Automatic Inspection of Punched Metal Plate Fasteners on Timber-to-Timber Joints with Image-Based 3D Reconstruction

Chris MICHEL[1,*] and **Markus ULRICH**[1]

[1] *Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany, (chris.michel@kit.edu, markus.ulrich@kit.edu)*

[*] *corresponding author*

Abstract

Roof structures of commercial buildings like supermarkets are often constructed with prefabricated timber trusses connected by punched metal plate fasteners. These fasteners are susceptible to failure due to mishandling during assembly or installation, as well as differing moisture content of the timber in use. Currently, a regular inspection of the joints is performed manually, which is time-consuming and prone to errors. We propose a methodology that performs the inspection automatically with high accuracy. This is achieved by performing an image-based 3D reconstruction, which allows determining deformations and distances of the fasteners to the timber surface. A handheld stereo camera that returns RGB and disparity images is used to acquire multiple frames of a joint. The images are acquired from different positions to minimise gaps and noise in the resulting reconstruction. After segmenting the fastener and timber in the images, they are transformed to 3D point clouds. Coarse and fine registrations estimate the poses of the individual measurements resulting in a combined point cloud. To increase the accuracy, we model the fastener and the timber with different mathematical surface representations and estimate their model parameters. Most fasteners are best described by B-spline surfaces, which are able to approximate local deformations and defects. We evaluate the proposed methodology on two fasteners: one in a laboratory context and the other in a real application scenario. The experiments show that the distance of fastener and timber can typically be estimated with a deviation to a reference measurement in the range of 0.5 mm.

Keywords: 3D reconstruction, stereo camera, B-spline surfaces, metal plate fasteners

Received: 8th December 2024. Revised: 20th February 2025. Accepted: 25th February 2025.

1 Introduction

A roof structure with punched metal plate fasteners on timber-to-timber joints is an inexpensive and efficient construction technique since large elements of the structure can be prefabricated off-site and installed easily. However, there are possible influences during construction, assembly, or in use, which make these type of fasteners susceptible to failure (Bouldin et al., 2014). For example, mishandling of the prefabricated elements during transport or installation could lead to deformations or defects. Even after correct installation, the timber can shrink or swell due to a varying moisture content over time (Paevere et al., 2009). Consequently, gaps between the fastener and the timber can occur or the fastener may partially detach leading to reduced load capacity. For example, a gap of 1 mm can result in a strength loss of over 25 % (Paevere et al., 2009). These possible influences require regular inspection, which is usually performed visually or by manually measuring the distances between fastener and timber. However, a visual inspection is costly, provides only a subjective assessment, and cannot accurately estimate the load capacity. Manually measuring the distance could be used to derive the load capacity but is time consuming.

Therefore, our objective is to develop a methodology for accurately and automatically reconstructing the 3D surfaces of timber joints in order to determine the distances between fastener and timber. Several different sensor types are available for 3D reconstruction (Steger et al., 2018, Chapter 2.5). However, not every sensor is applicable for our task since the measurements and the surrounding environment can be challenging. The sensor must meet the following requirements:

- The sensor must allow a simple measurement process with a handheld device to quickly assess multiple joints, even if they are difficult to access.
- The sensor must provide colour information for an automatic segmentation of fastener and timber.
- The sensor must be suitable for the reconstruction of reflective surfaces.
- The accuracy for the measured 3D points must not exceed 1 mm.

Based on these requirements, we opted for a stereo camera. Other sensors, for example laser scanners, have to be mounted on a tripod and do not reach the required accuracy. Structured light sensors would be much more accurate than a stereo camera, but also have to be mounted on a tripod because of their longer acquisition time. A stereo camera can be used during movement and captures colour images at the same time as disparity maps. Since the camera is not an active sensor, direct reflections from surfaces such as metal can be avoided with external light sources. We use the stereo camera Nerian Ruby (Allied Vision), which should be able to achieve the accuracy requirement with a theoretical depth error of 0.5 mm at a distance of 40 cm.

In this study, we develop an approach to reconstruct the surfaces of a timber joint with a stereo camera. The distance between fastener and timber is estimated from parametric approximations of these surfaces after an image segmentation of the separate materials. We apply the approach in a laboratory context with a comparison to a reference measurement, as well as in a real application at a supermarket's roof structure.

2 Methodology

Distances between the fastener and the timber are estimated by approximating the surfaces of a 3D reconstruction with parametric models. This section first discusses the generation and registration of a 3D reconstruction from stereo images, followed by the representation of the reconstruction's surface by B-splines.

2.1 3D Reconstruction

The stereo camera Nerian Ruby captures images with three sensors: two monochrome sensors in a stereo configuration capture grey value images and an RGB sensor captures colour images. All sensors have a resolution of 1440x1056 pixel at a field of view of 62° by 48°. A frame rate of 8 Hz is achieved at full resolution. Internally, the stereo camera rectifies the monochrome images and applies a variant of Semi-global matching (Hirschmüller, 2005) to produce a disparity map. The matching accuracy and robustness are improved by an internal infrared dot projector, which artificially adds texture to the captured surfaces. Thus, the start of the reconstruction process consists of a colour image, two monochrome images, and a disparity map, as exemplarily shown in Figure 1.

In a first step, the colour image is segmented into three semantic labels for fastener, timber, and background. While the colour of the two materials, timber and metal, is fairly consistent between different trusses and fastener combinations, their reflectivity and texture vary considerably. Additionally, the intensity, contrast, and other image parameters also differ depending on the pose of the camera and the measurement setup. A deep learning approach such as U-net (Ronneberger et al., 2015) can be trained to robustly segment the materials in these conditions. In this work, we train U-net on 450 labelled images from 150 different timber joint samples. We apply data augmentation techniques to artificially increase the variability of the training images leading to a better generalisation ability of the approach. For example, background replacement has a significant impact on the segmentation results' accuracy for unseen images and is especially important for our training images with a mostly consistent background between images.

A single capture is generally not sufficient for creating a dense, accurate, and robust 3D reconstruction due to noise and gaps in the disparity maps. Both noise and gaps mainly result from low surface texture, overexposed image sections, or specular reflections, which deteriorate or prevent the correspondence search during stereo matching. This is especially relevant for reflective surfaces such as the metal plate fastener. However, these effects highly depend on the specific lighting conditions which can easily be varied by changing the pose of the cam-

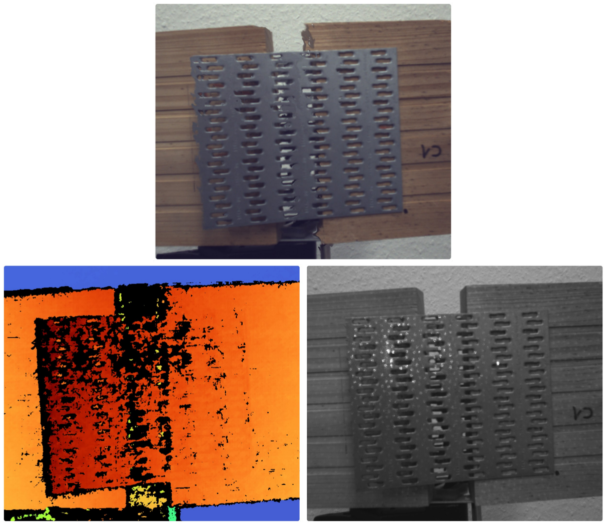

Figure 1. Example of a single capture with colour (top), disparity (bottom left), and monochrome (bottom right) images.

era. Multiple captures from different poses are then combined to minimise gaps and noise over the entire timber joint.

Combining different captures by registration requires the estimation of the individual camera poses. At first, the disparity maps are transformed to 3D point clouds located in the individual camera coordinate systems. The poses are then estimated by registering the point clouds with iterative closest point (ICP), specifically the multiway registration approach implemented in Open3D (Zhou et al., 2018; Choi et al., 2015). Faster and more robust convergence of the ICP iteration are achieved by providing coarse estimates of the poses, in our case by using the image-based SLAM algorithm ORB-SLAM (Campos et al., 2021). Alternatively, approaches for 6D object pose estimation can be used for scene registration (Drost et al., 2010). Furthermore, uncertainty quantification might be helpful to assess the quality of the registration result, and hence to identify registration problems (Wursthorn et al., 2024).

Finally, the result of this approach is a 3D point cloud with colour information and labels for each point.

2.2 Surface Approximation

In order to determine the distance between fastener and timber, the surfaces of all components are estimated by fitting parametric models to the labelled point cloud. The timber surfaces are sufficiently represented by planes. However, each timber element is modelled individually to account for a misalignment during the construction of the joints. The fastener's surface requires a more complex model, which also allows to take deformations of varying magnitude into account. A polynomial surface with a degree of 2 or 3 is sufficient to model most types of deformations. In some cases, the deformations are localised to a section of the fastener, which cannot be accurately represented by polynomial models. Surface models based on B-splines are able to better approximate these local deformations while also accurately representing non-deformed sections of the fastener in a single model.

The B-spline surfaces are constructed by a grid of control points P_{ij} and two one-dimensional cubic basis functions $N_i(u)$ and $N_j(v)$ (Harmening and Neuner, 2015). Knot vectors divide the parameter space into knot spans, which determine the local support of the basis functions. Thus, a point S on the surface results from

$$S(u,v) = \sum_i \sum_j N_i(u) N_j(v) P_{ij} \tag{1}$$

and depends on the parameters u and v in two directions. In an unordered point cloud, the parametrisation for each point is estimated by projecting the points onto a parametric base surface. Most surfaces can be approximated by a Coons patch, which is defined by four boundary curves (Harmening and Neuner, 2015). At first, the boundary points of the surface have to be identified. Since this identification can be complex in 3D space, the point cloud is projected to 2D space. This simplification is only applicable for surfaces with minor curvatures and deformations, which generally is the case for metal plate fasteners. In 2D space, an alpha-shape, which is the generalisation of the convex hull, identifies the border points of the point cloud (Edelsbrunner et al., 1983). The density of the border points is controlled by the alpha parameter. Additionally, the four corner points are identified in order to define the start and end points of the boundary curves. The first two corner points are the border points with maximum distance to each other, while the other two points result from the maximum distances to the line between the first two points.

Four B-spline curves are now estimated from the border points to construct the Coons patch, which serves as a base surface for estimating a first approximation of the parameters. The B-spline surface is estimated with these parameters by least squares. Since the Coons patch is only a rough approximation, the parameters are iteratively improved by using the estimated B-spline surface as the new base surface. As shown by Harmening and Neuner (2015), artefacts can develop at the surface's borders, which are reduced or eliminated by constraining the estimation to the previously determined border splines. The control points of these splines are introduced in the least squares estimation as additional observations with a higher weight than the original points of the point cloud.

The shape of a B-spline surface is predominantly defined by the number of control points. Generally, a surface with a higher number of control points leads to a better approximation of local deformations, but is also more susceptible to noise and can introduce unwanted ripples in planar sections.

3 Results

We evaluate the proposed methodology by applying it to two different timber-to-timber joints. The first joint is a test sample, where the fastener is partially disconnected from the timber, as seen in Figure 1. Since the surface of this sample was also captured with very high accuracy by a laser tracker (Leica AT901 with T-Probe), it can serve as a reference measurement to evaluate the accuracy of the methodology in a laboratory context. In the following, the joint is designated as Sample A. The second joint is part of a supermarket's roof structure and therefore constitutes a real application scenario. In Figure 2, it can be seen that the metal plate fastener of this joint exhibits deformations and is not fully pressed into the timber. Thus, gaps remain between the fastener and the timber with varying degree at different parts of the fastener. In the following, the joint is referred to as Sample B.

3.1 Measurement Setup And Common Processing Steps

For both timber joints, the same basic measurement setup is applied. The stereo camera always faces the timber joint perpendicular with a distance of approximately 40 cm, which is near the minimum

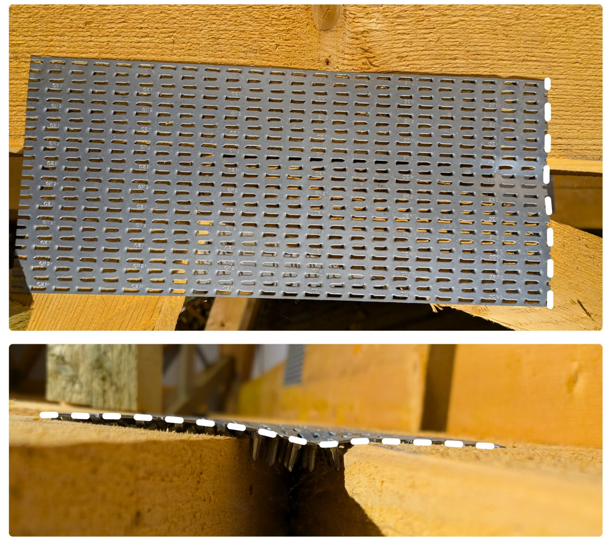

Figure 2. Frontal view (top) and side view (bottom) of a supermarket's roof joint with deformations of the fastener highlighted by the dashed line.

measurable depth of the camera. This ensures the highest precision, but also limits the captured area. Therefore, the camera is moved in a continuous motion from side to side, capturing the entire fastener and parts of the connected timber. At full resolution, the camera achieves a frame rate of 8 Hz, which results in about 30 to 50 images depending on the movement and the timber joint's size. However, a subset of about 10 images evenly distributed over time is usually sufficient and reduces the amount of redundant data for the subsequent processing steps.

Processing of this subset starts with the segmentation of metal plate, timber, and background with the deep learning network U-net. As exemplarily shown on the left of Figure 3, the two materials are easily distinguishable by colour, except for the metal plate's punched holes. Depending on the lighting, the visibility of the timber or the background through these holes varies considerably. Consequently, the disparities estimated by the stereo camera for these pixels have a low accuracy. Therefore, the applied U-net is extended with additional training images to predict the punched holes as background, as can be seen from the resulting segmentation masks on the right of Figure 3.

Additionally, due to the smoothing effect of the stereo correlation window, disparity maps are typically less accurate at depth discontinuities. This is the case, for example, at the transition from fas-

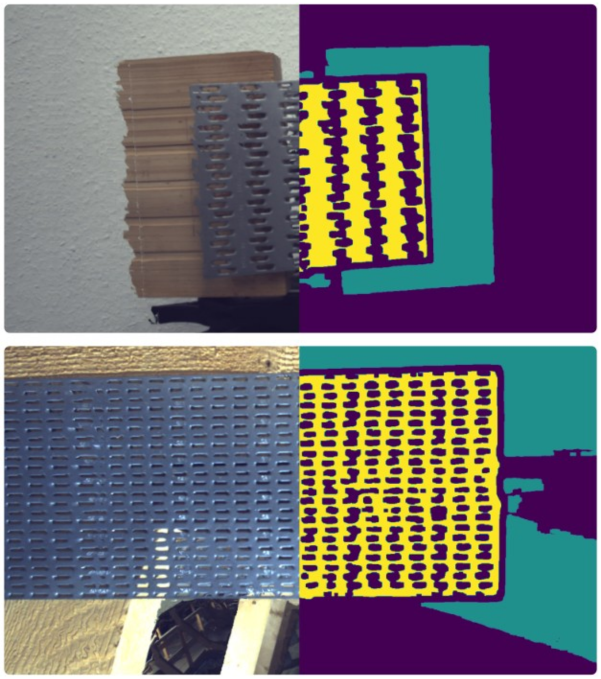

Figure 3. Example for segmentation results (right) based on colour images (left) for Sample A (top) and Sample B (bottom).

tener to timber. In order to remove these disparities, a morphological opening filter (Steger et al., 2018, Chapter 3.6.1) generates a buffer around the fastener's edges. To further assess the segmentation results of the entire image subset for Samples A and B, we determine the pixel accuracy and the intersection over union (IOU) of the prediction to a ground truth. At first the punched holes are not considered, which results in an accuracy of 96 % for the timber and 99 % for the fastener. The IOU is 0.87 and 0.99. If the holes are additionally segmented as background, the accuracy for the fastener decreases to 92 %. The IOU is reduced to 0.75, which likely results from the limited number of training images and the highly variable lighting conditions of the punched holes.

The individual captures are now combined by estimating their camera poses and transforming the 3D points to a common coordinate system. In the case of Sample A, the pose estimate of ORB-SLAM results in a good approximation, leading to a quick and robust convergence of the ICP algorithm. However, in the case of Sample B, the pose estimation does not converge for all point clouds, which is probably caused by inaccurate initial poses or the greater size of the timber joint. The reasons and improvements are further discussed in Section 4. By manually providing a more accurate initialisation of ICP, the pose estimation also converges for Sample B.

3.2 Evaluation With Sample A

In order to evaluate the accuracy of the stereo camera and the proposed methodology, we compare the measurements of Sample A with a reference measurement from a laser tracker. The laser tracker's sensing probe is guided over the surfaces, generating 3D points at regular intervals along the scanned lines. While the accuracy of these points is in the micrometer range, their density is too low for a direct comparison. Thus, the surfaces are approximated with planes for the timber elements and a polynomial model for the fastener, which servers as the reference surface. A polynomial with degree 2 in x- and y-direction shows residuals with a standard deviation of 0.06 mm and thus sufficiently describes the slight curvature of the fastener in the context of the stereo camera's targeted accuracy. The same models are also fitted to the stereo camera's point cloud for better comparability. Lastly, the reference model is registered to the stereo camera's point cloud by manually providing an initial pose and improving it by ICP.

Figure 4 shows the distances between points on the surface that represents the stereo camera point cloud to the reference surface. The distances are orthogonal to the reference surface. A mean distance of $+0.21$ mm indicates a systematic deviation between the two surfaces. The deviation could result from uncertainty in the registration or the camera calibration, as further discussed in Section 4. However, with a maximum distance of $+0.36$ mm and a minimum distance of -0.18 mm, the results are within the expected and required accuracy of the approach.

3.3 Model Comparison with Sample B

While the polynomial surface is sufficient to describe the fastener of Sample A, the real application scenario Sample B requires a B-spline representation, which is able to approximate local deformations. Therefore, we fit a B-spline surface with 5 control points in the u- and v-directions and compare the results to a polynomial fit with degree 2 in the x- and y-directions. We also fit a plane as a baseline and for visualisation of the fastener's curvatures

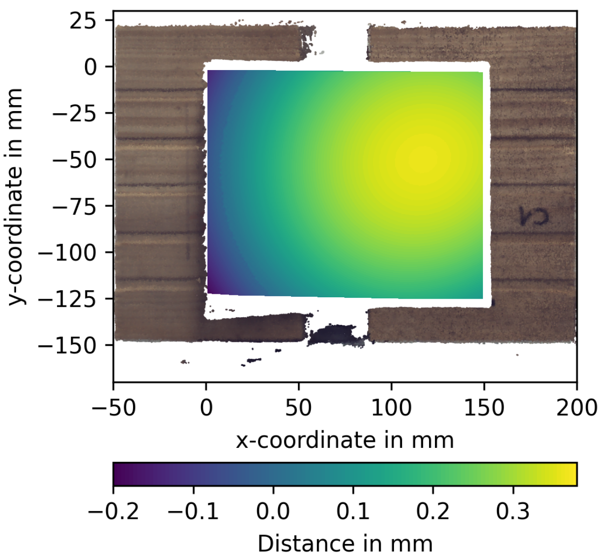

Figure 4. Orthogonal distances between polynomial approximations of the stereo camera's measurement to the reference measurement.

and local deformations.

Figure 5 shows the absolute residuals for every point in the stereo camera's point cloud to the fitted surface. In the case of the plane fit, the highest residuals with about 4 mm are located at the right side of the fastener, which is already visually identified as a local deformation in Figure 2. Other areas of the fastener also show residuals in the range of 1 to 2 mm, which highlights its non-planar characteristic. With the polynomial model, the residuals decrease slightly at the local deformation on the right side. However, an increase can be seen at the bottom left corner with values in the range of 2 to 3 mm. The B-spline surface shows a much more homogeneous distribution of residuals indicating a better fit. Residuals at the local deformation are not significantly higher than at other areas of the fastener.

Finally, the B-spline surface enables an estimation of the distances between the fastener and the timber. As the timber elements do not lie in the same plane, the fastener is split into three corresponding segments. In each segment, the distances between points on the B-spline surface to the individual timber plane are calculated. The distances are orthogonal to the timber plane. As seen in Figure 6, the bottom right corner exhibits a maximum distance of about 4 mm, which matches with the visual inspection in Figure 2. Additionally, the fastener's centre shows distances of up to 5.4 mm. Near zero or

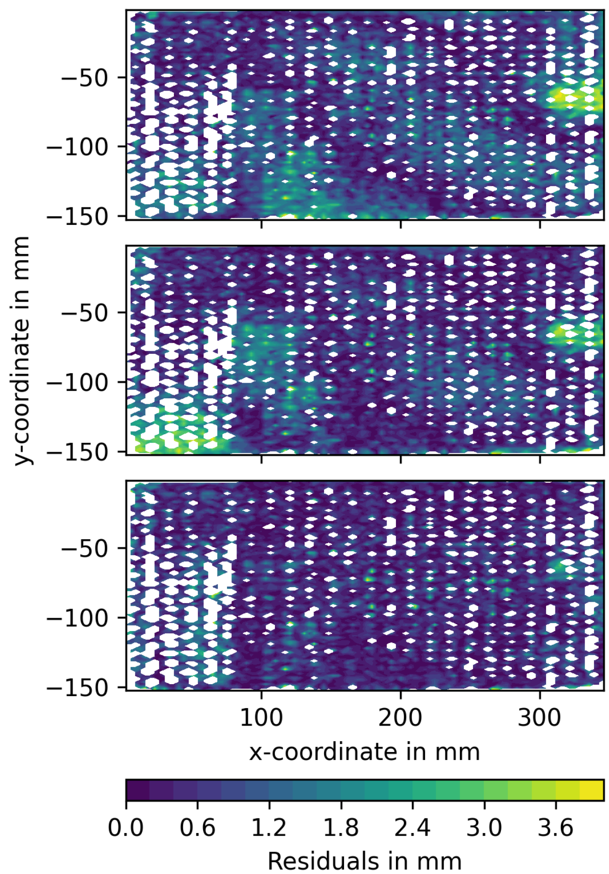

Figure 5. Absolute residuals of the plane (top), polynomial (middle), and B-spline (bottom) approximation for the fastener of Sample B.

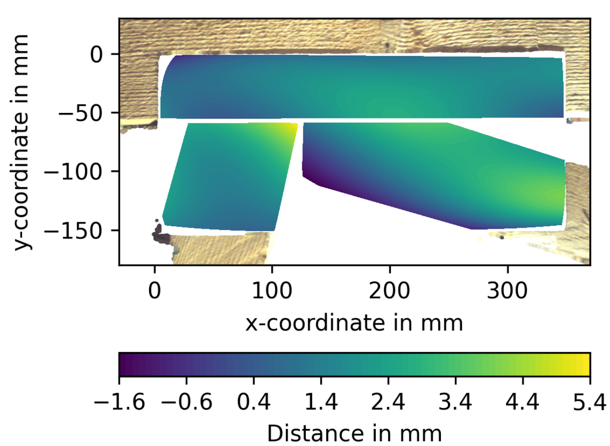

Figure 6. Orthogonal distances between the B-spline surface and the plane surfaces of Sample B.

even negative distances can be seen at the bottom right timber element. This element is much smaller than the top element and values for the area under the fastener have to be extrapolated from the fitted plane, which has a low accuracy. Future measurements should consider this influence, for example, by capturing a wider area around the fastener.

4 Discussion

The results of the two sample measurements show that the stereo camera and our proposed methodology are capable of capturing and modelling the surface of a timber-to-timber joint with sufficient precision. While the individual processing steps generate valid results, there are some aspects and challenges to consider for an entirely automatic evaluation of timber joints. For example, the training data for the image segmentation is limited and likely not diverse enough to achieve robust results in new measurement scenarios. However, a possibly lower segmentation accuracy could be compensated by the redundancy of the acquired images. Currently, the images are segmented individually and the significant overlap between consecutive images is unused. After registration, the point cloud segmentation could be filtered by generating a voxel grid and applying the segmentation label with the majority for each voxel to the corresponding points.

Registration of the separate point clouds has a considerable influence on the resulting accuracy of the estimated distances. Especially for larger timber joints, the captured area per image only contains a part of the fastener. In this case, the registration becomes ambiguous in at least one coordinate axis since there are no clear geometric features and the fastener's pattern is periodic. Initial poses with higher accuracy would be required, as shown with Sample B.

The measurement of Sample A exhibits a minor systematic deviation to the reference, which results from an uncertain registration of the reference to the stereo camera result. An uncertainty in the camera calibration could also cause this systematic deviation and is not discernible from the influence of the registration. Thus, the evaluation cannot determine the accuracy of the stereo camera independently from the registration. The camera poses have to be known in the coordinate system of the laser tracker, so that a registration is unnecessary.

The analysis of Sample B shows that B-spline surfaces can approximate local deformations and planar areas at the same time. A polynomial surface has much larger residuals, resulting in under- or overestimated distances. Underestimated distances could potentially be dangerous if the load capacity of a timber joint is derived from these values. Similar to the polynomial degree, the number of control points of the B-spline surface has to be defined beforehand. A B-spline surface with 5 control points should be suitable for most metal plate fasteners. However, an optimisation of this number depending on the specific sample could lead to more accurate results and would be important for an automatic evaluation of timber-to-timber joints.

5 Conclusion

Our study shows that stereo cameras can be used in an automatic reconstruction process for the assessment of punched metal plate fasteners on timber-to-timber joints. The automatic process first requires the segmentation of fastener and timber based on the camera's colour images. Afterwards, individual images and disparity maps are transformed to a combined point cloud with labels for fastener and timber. In two experiments, we have shown that the surface of a fastener can be approximated with a polynomial model or a B-spline surface. The B-spline surface can better approximate localised deformations in comparison to a polynomial model. The developed approach achieves deviations to a reference measurement of less than 0.5 mm.

An assessment of the distance of the fastener to the timber trusses still involves challenges, mostly in the automatisation of some processing steps. The image-based segmentation requires more training data for a generalisation to different fastener and timber combinations. Alternatively or additionally, methods for quantifying the uncertainty of the segmentation result could be used to assess the trustworthiness of the individual results (Landgraf et al., 2024b,a). This could help to select the most reliable segmentation result over different camera views. Improvements to the automatic registration of individual point clouds are also essential for an accurate and robust determination of the distances.

Acknowledgement

Funded by: Federal Ministry for Economic Affairs and Climate Action based on a resolution of the German Bundestag. We thank our colleagues of the Geodetic Institute at Karlsruhe Institute of Technology for supporting us with their laser tracker measurements.

References

Bouldin, J. C., Loferski, J. R., and Hindman, D. P. (2014). Inspection of metal plate–connected wood trusses in residential construction. *Practice Periodical on Structural Design and Construction*, 19(2).

Campos, C., Elvira, R., Gomez Rodriguez, J. J., Montiel, J. M. M., and Tardós, J. D. (2021). ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Transactions on Robotics*, 37(6):1874–1890.

Choi, S., Zhou, Q.-Y., and Koltun, V. (2015). Robust reconstruction of indoor scenes. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.

Drost, B., Ulrich, M., Navab, N., and Ilic, S. (2010). Model globally, match locally: Efficient and robust 3D object recognition. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 998–1005. IEEE.

Edelsbrunner, H., Kirkpatrick, D., and Seidel, R. (1983). On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29(4):551–559.

Harmening, C. and Neuner, H. (2015). A constraint-based parameterization technique for B-spline surfaces. *Journal of Applied Geodesy*, 9(3).

Hirschmüller, H. (2005). Accurate and efficient stereo processing by semi-global matching and mutual information. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, volume 2, pages 807–814. IEEE.

Landgraf, S., Hillemann, M., Kapler, T., and Ulrich, M. (2024a). Efficient multi-task uncertainties for joint semantic segmentation and monocular depth estimation. In *DAGM German Conference on Pattern Recognition (GCPR)*. Springer.

Landgraf, S., Wursthorn, K., Hillemann, M., and Ulrich, M. (2024b). DUDES: Deep uncertainty distillation using ensembles for semantic segmentation. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 92(2):101–114.

Paevere, P., Nguyen, M., Syme, M., Leicester, R., and Ho, K. (2009). Mechano-sorptive nailplate backout in nailplated timber trusses. *Forest & Wood Products Australia*.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation.

Steger, C., Ulrich, M., and Wiedemann, C. (2018). *Machine vision algorithms and applications*. John Wiley & Sons.

Wursthorn, K., Hillemann, M., and Ulrich, M. (2024). Uncertainty quantification with deep ensembles for 6D object pose estimation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-2–2024:223–230.

Zhou, Q.-Y., Park, J., and Koltun, V. (2018). Open3D: A modern library for 3D data processing.