Real-time Defect Detection in Conveyor Belts using Point Clouds Generated by a ToF Camera

Xaime Pardal Cardama, Carlos V. Regueiro, and Miguel R. Luaces

Grupo de Arquitectura de Computadores (GAC), CITIC Research Center, Universidade da Coruña, 15071 A Coruña, Spain
Database Lab., CITIC Research Center, Universidade da Coruña, 15071 A Coruña, Spain
Correspondence: xaime.pardal@udc.es

DOI: *https://doi.org/10.17979/spu.23.c42*

Abstract: A three-dimensional analysis system was developed using Time-of-Flight (ToF) infrared cameras to detect conveyor belt teeth and estimate their speed. From the captured point clouds, structural patterns are identified through geometric fitting and spatial segmentation. As a novel contribution, the system detects belt teeth, locating absences and deformations. Detection is performed by comparing the expected distribution with the observed one. The speed of the conveyor belt is estimated by calculating the relative displacement between consecutive point clouds. The system was successfully tested, providing accurate and reliable analysis in dynamic industrial environments.

1 Introduction

Three-dimensional vision through volumetric sensors has emerged as an alternative to traditional 2D vision systems and contact-based sensors. Among the available technologies, LIDAR systems operate at medium and long ranges but do not achieve millimeter-level accuracy. Stereo cameras, on the other hand, have a resolution that is inversely proportional to distance, their operating range is limited, and most importantly, they do not function properly without good illumination and texture presence. In contrast, infrared *Time-of-Flight* (ToF) cameras provide real-time depth maps at short and medium distances, with a good balance between accuracy, cost, and robustness against variations in illumination and texture. These features make them a competitive solution for industrial applications (Chen et al. (2022)).

In this work, we propose a system based on ToF cameras for defect detection in conveyor belts and speed estimation through point clouds. One of the most relevant problems arises in conveyor belts installed on slopes, which require the presence of teeth on their surface to prevent material from sliding downhill. A defect in one of the teeth can reduce the belt's capacity to move material, leading to blockages. Our method allows the identification of anomalies in the teeth, such as absences or deformations, and enables non-intrusive calculation of conveyor belt speed. Validation carried out in a real industrial environment confirms its usefulness.

The remainder of this paper is organized as follows: Section 2 reviews the related work, Section 3 describes the proposed method in detail, Section 4 presents the results obtained under real operating conditions, and Section 5 summarizes the main conclusions and future research directions.

2 Related work

In the current literature, no studies have been identified that address the detection of conveyor belt teeth through the analysis of high-density, high-resolution 3D point clouds. There are

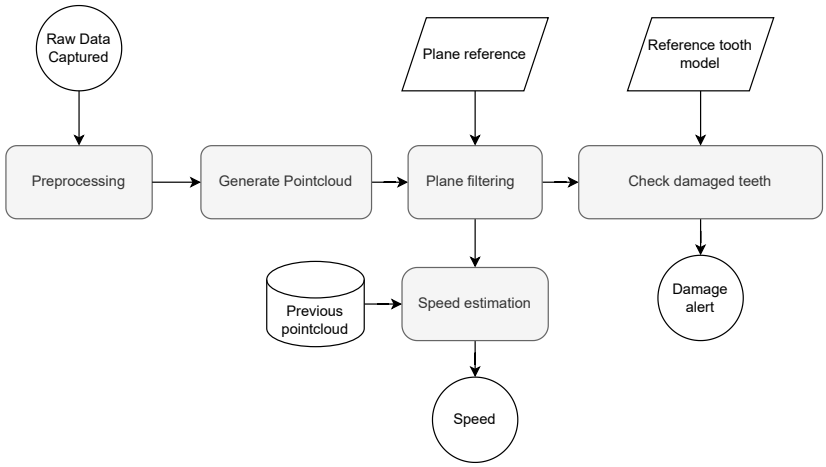


Figure 1: Processing pipeline of the proposed system

vision-based approaches, either in combination with a laser or by employing neural networks. In the first case, the laser must be synchronized with the belt to generate 3D measurements. Furthermore, it has only been tested with synthetic data (Netto et al. (2021)). In the second case (Guo et al. (2023); Xie (2023)), the images do not provide distance information, large computational resources (GPUs) are required, and the method is highly sensitive in industrial environments (low lighting and lack of textures). Futhermore, hundred of examples are needed to train the neural networks, so it has only been tested in a laboratory simulation environment.

Measuring material speed on conveyor belts is a challenge in industrial settings. The literature does not identify studies that explicitly apply the Iterative Closest Point (ICP) algorithm (Besl and McKay (1992)) to estimate conveyor belt speed. Two variants exist: *point-to-point* and *point-to-plane*. The former minimizes the distance between point pairs, making it simple to apply, but it is sensitive to noise and not robust on flat surfaces. In contrast, the *point-to-plane* variant minimizes the projection of the distance onto the local normal, converging faster and providing higher accuracy in geometries with defined structure, such as planes and belt teeth. In our context, the *point-to-plane ICP* is more suitable for stable and accurate speed estimation. ICP and its variants have demonstrated robust performance in point cloud registration under noisy conditions, partial occlusions, and incomplete geometries, such as surface inspection and mining (Li et al. (2022)).

3 Method

Figure 1 shows the pipeline of the proposed system. The raw data captured by the camera (Section 3.1) are preprocessed to generate a point cloud (Section 3.2), on which a plane filtering is applied (Section 3.5). Subsequently, teeth are checked using a reference model of one of them (Section 3.6). In parallel, speed estimation is performed using ICP (Section 3.7), based on the stored point cloud from the previous iteration. The reference plane, common to all iterations, is generated once and used throughout the process (Section 3.3). A reference tooth model is also generated once (Section 3.4), which allows the identification of missing or deteriorated teeth.

Figure 2: Example of raw data (from left to right): distance (pseudocolor), intensity and confidence.

3.1 Raw data capture

For the development of the system, a *Sick Visionary-T Mini* Time-of-Flight (ToF) camera (Sick AG (2024)) was used. This device is designed for industrial use and provides three-dimensional representations by emitting infrared light pulses and measuring their time of flight. Since it does not depend on ambient lighting, the camera can operate under adverse conditions or in darkness.

The sensor has a resolution of 512 $\times$ 424 pixels, with a maximum capture rate of 30 frames per second. Its working range extends up to 16 meters and its field of view is 70° $\times$ 60°. The depth resolution is 0.25 mm, although the spatial resolution is 1 mm. Regarding physical installation, the camera was mounted in an elevated and relatively stable position with respect to the conveyor belt.

The camera output consists of three matrices of equal size (Figure 2):

- **Distance**: distance in millimeters to each measurement point.
- **Intensity**: reflection of infrared light at each point.
- **Confidence**: identifies points with measurement errors.

3.2 Preprocesing and generate pointcloud

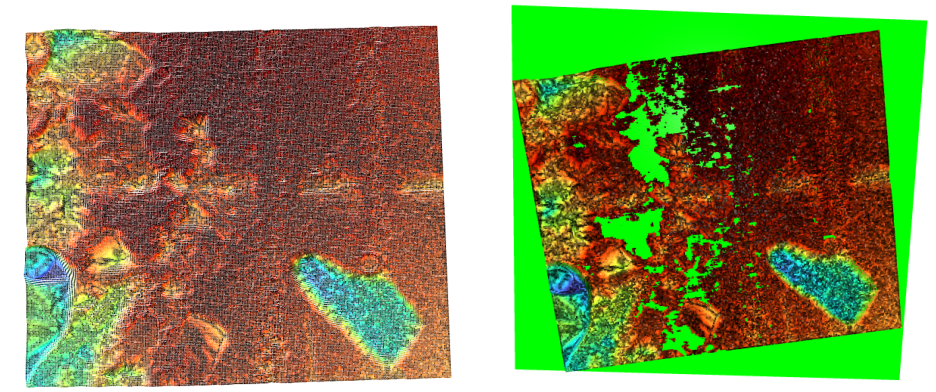
As the first step in processing the data captured by the ToF camera, a filtering based on the confidence map is applied. Subsequently, a distance-based filter is used to roughly discard regions clearly outside the area of interest. By setting a distance range, points are removed from the distance map that, due to being too close (possible sensor noise or unwanted reflections) or too far (entities unrelated to the object of study), do not provide useful information for the analysis.

Once the irrelevant regions are discarded, a third filtering is applied based on a closed polygon defined by four vertices, referred to as the ROI (Region of Interest). This ROI removes points belonging to irrelevant areas, such as the edges of the conveyor belt or structural elements of the environment.

Finally, a 3D point cloud is generated using the distance and intensity data returned by each pixel. The normal vectors of each point are then estimated. This information is particularly useful for subsequent algorithms that require knowledge of the local orientation of surfaces. Figure 3a shows a top view of the resulting point cloud.

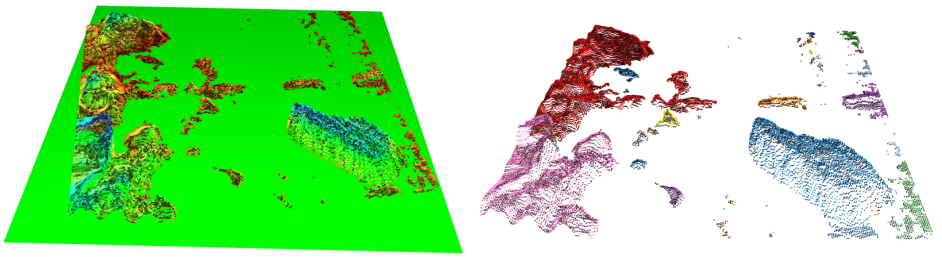
3.3 Generate plane reference

Before starting the execution of the pipeline, the reference plane used for filtering is generated. To identify the reference surface of the conveyor belt, the RANSAC (Random Sample Consen-



(a) Pointcloud preprocessed

(b) Detected plane



(c) Points filtered by plane

(d) Clustering

Figure 3: Plane estimation, filtering and clustering examples

sus) algorithm (Fischler and Bolles (1981)), implemented in the Open3D library, is used. This method is suitable because it is robust to noise and outliers, allowing the extraction of a dominant plane from the preprocessed point cloud.

The procedure consists of randomly selecting minimal subsets of points from the point cloud generated in Section 3.2, fitting a plane model to these subsets, and checking how many points of the global cloud fit this model within a tolerance threshold. After a sufficient number of iterations, the model with the highest consensus is accepted as the base plane. The final plane is selected according to the criterion of greatest extension. Figure 3b shows the preprocessed point cloud and the generated plane in green, demonstrating a correct fit to the surface.

3.4 Generate reference tooth model

A tooth model is also generated beforehand in order to identify them in the subsequent workflow. From a well-preserved tooth, previously segmented and manually validated through visual inspection of the clustered point cloud in Section 3.6, a reference model is created. This model represents its relative height with respect to the belt's base (reference plane). Since the teeth repeat periodically along the belt and share the same axis, locating one tooth in a frame allows us to determine the positions of the rest provided that the tooth spacing and count for that belt type are known in advance. These parameters must be supplied manually for each belt type as they are not estimated from the image.

3.5 Plane filtering

For the subsequent tooth detection process, a second filtering step is required to remove points below the plane and those exceeding a certain offset above it, as shown in Figure 3c. The result is a representation of the material circulating on the belt and its teeth, thus eliminating the structural part of the conveyor.

3.6 Check damaged teeth

The conveyor belt tooth detection process is divided into three phases: clustering, identification of a reference tooth, and checking for missing/damaged teeth.

Clustering To segment the point cloud into coherent regions, the DBSCAN algorithm (Density-Based Spatial Clustering of Applications with Noise) (Ester et al. (1996)) was used, in its implementation from the Open3D library. The data source used for clustering is the plane-cropped cloud, selected at runtime. With these characteristics, DBSCAN consistently isolates the conveyor belt teeth, as shown in Figure 3d, serving as a basis for their subsequent comparison with the reference model.

Identification of a reference tooth Subsequently, each cluster is compared with the reference model using the Hausdorff distance, defined as:

$$d_H(A,B) = \max\left\{ \sup_{a\in A} \inf_{b\in B} \|a-b\|,\ \sup_{b\in B} \inf_{a\in A} \|b-a\| \right\}, \tag{33.1}$$

where A is the reference model and B the observed cluster. This metric makes it possible to assess the degree of similarity between the expected tooth and the observed structure.

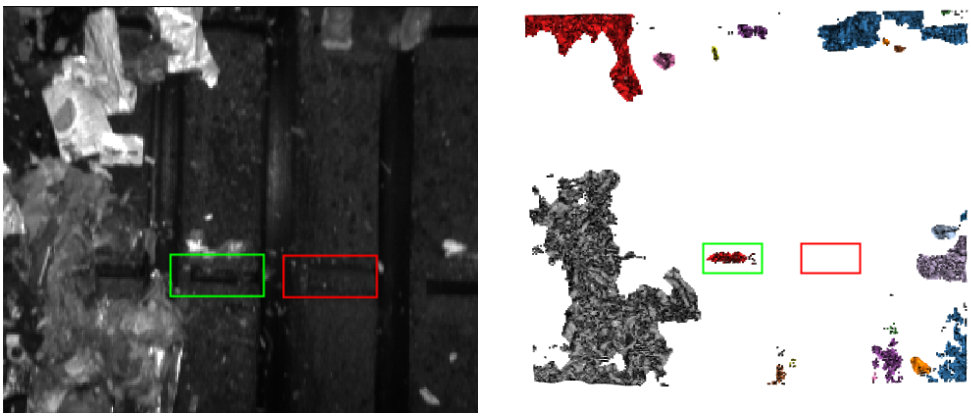
Check missing/damaged teeth When a tooth is detected on the belt, it is cropped and the remaining teeth in the group are searched. The process then proceeds as follows:

- **Confirmed presence**: when material of equal or greater height than the reference tooth height is detected in the expected region, the tooth is considered present and in good condition.
- **Absence or damage alert**: issued when a detected tooth differs sufficiently from the reference tooth. We calculate similarity using the Hausdorff distance (Eq. 33.1) and set the threshold at twice the sensor accuracy (i.e., 2 mm).
- **Positive validation**: if the Hausdorff distance is less than the set threshold, the tooth is classified as correct.

Figures 4a and 4b show an example frame where a missing tooth is detected. The algorithm identifies the green-marked area in both images as a correct tooth. By comparing the cluster of the tooth with the previously stored reference, the algorithm scans the red-marked area by moving along the tooth axis in the point cloud. Since no point cluster is detected at the expected location, it concludes that a tooth is missing, and an alert is generated.

3.7 Speed estimation

The speed of the conveyor belt is calculated from the translation obtained by the *Iterative Closest Point* (ICP) algorithm when aligning consecutive point clouds (Figure 5). Given two captures separated by a time interval Δt, ICP computes the rigid transformation $T = (R, \mathbf{t})$ that minimizes the distance between the cloud at instant t and that at instant $t + \Delta t$. The translation component Δx aligned with the conveyor's forward axis is used to calculate the average speed as $v = \frac{\Delta x}{\Delta t}$.



(a) Intensity image (b) Point cloud

Figure 4: Example of missing (red) and detected (green) tooths in a specific frame.

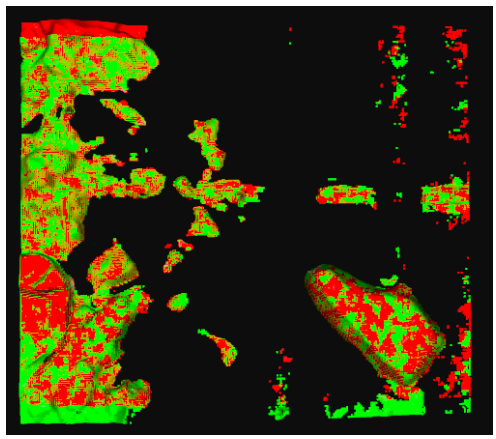


Figure 5: Example of two consecutive point clouds (red and green) aligned by point-to-plane ICP.

4 Results

The validation of our algorithm was carried out on a toothed conveyor belt of a waste separation plant under real operating conditions, since the ToF camera does not interfere with its operation. The recording used lasted 3 minutes and 22 seconds, with a total of 998 frames.

The system achieved satisfactory results in tooth detection within the evaluated frame groups, validating its effectiveness under real operating conditions. Verification was performed through visual inspection, comparing both the intensity images and the point clouds generated by the camera. This validation confirmed the correct identification of teeth and the absence of false positives in the analyzed sets. In the evaluated video, a total of 160 teeth were counted, of which only 53 were fully visible without material covering them. In addition, 8 missing teeth were identified. The system correctly detected all visible teeth and identified all missing teeth (100% accuracy rate) without recording any false positives.

The results of the speed estimation (Table 1) show a range of speeds between 1.05 and 1.53 m/s, with an average of 1.28 m/s and a standard deviation of 0.17 m/s. The variability in the measurements can be explained by the fact that the camera was not rigidly fixed to the structure, causing vibrations during recording that affect the stability of the point clouds, in addition to the inherent limitations of the ICP algorithm.

5 Conclusions

The developed system proved to be fully effective in monitoring conveyor belts using ToF cameras, achieving a 100% success rate in detecting missing or damaged teeth, without recording any false positives under real operating conditions, while providing a useful and realistic estimate of belt speed. These results consolidate its feasibility as an alternative with a clear industrial impact on predictive maintenance and quality control tasks.

As future work, we propose the application of deep learning for the detection of complex or difficult-to-model defects, the implementation of online calibration mechanisms, and a feasibility study to deploy solutions directly on edge devices, bringing the analysis capabilities closer to the sensor itself in order to optimize processing efficiency and real-time performance.

Bibliography

P. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

R. Chen, J. Xu, and S. Zhang. Comparative study on 3D optical sensors for short range applications. *Optics and Lasers in Engineering*, 149:106763, 2022.

M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231, 1996.

Table 1: Results of conveyor belt speed estimation by frame pair

Frames	Distance (m)	Speed (m/s)	dt (s)
62837→62838	0.0457	1.3450	0.033
62838→62839	0.0506	1.5329	0.033
62839→62840	0.0448	1.3585	0.033
62840→62841	0.0483	1.4211	0.033
62841→62842	0.0348	1.0535	0.033
62842→62843	0.0379	1.1476	0.033

M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24 (6):381–395, 1981.

X. Guo, X. Liu, P. Gardoni, A. Glowacz, G. Królczyk, A. Incecik, and Z. Li. Machine vision based damage detection for conveyor belt safety using fusion knowledge distillation. *Alexandria Engineering Journal*, 71:161–172, 2023.

J. Li, Q. Hu, Y. Zhang, and M. Ai. Robust symmetric iterative closest point. *ISPRS Journal of Photogrammetry and Remote Sensing*, 185:219–231, 2022.

G. Netto, B. Coelho, S. Delabrida, A. Sinatora, H. Azpúrua, G. Pessin, R. Oliveira, and A. Bianchi. Early defect detection in conveyor belts using machine vision. In *Proceedings International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 303–310, 2021.

Sick AG. Visionary-t mini data sheet. *https://www.sick.com/media/pdf/3/83/983/dataSheet_V3S105-1AAAAAA_1112649_en.pdf* , 2024. [Accessed 17-September-2025].

M. Xie. Analysis of machine vision-based defect detection for logistics conveyor belts. In *International Conference on Computer Vision and Pattern Analysis (ICCPA 2023)*, page 91. SPIE, 2023.