visible image fusion methods and applications: A survey. *Inf. Fusion, 45*, 153–178.

Meindl, B., Ayala, N. F., Mendonça, J., & Frank, A. G. (2021). The four smarts of industry 4.0: Evolution of ten years of research and future perspectives. *Technological Forecasting and Social Change,168*, 120784.

Mutlag, W. K., Ali, S. K., Aydam, Z. M., & Taher, B. H. (2020). Feature extraction methods: A review. *Journal of Physics: Conference Series, 1591*(1), 012028.

Nixon, M. S., & Aguado, A. S. (2020). 5 - high-level feature extraction: fixed shape matching. In M. S. Nixon & A. S. Aguado (Eds.), *Feature Extraction and Image Processing for Computer Vision* (pp. 223–290). Academic Press.

Olhede, S. C., & Walden, A. T. (2004). The Hilbert spectrum via wavelet projections. *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences, 460*, 955–975.

Pajares, G., & de la Cruz, J. M. (2004). A wavelet-based image fusion tutorial. *Pattern Recognition, 37*, 1855–1872.

Prewitt, J. M., et al. (1970). Object enhancement and extraction. *Picture processing and Psychopictorics, 10*(1), 15–19.

Putra, B. T. W., Ramadhani, N. J., Soedibyo, D. W., Marhaenanto, B., Indarto, I., & Yualianto, Y. (2021). The use of computer vision to estimate tree diameter and circumference in homogeneous and production forests using a non-contact method. Forest Science and Technology.

Qian, J., Zhang, Z., Shi, L., & Song, D. (2021). An assembly timing planning method based on knowledge and mixed integer linear programming. *Journal of Intelligent Manufacturing,* 1–25.

Singh, S., & Anand, R. S. (2018). Ripplet domain fusion approach for CT and MR medical image information. *Biomedical Signal Processing and Control, 46*, 281–292.

Song, R., Zhang, Z., & Liu, H. (2017). Edge connection based Canny edge detection algorithm. *Pattern Recognition and Image Analysis, 27*, 740–747.

Sundararajan, D. (2016). *Discrete Wavelet Transform: A Signal Processing Approach*. NY: Wiley.

Suzuki, S., et al. (1985). Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing, 30*(1), 32–46.

Tafti, A. P., Baghaie, A., Kirkpatrick, A. B., Holz, J. D., Owen, H. A., D'Souza, R. M., & Yu, Z. (2018). A comparative study on the application of sift, surf, brief and orb for 3D surface reconstruction of electron microscopy images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 6*(1), 17–30.

Tong, X., Ye, Z., Xu, Y., Gao, S., Xie, H., Du, Q., Liu, S., Xu, X., Liu, S., Luan, K., et al. (2019). Image registration with fourier-based image correlation: A comprehensive review of developments and applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12*(10), 4062–4081.

Tong, X., Ye, Z., Xu, Y., Gao, S., Xie, H., Du, Q., Liu, S., Xu, X., Liu, S., Luan, K., & Stilla, U. (2019). Image registration with fourier-based image correlation: A comprehensive review of developments and applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12*(10), 4062–4081.

Wang, P., Lin, Y., Muroiwa, R., Pike, S., & Mihaylova, L. (2019). Computer vision methods for automating high temperature steel section sizing in thermal images. In Proceedings of the Sensor Data Fusion: Trends, Solutions, Applications (SDF). IEEE, pp. 1–6.

Wang, P., Lin, Y., Muroiwa, R., Pike, S., & Mihaylova, L. (2020). A weighted variance approach for uncertainty quantification in high quality steel rolling. In Proceedings of the IEEE 23rd International Conference on Information Fusion (FUSION'23). IEEE, pp. 1–7.

Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing, 13*(4), 600–612.

Zheng, L., Yang, Y., & Tian, Q. (2017). Sift meets cnn: A decade survey of instance retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(5), 1224–1244.

Zhou, Y., Wu, Y., & Luo, C. (2018). A fast dimensional measurement method for large hot forgings based on line reconstruction. *The International Journal of Advanced Manufacturing Technology, 99*(5), 1713–1724.

Zitova, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing, 21*(11), 977–1000.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.