POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Robotic sole Deburring: from Burrs Identification to Path Planning from Human Demonstration

Laurea Magistrale in Mechanical Engineering - Ingegneria Meccanica

Author: Luigi Cacciani

Advisor: Prof. Andrea Maria Zanchettin

Co-advisors: Prof. Paolo Rocco, Ing. Alessandra Tafuro

Academic year: 2022-2023

1. Introduction

Shoe manufacturing is a dynamic blend of craftsmanship, technology, and innovation and is continuously evolving. Sole deburring, the removal of unintended protrusions of material on the edges of the sole called *burrs*, is crucial for the quality of the final product. Even if it is still conventionally done by skilled workers, automation with CNC machines has been exploited for extremely high production rates despite facing adaptability issues. Industrial robotics seeks to bridge the gap between human flexibility skills and high volume production. The integration of new-generation AI technologies, mainly machine learning (ML) and deep learning (DL), plays a pivotal role in enhancing robotics across various sectors. This fusion of advanced manufacturing, DL, and robotics sets the stage for exploring intelligent robotic deburring. This thesis delves into this transition, emphasizing the synergy of human expertise with cutting-edge technologies. The research aims to address the following key challenges:

- **Sole detection and segmentation:** precise identification of the sole with burrs among industrial scenarios complexities;
- **Burr's identification:** develop an efficient automated method for the detection of the burrs;
- **Cutting tool orientation:** teach the robot the optimal tool orientation from expert demonstrations;
- **Path planning:** create an automated path planning pipeline for efficient deburring.

The thesis recognizes the challenges of industrial conditions and proposes methods that ensure precision, robustness, and adaptability. As the manufacturing landscape transitions into the intelligent era, this research attempts to contribute to advancements in robotic deburring for the footwear industry.

2. Related Works

Burrs identification: The identification and measurement of burrs in manufacturing processes are crucial for maintaining product quality. A burr can be defined as a material protrusion beyond the designed dimensions, necessitating precise detection and removal. Current techniques span from *contact* to *non-contact* methods and the selection of the most suitable system is highly application-specific.
Touch sensors for measurements can be disturbed by the non-uniformity and considerable

size variability of the burrs: the presence of nearby burrs can disturb the measurement due to the conical shape of the tracer point.
In the field of non-contact systems, several methodologies employ laser displacement sensors (LDS) for direct burr size measurement: when mounted on the deburring robot, LDS allow to precisely measure the distance from the workpiece and, knowing all the other dimensions and the nominal geometry of the workpiece, the burr's height can be found point-wise.
Some very application-specific effects can be exploited for burrs identification like the temperature variance explored by Wulf [4]: utilizing a high-temperature thermographic camera during steel slab cutting burrs are detected based on thermal contrast.
Vision systems coupled with image processing techniques, such as thresholding, represent a promising non-contact solution. For instance, a 2D image of the workpiece can be transformed into binary form, enabling burr contour extraction and subsequent burr data derivation. Thresholding can also be applied for detection starting from 3D scans of the object of interest. In recent studies, Neural Networks are joined to visual systems for burrs identification from images.
Contour matching is another possible procedure: ideal geometric profiles are compared to identify peripheral defects.
Deburring path planning: Robotic deburring involves planning and motion execution. The planning phase focuses on obtaining the robot path considering machining parameter estimation. The subsequent motion execution stage translates these planned trajectories into physical deburring actions.
Path planning for robotic deburring encompasses three primary approaches. The CAD/CAM-Based approach utilizes CAM software with the CAD model to select and plan deburring paths. The Sensor/Vision-Based approach employs specialized sensors or vision systems for detecting workpiece geometry and defining deburring paths. Teaching through Human Demonstration involves physically guiding the robot through direct, indirect, or non-interactive methods, recording the demonstrated path for autonomous execution.
Literature showcases various path-planning methods integrating these approaches. In vision-assisted methodologies, 2D vision cameras are employed to identify workpiece features, serving as the basis for deburring paths. Innovative solutions leverage human expertise, like manual path drawing on the workpiece followed by digitalization through image processing or utilizing impedance control for the manual selection of key points on the workpiece.
Additional approaches involve offline teaching on a reference workpiece and online adaptation through 3D scans made by laser displacement sensors or alignment with the CAD model through convolutional neural networks.

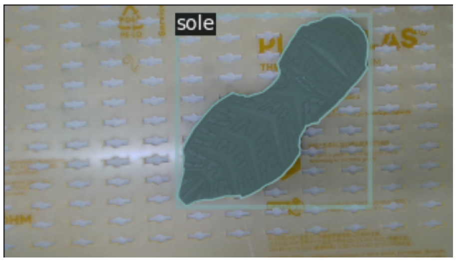

Figure 1: Sole detection annotations.

3. Sole detection and segmentation

The process of robotic sole deburring proposed in this study starts with a sole with burrs positioned on the working plane. An *Intel RealSense D453i RGB* camera located on the robot's end effector captures an RGB 2D image of the sole. Classic computer vision image processing techniques, namely simple global thresholding, Otsu's global thresholding and adaptive (local) thresholding have been tested for segmentation purposes but lack of robustness against varying lighting conditions and background changes has led to the choice of a Deep Learning approach. Moreover, DL methods can be exploited for their classification capabilities. The proposed approach for detection, classification and segmentation relies on **Detectron-2** [3], a state-of-the-art deep learning framework for object detection, built upon Mask R-CNNs.

3.1. Dataset generation and results

A **novel dataset** has been created for the detection and segmentation of shoe soles. 71 images have been captured with a resolution of (640x480): soles with and without burrs, with variations in orientation, background and

Model	Threshold		
	0.95	0.975	0.99
R50-FPN	100	91.7	73.4
R101-FPN	100	100	71.6
X101-FPN	100	97.8	77.5

Table 1: Segmentation results with Detectron-2.

clamping configuration have been included. Manual annotations of the bounding box and segmentation mask have been applied to each image (Figure 1): the bounding box label consists of class designation (only one class considered, *"sole"*), center coordinates and box dimensions, while segmentation labels provide the pixel coordinates defining the sole region. The dataset has been split into training (80%) and testing (20%) subsets.
Experiments with three backbone models from the Detectron-2 model zoo have been carried out, namely *R50-FPN*, *R101-FPN* and *X101-FPN*. Table 1 summarizes the results in terms of segmentation Average Precision (AP) values, with Intersection over Union (IoU) thresholds of 0.95, 0.975 0.99. Despite all models showing excellent segmentation accuracy, the *X101-FPN* one has been chosen due to the best performance under the most stringent demand.

4. Reconstruction of the occlusions

The operating conditions of robotic deburring, both in the study's experimental setup and real-world industrial settings, introduce a crucial challenge. The sole must be clamped securely to a fixed base, necessitating a gripping system that unavoidably obstructs parts of the sole's profile. The introduced occlusions may affect the accuracy of the burr identification method.
To reconstruct the occluded portions, image-to-image translation has been exploited, specifically using **Pix2Pix** [2], a Conditional Generative Adversarial Network (cGAN).

4.1. Dataset generation and results

The training dataset for Pix2Pix comprises pairs of images: a ground truth image representing the unobstructed sole segmentation and an image featuring randomly created occlusions on the sole profile.
Since the objective is to reconstruct the profile independently of the specific shape of the burrs, ten distinct profiles of soles with burrs have been used as ground truths. Random rotations and translations within the image domain have been introduced to enhance the model's generalization capabilities.
The results, shown in Figure 2, report that the trained network is able to correctly reconstruct the profiles provided in the validation dataset. However, challenges arise when integrated with the segmentation obtained from Detectron-2 (Figure 3). Irregularities in the segmentation profile are misinterpreted as occlusions and Pix2Pix tries to address them, leading to blurred and incorrect reconstructions.
For future developments, it is crucial to introduce a post-processing step for the segmentation generated by Detectron-2. This step should distinguish sections with occlusions from those without, enabling the smoothing of the profile in areas lacking occlusions. The goal is to refine Detectron-2's segmentation, eliminating irregularities and ensuring Pix2Pix's accurate reconstruction of occluded portions.

5. Burrs identification

A novel burrs identification method has been developed leveraging computer vision and image processing techniques. The objective is to correctly overlay the nominal designed sole profile onto the burr-containing sole profile. This method is independent of the size, orientation and position within the image of the sole with burrs.
Firstly, a **template of the nominal profile** has been extracted from an image of a sole without burrs through image simple global thresholding and contour detection. Its scale and position have been standardized: it has a unitary scale and is positioned at the origin of the image. When the image with the segmented sole with burrs is obtained, its oriented bounding box is calculated: the nominal template is scaled, rotated and translated to align with the sole's dimensions, orientation and position provided by the oriented bounding box (Figure 4).
An **optimization process** is developed to further refine the contour matching, maximizing the ratio of the intersecting area between the nominal profile and the profile with burrs and the nominal profile area. The nominal profile is scaled, translated and rotated to directly search

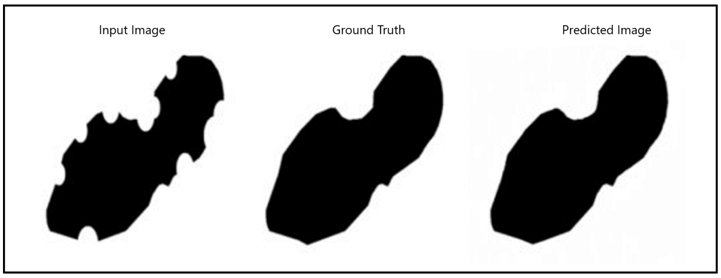

Figure 2: Pix2Pix result from validation dataset.

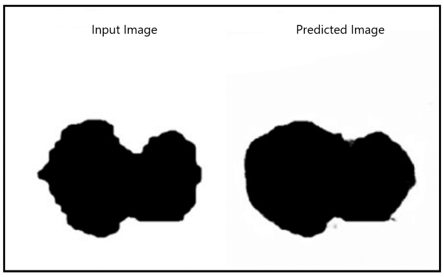

Figure 3: Pix2Pix result on Detectron-2 segmentation.

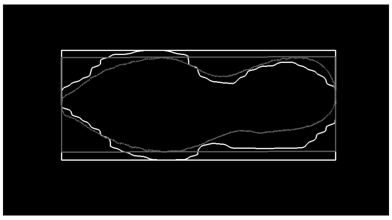

Figure 4: Initial guess of overlap

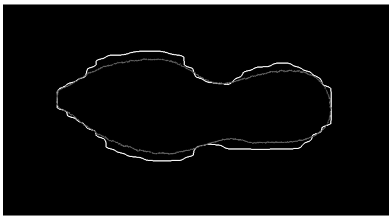

Figure 5: Optimized contour matching

Figure 6: Segmentation of the burrs

the solution. To enhance the convergence speed and final accuracy, the search space and the utilized steps are enlarged or diminished depending on the actual value of the areas' ratio: bigger steps and ranges when the solution is far from the optimal, to speed up the search, and smaller ones when close to the optimal for refinement. The process stops as soon as the overlapping percentage is higher than 99.5% (Figure 5).
The burrs, then, correspond to the area in between the two profiles: a pixel-wise XOR logical operator has been used to obtain the segmentation of the burrs. Subsequent filtering retains only the thickest physical portions, eliminating small areas coming from errors from the segmentation and the image processing steps (Figure 6).

6. Tool orientation through LfD

To achieve effective deburring, the optimal orientation of the cutting tool is crucial. In this study, the proposed approach involves teaching the robot the ideal orientation through *human demonstration*, employing videos of experts performing the task. Direct teaching methods, like kinesthetic teaching, have been discarded to enhance the spontaneity of the demonstrated deburring gesture, enhancing the precision in showing the best pose. To extract the pose from RGB images a Deep Learning approach has been used, precisely EfficientPose [1], known in the literature for its exceptional accuracy of pose estimation.

6.1. Dataset generation and results

Since precise annotations of the full 6D pose, in a large volume of images, are required to create an effective dataset, the process of precisely annotated data generation for pose estimation can be an exceedingly time-consuming and often costly task. To address these challenges, this study has opted to use **synthetic data**, employing BlenderProc, an open-source framework for synthetic image rendering (Figure7a): the 3D model of the tool can be positioned in a known pose, creating a precise ground truth annotation. To address the reality gap, backgrounds are composed of soles and hands, enhancing the realism of the data and the similarity with the real working environment. Physical features like surface roughness and a varying point light source have been also added for realism and to ensure adaptability to varying working conditions. The dataset comprises 10,000 scenes, split into 80% training and 20% validation. The trained model achieves an ADD-S value of 64.85%, showing the capability of pose estimation both on the validation dataset (Figure 7b) and also on real images of a tool held by a hand during the demonstrative video (Figure 7c). For experimental purposes, **three videos** of deburring operations have been recorded and **116** frames have been analyzed by EfficientPose from each video. The system learns the orientation of the Tool Center Point (TCP) reference frame (defined coherently with the cutting tool shape) with respect to the burr's local reference frame (defined at *each point* of the nominal profile where a burr is present, composed by a tangent and a normal to the profile). This orientation has been estimated in every acquired frame, and an average of the quaternion components has been calculated. The learned orientation is (X,Y,Z Euler angles in degrees): $[178.70 \ \ -9.61 \ \ -178.77]$.

(a) Dataset element sample.

(b) Validation dataset prediction.

(c) Real image prediction.

Figure 7: EfficientPose results.

7. Path planning pipeline

Each step in this study has proven effective in its required contribution. The ultimate objective is to integrate these diverse outputs cohesively, deriving the final path.

7.1. Experimental setup

The setup used for the experiments made in the *Mechatronics and Robotics* laboratory of *Politecnico di Milano*, illustrated in Figure 8, comprises several key elements, including the sole requiring deburring, the *Intel RealSense* RGB camera, positioned with a suitable support in correspondence of the final joint of the robotic arm, and the robotic arm itself, equipped with a mock deburring tool with the cutting blade at its tip. The tool has been designed and 3D printed, in order to simulate the deburring behavior, taking inspiration from real robotic actuated deburring tools.

The RGB image of the sole with burrs is acquired when the robotic arm is in *home position*, a known configuration, easily accessible as

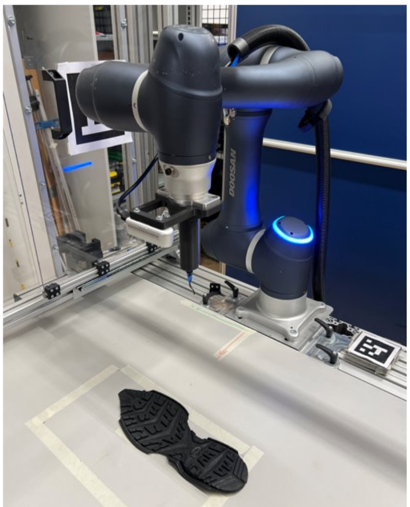

Figure 8: Experimental setup

starting point for all the deburring operations. The segmentation and the burrs identification are performed as described in the previous sections.

The deburring path in the plane is determined by intersecting the segmentation of the identified burrs with the nominal profile, obtaining a collection of deburring points. The result is calculated in the image reference frame (coordinates of pixels). These coordinates are then transformed from the image frame to the camera frame exploiting the intrinsic parameters of the camera (z is the vertical distance in home position, known):

$$x = \frac{(u - c_x)z}{f_x} \tag{1}$$

$$y = \frac{(v - c_y)z}{f_y} \tag{2}$$

From the camera frame, the coordinates are transformed into the robot base frame through a known transformation, calculated in the robot home configuration.

In every point of the nominal profile that intersects the burrs, a local reference frame is defined. Normal and tangent vectors to the profile are identified: the normal points outside the sole and the tangent rotates clockwise along the profile (found by imposing that the vertical axis points upwards).

The deburring tool's orientation is then set by imposing the learned orientation (section 6) with respect to these local frames.

Lastly, a clockwise deburring direction is imposed.

	Burr				Avg. Eucl. dist.
	1	2	3	4	
Test *I*	8.7	27.1	5.9	35.9	
Test *II*	5.2	7.1	11.5	24.7	15.8

Table 2: Average distance in mm between calculated and demonstrated in-plane trajectories

7.2. Experimental results

The experimental validation aims to quantitatively assess the accuracy of the computed deburring path by comparing it with the path demonstrated by an expert who has **physically manipulated the robot** to demonstrate the path. Two tests have been conducted, considering only the in-plane trajectory (x, y) due to challenges in precise robot positioning during physical movement. For every sole, the deburring paths of four burrs have been considered. The results are shown in Table 2. The average Euclidean distance between calculated and measured points has been 15.8 mm, attributed to accumulated errors from image processing and accuracy in the transformation matrices' calculation. The result is deemed highly acceptable for the study's initial deburring path objective, emphasizing the need for a control algorithm in practical industrial scenarios to address inherent errors and uncertainties.

8. Conclusions

This thesis deals with the development of a robotic deburring path planning pipeline, integrating learning from demonstration, and addressing challenges such as sole detection and segmentation, occlusion reconstruction, burrs identification, and tool orientation learning.
Detectron-2 is employed for sole detection and segmentation, demonstrating exceptional accuracy in recognizing the precise mask, even under diverse conditions.
To handle occlusions in the detected profile, Pix2Pix is employed for image-to-image reconstruction.
The burrs identification method involves obtaining a template of the nominal profile and a novel contour matching algorithm, composed of the identification of a first guess and the subsequent optimization, to find the position of the nominal profile inside the one with burrs. The segmentation of the burrs is finally found.
The optimal orientation of the tool during the task is learned from videos of expert demonstrations, exploiting the EfficientPose pose estimation neural network.
The automated pipeline demonstrates the joining of the aforementioned results to compute the complete deburring path.
The experimental results highlight the high precision achieved by the designed procedure, showing the potential of these technologies to contribute to advancements in the field.

References

[1] Yannick Bukschat and Marcus Vetter. Efficientpose: An efficient, accurate and scalable end-to-end 6d multi object pose estimation approach, 2020.

[2] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.

[3] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. `https://github.com/facebookresearch/detectron2`, 2019.

[4] Christian Wulf and Manfred Hayk. Automatic residual burr detection on steel slabs by a thermographic system. *Stahl und Eisen*, 30:36–37, 04 2007.