IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 20, 2021, accepted June 22, 2021, date of publication June 30, 2021, date of current version July 20, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3093526

Defect or Design? Leveraging the Angle of Opportunity for Detecting Scratches on Brushed Aluminium Surfaces

ANNE JUHLER HANSEN, THOMAS B. MOESLUND, AND HENDRIK KNOCHE
Department of Architecture, Design, and Media Technology, Aalborg University, 9000 Aalborg, Denmark
Corresponding author: Anne Juhler Hansen (ajha@create.aau.dk)

This work was supported by the Manufacturing Academy of Denmark/Innovation Fund Denmark.

ABSTRACT Design features such as polishing strokes share similarities with defects; this makes defect detection and quality assessment difficult to perform both manually and automatically. Human assessors rotate objects to probe different incoming illumination angles and evaluate the defect dimension to limits samples i.e. decide whether differences between defect candidates and design features qualify as a defect. This process has poor access to quantifiable defect descriptors needed for automation and expose a gap in the existing evaluation of defects. To integrate this notion into automated defect detection we propose a spatio-temporal image acquisition setup capturing the defect descriptor *Angle of Opportunity* (AoO) which can be used as a feature for image-based classification. The Random Forest approach classified defects with an area under the ROC-curve of 92%.

INDEX TERMS Aesthetic quality, defect inspection, machine vision, visual appearance.

I. INTRODUCTION

Premium products with low volume require expensive human inspection to avoid customer rejection especially when design features such as polishing strokes share similarities with defects [1]. Defects are defined based on visual descriptors and can be quantified and separated from design features due to irregularities in area, contrast, width, length, frequency etc. [2]. In their search for defects, human assessors rotate objects to probe different incoming illumination angles to make defects visible and compare them with design features [2]. This technique is especially used when assessing reflective surfaces. However, standard automatic visual inspection (AVI) processes [3] identify defects from single images. At inopportune angles this can render defects invisible or at many angles result in many defect candidates that in fact are design features. Consequently, we propose using the *Angle of Opportunity* (AoO), i.e. the total subtended angle during which defects are visible, as a defect descriptor. The contribution of AoO in defect detection tasks and whether it can improve results with standard defect descriptors, such as contrast, is currently unknown. The problem statement include how AoO can improve defect detection on reflective surfaces where high similarity between defects and design features exists. Our results illustrates the importance of using AoO in combination with other features as contrast, area, width, length, and rotation for human-in-the-loop processes. Additionally we quantify scratches based on different defect descriptors and compare these to the human ordinary acuity, existing quality limits and non-defects in our data set. The non-defects include both design features such as polishing, defect candidates just above the quality threshold and noise in the images e.g. from reflections.

This paper contributes the defect descriptor Angle of Opportunity (AoO) which quantifies the subtended angle across which defects are visible. Leveraging the AoO our model provides recommendations in terms of picture acquisition angles for automatic visual inspection tasks on brushed aluminium surfaces. Our results illustrate how capturing a series of images and using spatio-temporal defect descriptors such as AoO can improve detection of defects.

II. DEFECT APPEARANCE AND QUALITY STANDARDS

Since humans are highly sensitive to material appearance [4], product designers meticulously craft the visual appearance of

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Sharif.

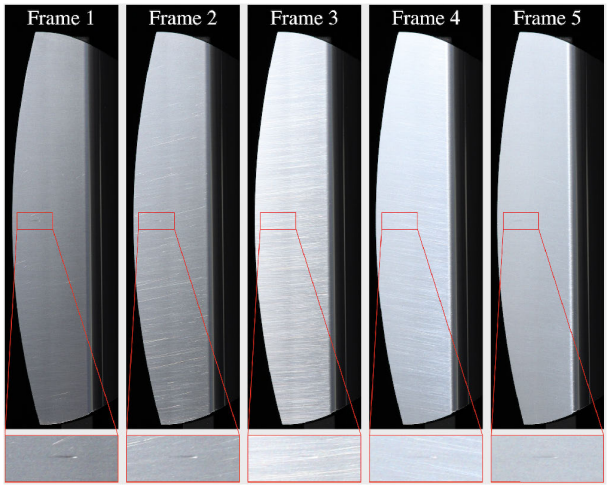

FIGURE 1. Example of the surface of a unit (top) in five frames rotated 1° apart and detailed view (bottom). From certain view angles the defect (red box) has high visibility (frame 1). As the surface is rotated the visibility of the polishing increases (frame 2) and the defect characteristics become more similar to the polishing (frame 3 and frame 4) and continuously changes based on rotation and view angle (frame 5).

consumer products to satisfy customer needs. Consequently, companies invest extensive amounts of time on the design and quality control of surface finishing. Levitt's definition of manufacturing quality is defined as *conformance to specification* [5]. In other words, quality should be measurable given the design specifications (including tolerances) and can be supported by standards providing universal methodologies and descriptions agreed upon by international experts (e.g. ISO8587 [6] and CIE TC1-65 [7]). Standards exists for describing and measuring the main dimensions of visual appearance: texture, color, translucency, light distribution, gloss (e.g. ASTM D523 [8] and ISO 2813 [9]) and support specifying visual appearance of different surface materials and illumination. The ISO8587 standard defines a large range of defect types linked to manufacturing, material purity and more (e.g. scratches, dents, holes, color changes, etc.) [6] but lack quantification of the defect descriptors. Some defects cause geometrical changes in the surface structure and other defect types constitutes only visually perceived differences (e.g. color changes) [10]. We address our problem of quality inspection of premium products by evaluating different visual features in the context of brushed aluminium surfaces. As follows, we narrow our scope to visually perceived defects and more specifically line defects [2]. These include common defect types as scratches, polishing fish etc. typically characterized by long thin shapes. This constitutes a challenging problem since line defects on brushed aluminium surfaces have high visual similarity to the unique polishing patterns that constitute a design feature.

Aesthetic quality assessment is a complex task to perform [11], and manual inspection can suffer from poor inter-assessor reliability due to variances in attention, level of training, individual state (mood, sleep etc.) and more [12]. Therefore, aesthetic quality assessment can benefit from objective and automated systems which can minimize the human bias and make this procedure more objective. Human assessors search for defect candidates and evaluate whether these are above or below the quality threshold of defects. Surface anomalies include anything that does not meet the designers' norm. Defect candidates that exceed the quality threshold are classified as *defects*. When assessors determine whether a defect candidate is above or below the quality threshold the design specifications are used as reference. The design specifications include measurement rules, tolerance levels and limit samples which are real-world examples of anomalies and defects close to the thresholds. Specular reflective metals, specifically brushed aluminium surfaces, possess brush strokes across the surface. Deep brush strokes can constitute anomalies since people perceive the brush strokes as scratches instead of homogeneous polishing, and thus this design feature can easily be mistaken for a defect. Other examples include impurity in the material (e.g. polishing fish) and handling defects (e.g. scratches) that can have an appearance similar to the design. Automatic systems should provide evaluations of defect in line with the design specifications and provide the same judgment performed by the expert assessors. This includes leveraging the current practice of assessors rotating objects and applying standard defect descriptors (location, size, shape etc.) [1], [2]. Current standards [6] are used for assessment of defect types within manual inspection but do not quantify the defect descriptors such than an automated system could use them to classify defects.

III. AUTOMATIC VISUAL INSPECTION

Visual inspection depends on illumination that causes varying appearance of both materials and defects in terms of surface properties such as texture, reflectivity, geometry etc. [13]. Nondestructive methods, as machine vision, can be used to inspect surfaces without permanently altering the inspected object and are exceedingly requested in industry [11]. Previous work involving machine vision for visual inspection focused on different materials including leather [14], ceramic [15], stone [16], plastic [17], and metals [18], [19]. Searching for defects on diffuse materials as leather and stone is easier than detecting defects on reflective surfaces as metal and plastic due to the specular reflections causing overexposure and appearance variance from certain viewpoints. The above approaches captured data from only a single viewpoint and lacked multi-view assessment of defects. Highly reflective materials, such as brushed aluminium surfaces, requires units to be inspected from various view angles [2]: a) for all unit surfaces in 3D (e.g. front and back sides) b) for defects only visible from certain view angles, and c) to account for specular reflections in the images. It is important to know the characteristics of different defect types to improve the defect detection process through better data capturing e.g. through multi-view assessment. As example, certain defect types, such as line defects, become more visible

as the image variance increases since these defects reflects the light differently than their adjacent areas [2], in this way, supporting our belief that brushed aluminium surfaces requires units to be inspected from various view angles [20].

Data-driven deep learning (DL) has increased in popularity and lead to increased performance in computer vision tasks such as image classification [21], object detection [22], segmentation [23] and tracking [24]. Real-time defect detection has previously been explored on different reflective surfaces. This includes inspection of highly reflective curved plastic surfaces in the automotive industry [25], diagnosing the penetration state of laser weld [26] or finding defects on highly reflective ring components [27]. Tiny casting defects can be detected with a CNN [28] or the surface quality of welds can be predicted [29]. Detection can be improved using information fusion [30], and has been considered for automatic inspection of thermal fuses where incorporating machine vision with artificial neural networks is used for detection of four common defect types [31]. A deep neural network can efficiently learn to recognise patterns in specific data sets, but problems arise when DL performs as a black box used for domain specific tasks where data is sparse (e.g. low-volume products) [20]. This provides little insight into the structure of the function being approximated and, therefore, no findings on the effects of different visual characteristics of defects. This is needed since companies defining aesthetics wish to articulate the measured descriptors, and for this reason a human-in-the-loop could be a first approach. Consequently, we cannot relate the feature extraction or decision process from the deep neural networks to the current manual inspection practice. Concretely for our problem, detecting defects (holes, stripes, scratches, dents, and pressure marks) on low-volume brushed aluminium surfaces, previous work showed that an off-the-shelf object detector (e.g. YOLOv5) yielded poor results (under the precision-recall curve of 0.67) due to many false positives [20]. The large number of false positives were, among other, linked to the reflective surface and were problematic since the false positives incorrectly indicated the presence of defects when no defects were present.

Other common methods used for defect detection in images include various texture analysis techniques. A significant amount of previous texture analysis methods are based on statistical (measuring and evaluating the spatial distribution of pixel values) and filter based approaches (computing the energy of the filter responses) [32]. These methods take advantage of the image characteristics (contrast and edges) in the spatial and spatial-frequency domain. The descriptors produced by the texture analysis models are often abstract and it is difficult to explain the relationship to the human visual inspection process. Many different defect types exist [6], where machine learning based on vision is limited by the size of corpus of defects when classifying different defect types. When working with low-volume production it is not possible to obtain an representative corpus. Existing MVI solutions are therefore often delimited; this includes limiting quality control based on material properties [33] or constraining the problem to individual product types in a controlled setup with fixed object placement and illumination [34]. Consequently, automatic quality inspection is lacking solutions for low volume high quality manufacturing due to the lack of data and the inadequacy of traditional statistical approaches [35]. In general comprehensive data sets are challenging to obtain due to the randomness and uniqueness of the extensive number of different defect appearances [33]. For proper multi-class classification of many different defect types industrial applications require a representative corpus (e.g. the MVTec AD data set used for anomaly detection of various surface defects [36]). Due to the vast amount of different defect types previous approaches reduce the defect space to a limited amount defect types e.g. welding, exfoliation, or sink marks [17], [18] and perform binary classification (defect versus non-defect) [31] and/or anomaly detection [36]. In other words, previous work limit the amount of classified defect types or apply anomaly detection looking for anything in-homogeneous on the surface. Limiting the classified defect types include; (a) using a large data set for training one classifier, or b) using different classifiers for different defect types as a binary classification task. We choose to focus on the defect detection of line defects including a binary classification of defects versus non-defects. We also narrow our scope to defect descriptors that can be easily explained as well as linked to the current manual inspection process as performed in industry i.e. using design specifications (written descriptions supported by standards), rulers (measuring defined metrics) and limit samples (physical examples of defect types near the quality threshold). We concentrated on working with hand-crafted features in our classification and link our defect descriptors to the common industrial practice. Building on the current manual inspection process where objects are rotated to assess defects, we found a lack in the general knowledge of the defect features, especially spatio-temporal defect descriptors as AoO. Along these lines we investigate six perceptual defect descriptors [2]: *AoO, rotation, contrast, area, length, width, minimum angle and maximum angle* (see Fig. 2) in relation to the human visual system (i.e. ordinary acuity limits common for all humans) and quality thresholds set by experts.

IV. VISUAL PERCEPTION OF SCRATCHES

Human visual acuity (VA) is the ability to discriminate details. When measuring VA, instead of considering absolute size (meters) we favor the angular size; minutes of arc (arcmin). Arc minutes describes the visibility of details of a particular frequency as a function of viewing distance. Reading characters (e.g. when measuring VA using a LogMAR chart) covering 5 arcmin of the visual angle requires a VA high enough to resolve 1 arcmin. This means that humans with normal vision can resolve contrast differences (depending on the spatial frequency) subtending an angle of approximately 1 arcmin (0.017°) [37] and under ideal conditions higher (we performed a simple desk test on our data and

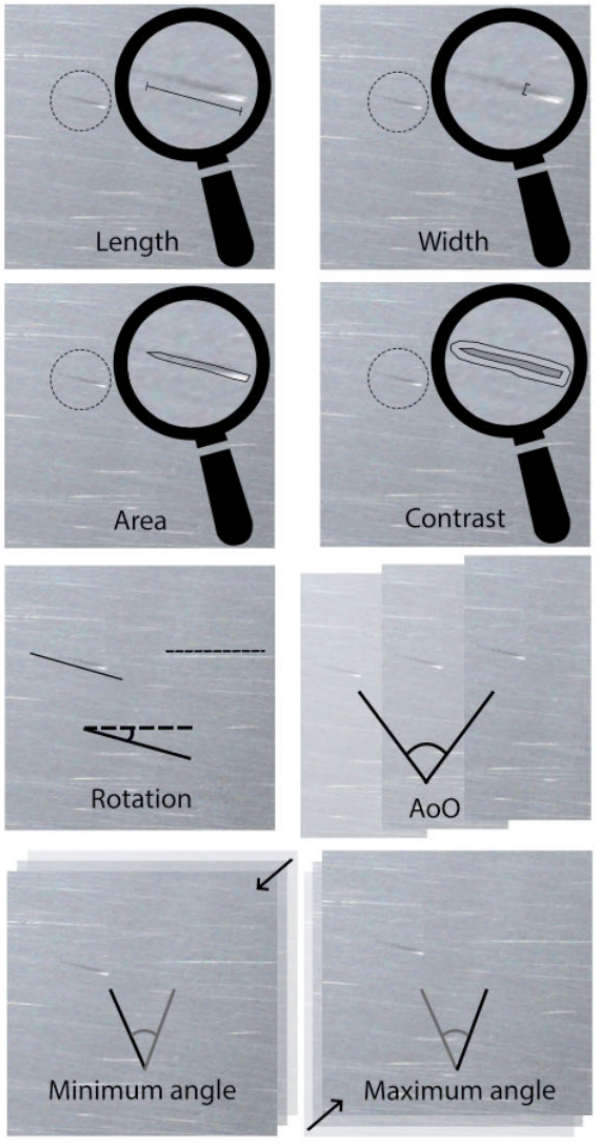

FIGURE 2. Defect descriptors; length, width, area, Weber contrast, AoO (angle where a defect candidate is visible), rotation (deviation from horizontal lines), minimum angle (first visible) and max angle.

calculated a visual acuity of 0.7 arcmin, however, we define the VA as 1 arcmin). Visual defects by definition need to be visible, hence irregularities below the VA are not visible and therefore not defects. Brushed surfaces contain polishing strokes varying in sizes with the smallest strokes often being below the resolution of both human VA. Based on contextual inquiry at a high quality manufacturing company we consider the typical viewing distance of our inspected product type to be equivalent to an arms length (50 cm) where 1 arcmin translates to a size of 0.15 mm. The spatial contrast sensitivity function (CSF) measure the amount of contrast required to detect an object. The CSF can be measured using sine-wave gratings (parallel lines that change in brightness over space) where the larger the difference in brightness between the dark and the lighter lines equals higher contrast. People can discriminate differences given Weber's law: constant $= \delta$I/I. Weber's law expresses that the human visual system is more sensitive to light intensity changes in low light levels than in strong ones and follows the Stevens effect which describes that contrast increase with luminance (i.e. the relationship between perceived brightness and measured luminance tends to follow a power function) [38]. When seeking to distinguish objects from a background (i.e. determining the luminance of a polishing stroke and the luminance of the area surrounding it) the human visual system is able to differentiate between different spatial frequencies (either low spatial frequencies composing generic shapes or high spatial frequencies including considerable detail and edges). In the context of brushed aluminium surfaces the contrast of the polishing can be considered the difference in illumination between the individual polishing strokes and the background. Thus the perceived contrast between defects, polishing strokes and the background surface will be affected by the contrast (i.e. spatial frequency of the surface and defect candidates) and light intensity changes depended on illumination and view angle. In addition, the polishing strokes of brushed aluminium surfaces can be evaluated based on both spatial frequency and orientation. The oblique effect describes that visual acuity is better for gratings oriented at 0° or 90° (relative to the line connecting the two eyes) than for gratings oriented at 45° [39]. Simple symmetric shapes and orientations in texture also exhibits an oblique effect [40]. We assume that line defects oriented at 90° relative to the polishing are easier to perceive by the human perceptual system than defects oriented at 45°.

V. DATA ACQUISITION

We obtained a set of 50 individual brushed aluminum units (i.e. an end cap for a speaker) from a Danish high quality production company. The shape of the front side can be described as a isosceles trapezoid (dimensions: 173mm×57mm). The backside can be estimated by a rectangle with rounded corners (dimensions: 133mm×70mm). All defects on the units were visually assessed and classified by expert visual assessors from a Danish high quality production company. The brushed aluminium surfaces possess handcrafted (i.e. unique) polishing.

A. ROBOTIC IMAGE CAPTURING

RGB images were captured using a Canon 5D (JPG with an image resolution of 6050 × 3300px) using a 70-300 f/4-5.6L IS USM lens. We used a fixed light source (Elinchrom Modeling Lamp 100W, color temperature: 3200K, luminous flux: 2700 lumen) at a 45 degrees angle from the camera following the CIE TC1-65 [7] and approximately 1.5 meter away from the object (see Fig. 3). The light source approximated a point light source, as polishing and defects will be most visible when the light rays are closely originating from the same point since this will produce less diffuse reflections moving in various directions. A robotic

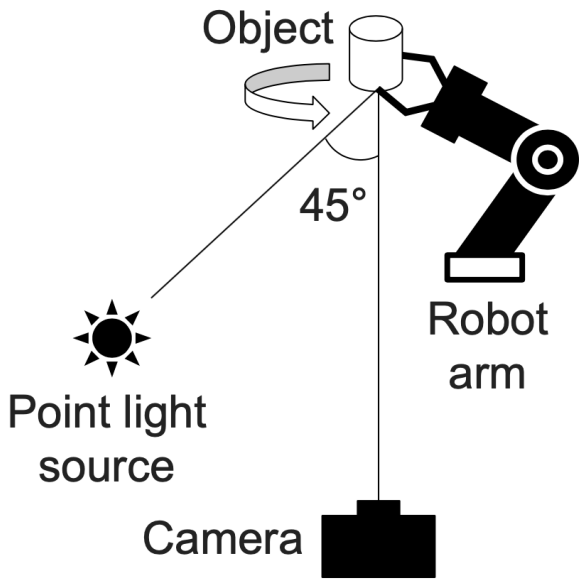

FIGURE 3. Leveraging the current practice of assessors rotating objects our camera setup consist of a camera, a point light source, and a robotic arm rotating the object.

arm (Universal Robots UR10e) held the brushed aluminium units upright (see Fig. 1 and 3) rotating the units around the yaw-axis. To replicate the human assessment process the rotation of the units were parallel to the polishing lines i.e. rotating around the yaw-axis. Rotation going in the same direction as the polishing is how the expert assessors checked the units during their visual inspection process. The robotic arm rotated the unit in the yaw-axis 1 degree between images for both front and back. Pre-test revealed that 1 degree rotation of the unit generated much image variance, and based on this we chose 1 degree as a first approximation for the resolution in rotation. Due to the curvature around the unit corner and the highly reflective surface the robot rotated the unit 5 degrees per shot when capturing the corner of each unit since these images were often overexposed. This resulted in 96 images per unit and a total of $96 \times 50 = 4800$ images. Images were later manually inspected were a few images got discarded due to motion blur etc.

B. DEFECT VISIBILITY IN RELATION TO HUMAN PERCEPTION

The camera resolution was 0.03 mm/pixel. 0.03 mm is equivalent to 0.21 arcminutes at a viewing distance of 0.5 meters (an arms length). For comparison the human eye has an ordinary visual acuity of about 1 arcminute. Therefore, defects equal to or less than one pixel are below the visual acuity of the human eye. There might be visible defects smaller than 1 pixel but they should hardly be recognised as shapes by assessors since we find our cameras angular resolution (0.21 arcminutes) smaller than the angular resolution of the human eye (1 arcminute) at a normal a viewing distance of 0.5 meters (an arms length). The polishing frequency of our surface was estimated using thread-counting and a

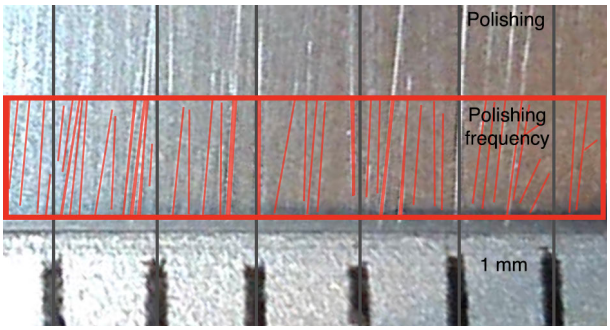

FIGURE 4. Polishing frequency under microscope. Based on thread-counting and pixel-wise counting our polishing strokes had a width between 0.04-0.14 mm.

microscope (see Fig. 4). This information was collected by sampling one unit through different angles of a surface (i.e. on 9 different images). With an average of 7 lines (individual polishing strokes) per millimeters we identify that 1 line $<$ 0.14 mm. To get a more precise estimate we also performed pixel-wise counting where we observed an average polishing stroke width of 6 pixels. With 1 mm $=$ 160 px we potentially had up to (160/6) 26 lines/mm thus an individual polishing stroke has a width of 0.04 mm. Thus, polishing strokes had a width between 0.04-0.14 mm. At a viewing distance of 0.5 meters the individual polishing strokes (0.04 mm $=$ 0.27 arcmin and 0.14 mm $=$ 0.96 arcmin) is near the ordinary acuity limit of 1 arcmin. Generally, surface polishing with high spatial frequencies (i.e. a very narrow polishing stroke width) must have significantly higher contrast than lines with lower spatial frequencies to be detected by the human visual system. Our polishing orientation had an estimated variation $\approx$ 20°($\pm$10°) and we expect high visual acuity for line defects oriented at either 0° or 90° (roughly parallel or perpendicular to the polishing strokes) than oblique angles [39]). Additionally, line defects oriented at 90° relative to the polishing should be easier to perceive than those following the lines of the polishing strokes according to the similarity principle in Gestalt theory applied to defect detection [2]).

C. GROUND TRUTH VALIDATION BY EXPERTS

Human expert assessors from Bang & Olufsen validated all defects on the physical surfaces. In general, defect candidates close to the quality threshold are what makes this evaluation process difficult. Defect candidates detectable by the human VA not necessarily entailed that the candidate was classified as a defect by the assessors. When expert assessors evaluated that the defect candidate was below the quality threshold the defect candidates were not annotated in our data set as a defects. After expert assessors evaluated all physical surfaces we mapped the assessors judgment onto the images based on good judgment.

VI. DEFECT TRACKING PIPELINE

The defect detection pipeline consist of edge enhancement, region of interest, tracking, and defect feature extraction

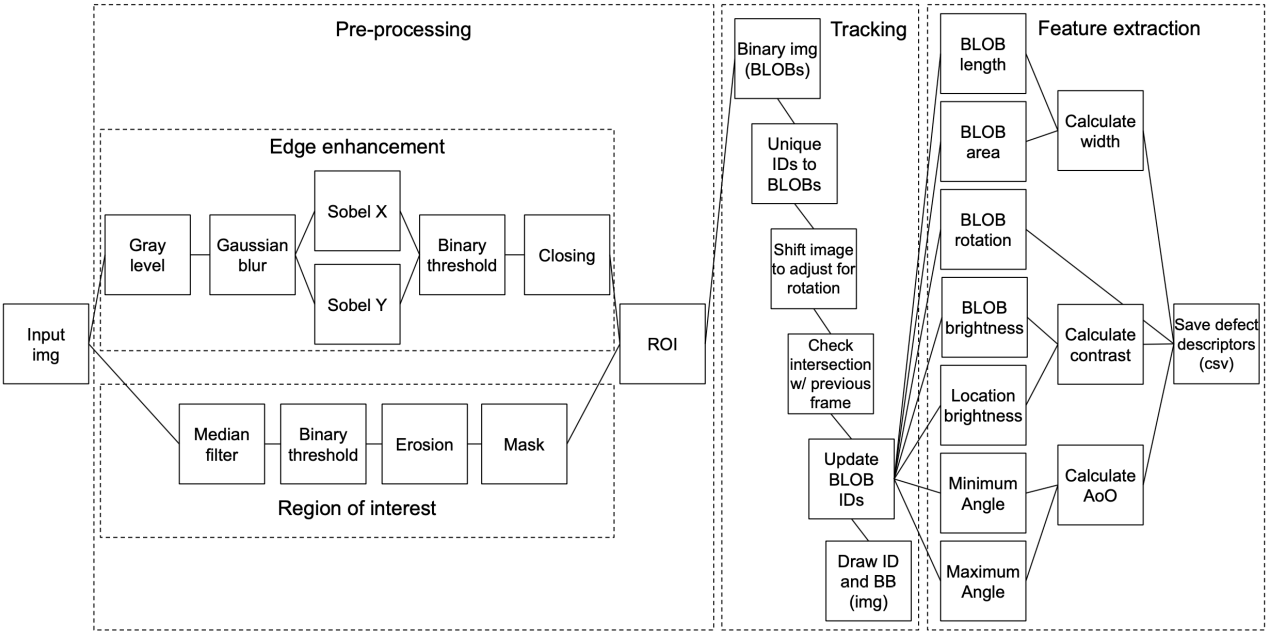

FIGURE 5. Pipeline for extracting features based on common defect descriptors (e.g. length) and spatio-temporal defect descriptors (e.g. AoO). The pipeline consist of 1) pre-processing including edge enhancement and region of interest, 2) tracking of all defect candidates and 3) extracting features based on set defect descriptors.

(see Fig. 5). The spatio-temporal defect detectors are based on rotation where we build the feature from tracking defect candidates in the captured image series. The inspected object consist of two flat surfaces connected by a round 90° corner but can be simplified as two flat surfaces. We limited our scope to detecting defects on flat surfaces where the light is fixed and even. Expanding our calculations to curved surfaces is possible but requires more complex estimations. As follows, we mask out the background including the border and the 90° corner of the objects and split the image series in two; front and back side.

A. EDGE ENHANCEMENT AND REGION OF INTEREST

Our pre-processing steps include edge enhancement and region of interest (ROI). We begin the edge enhancement operation by converting our RGB image into a gray level image (see Fig. 5, edge enhancement). Next a Gaussian blur is applied before using a vertical and horizontal Sobel filter. Different fixed binary thresholds are manually chosen based on the rotation angle of the object (i.e. the image number in the image series). We finalize the edge enhancement with morphological operations (i.e. closing; dilation, erosion) to remove noise and small holes. We cut out the ROI after edge enhancement to avoid noise around edges. First we applied a median filter on the input image as a noise reduction pre-processing step (see Fig. 5, region of interest). Subsequently, we use binary thresholding and perform a morphological erosion (to cut off the borders of the object) resulting in a mask used for cutting out our region of interest (the flat surfaces) in our output image from the edge enhancement.

B. TRACKING

Our input image consist of enhanced edges on the flat surface of the object for all frames in our image series. Using OpenCV [41] we find the contours of all Binary Large Objects (BLOBs) and assign unique IDs to all BLOBs in the current frame (see Fig. 5, tracking). We provide this temporary ID for all BLOBs in the frame which will later be updated if the area of the BLOB is overlapping with BLOBs in the previous frames. We assign unique IDs to all new BLOBs and try to match new BLOBs with previous BLOBs based on area intersection. We then perform image shifting on frames in our image series based on the 1° rotation of the object to match alignment with the previous frame. The next step of assigning IDs was not optimized as it checked for overlaps of all BLOBs in the current with all BLOBs in the previous frame (see Fig. 6); 1) if a BLOB overlaps with only one BLOB in the previous frame we assign the current BLOB with the same ID as in the previous frame, 2) if a BLOB in the current frame overlaps with several BLOBs from the previous one we check area intersection and the ID is then assigned based on the largest area intersection, 3) if a BLOB does not overlap with any BLOBs in the previous frame we maintain a unique ID. This will allow us to track all defects candidates on the rotated surface using a simple implementation of tracking.

C. FEATURE EXTRACTION

We provided a unique ID for all BLOBs in the frame and calculate the polygon shape including features as area, length, etc. (see Fig. 5/feature extraction and Fig. 7). The area is calculated from the polygon shape approximating the BLOB.

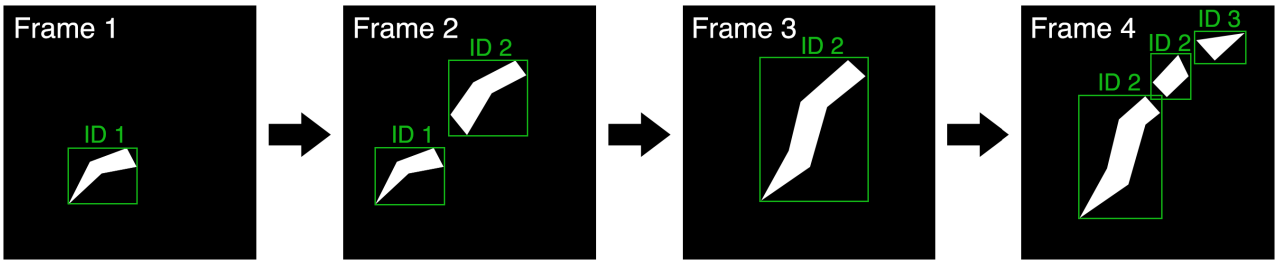

FIGURE 6. Tracking and assigning unique IDs to BLOBs. All new BLOBs in a frame are provided with a unique ID (Frame 1). The current frame (Frame 2) is compared to the previous frame and IDs will be assigned to match the previous frame. If several BLOBs overlap IDs will be assigned based on the largest area intersection (Frame 3). Moving forward several BLOBs can be assigned with the same ID even though they are no longer connected (Frame 4).

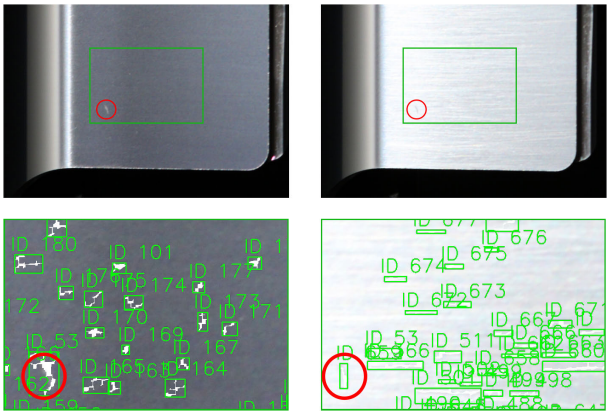

FIGURE 7. Polishing has vague visibility (left) but when light was reflected at the right view angle polishing becomes visible (right). Most of the false positives came from the majority of images where polishing had reduced visibility (left) and consequently the false positives were on average smaller in length, width and area. The defect is marked with a red circle and (in this case) clearly has an orientation perpendicular to the polishing strokes.

The length is given by the largest distance between any two points in the polygon. The rotation is calculated from the slope between these two points. The width is calculated as area/length. The contrast is calculated based on the difference between the brightness of the BLOB and the brightness of the area surrounding the BLOB. We draw out all IDs and their bounding boxes to visualise our results. We finalize the pipeline by saving all features (BLOB IDs, length, area, rotation, and contrast) in a text file.

VII. DEFECT DESCRIPTORS

The data set consist of 395 true positives (i.e. individually tracked defects) and 14.017 false positives (i.e. non-defects as individually tracked design features, anomalies below the quality threshold, noise etc.). Whenever we loose a tracklet in our tracking pipeline we introduce a new tracklet for the same defect candidates, thus, these numbers (395/14.017) are artificially higher than the total number of defects in the data set.

For all tracked defect candidates we extract the defect descriptors; length, width, area, contrast, rotation, AoO, minimum angle and maximum angle. Scratches typically represents a line (i.e. they are long and thin) hence related to the length of the defect candidates. The length of all defect candidates is measured in pixels. With a camera resolution of 0.03mm/pixels this can be converted into millimeters (mm) and arc minutes (arcmin). The area is measured in pixels and the width is approximated based on $width = area/length$. The contrast is calculated as spatial contrast in cycles per degree (cpd) and the Weber contrast. Weber contrast is given by $C = (I - I_b)/I_b$. Traditionally I represents the luminance but we calculate the contrast based on the measured lightness (L) in the CIE L*a*b* color space. The rotation is defined as the deviation from the polishing pattern. The polishing pattern was horizontal with an estimated variance of $\pm$ 30 degrees. Based on our defect tracking we compute the angle of opportunity (AoO) i.e. how many degrees a potential defect candidate is visible. With a rotation of 1° between each image capture each image counts 1 degree ($\pm$ 0.5°). The minimum and maximum angle corresponds to the minimum and maximum rotation of the object surface. When the angle equals 0° the surface is orthogonal to the view direction. The AoO is then calculated based on the difference between the minimum and maximum angle where the defect candidate is visible.

For each defect descriptor we provided the set quality limit i.e. the threshold between non-defects and defects demanding product rejection (see Table 1). We based the limit of ordinary acuity on normal vision (20/20 or 6/6 = 1.0 acuity). As visual acuity has high variance and depends on several factors (e.g. age) we let our ordinary acuity limit act as a rule of thumb and recognize that the limit is not uniform across individuals. The defect minimum was the lowest measured score per defect descriptor measured in our complete data set. The defect and non-defect average was calculated based on the mean value per defect descriptor. The non-defects consist of all defects below the quality limit, brush strokes from polishing and noise in the edge-enhanced images. Most data captured presents the surface with diffuse reflection. Non-defects on average had a shorter width (0.43 arcmin) compared to defects (0.50 arcmin) and are close to the quality limit (0.34 arcmin) though a bit under the ordinary acuity limit (1.00 arcmin) at the calculated view distance. The length

TABLE 1. Quality limits, limits for ordinary acuity and data breakdown. Numbers in parenthesis are based on single measures that are repeated to calculate an estimated spatial contrast. The second line in the rows displays the standard deviation (SD).

	Length (arcmin)	Width (arcmin)	Area (mm2)	Spatial contrast (cpd)	Weber contrast	Rotation (deg)	AoO (deg)	Minimum angle (deg)	Maximum angle (deg)
Quality limit	68.76	0.34	0.40	-	-	-	-	-	-
Ordinary acuity	1.00	1.00	0.02	30	0.01	-	-	-	-
Defects									
minimum	4.47	0.22	0.03	(40)	0.00006	0.00	1.00	-	-
mean	38.34	0.50	0.45	(17.45)	0.08	53.60	3.56	20.03	22.60
SD	5.24	0.03	0.55	-	0.16	55.94	4.52	12.69	12.31
Non-defects									
mean	28.20	0.43	0.27	(60)	0.03	48.92	1.11	16.86	16.97
SD	2.50	0.03	0.39	-	0.12	43.22	0.49	10.81	10.84

of defects (38.34 arcmin, SD: ±5.24) was similarly longer than non-defect (28.20 arcmin, SD: ±2.50). In most images the surface reflection was diffuse and the non-defect were short since polishing strokes were less visible. However, in those view angles where specular reflection was high and the individual polishing strokes are clearly visible those individual strokes are typically longer than a common scratch (see Fig. 7). The length of the non-defects were either very short (noise, defect candidates below the quality limit etc.) or very long (polishing strokes). On average defects had a larger valued length, width, area, Weber contrast, rotation, AoO, minimum angle, maximum angle than non-defects (see Table 1). On average defects are most visible when the surface normal is rotated between 20°-23° (±13°) away from the view direction. The AoO is on average 3.56° and thus visible for 2.45° more than non-defects.

VIII. RESULTS

We present a comparison of different classification models and evaluate their performances. Subsequent we inspect the performance of the defect descriptors with individually trained classifiers. We use our defined defect descriptors as features for our classification models and compute the feature importance based on mean decrease accuracy for our best performing model including all features. Based on the feature importance we rank the features and evaluate the top ranked features for feature selection. We investigate different data splits, review the differences in the data sets and evaluate the performances. In conclusion we assess feature interpretation using a single decision tree to explain the classification cut-off limits to better understand the relation to our quality and ordinary acuity limit.

A. MODEL COMPARISON

Classifying defect vs. non-defects constitute a binary classification problem. The Receiver Operating Characteristics (ROC) curve evaluates model performance across all possible thresholds and the model ROC curve is quantified as a single metric; the Area Under the Curve (AUC) [42]. A large AUC indicates that a high sensitivity (true positive rate) and specificity (true negative rate) can be achieved concurrently. We tested four different classification models;

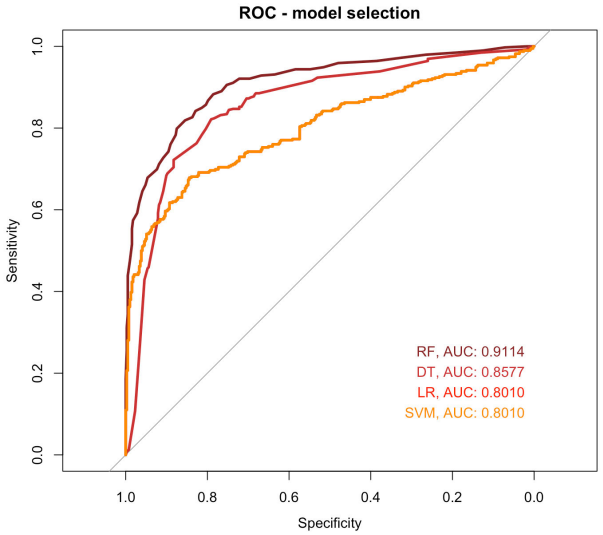

FIGURE 8. ROC curve for model comparison: Random Forest (RF, AUC = 0.91), Decision Tree (DT, AUC = 0.86) Logistic Regression (LR, AUC = 0.82), and Support Vector Machine (SVM, AUC = 0.80).

TABLE 2. Model comparison: The Random Forest classifier provides the best performance with an accuracy of 83% and an AUC of 91%.

	RF	DT	LR	SVM
Accuracy (%)	**83.16**	80.61	77.68	74.74
AUC (%)	**91.14**	85.77	82.19	80.10

Random Forest (RF) using 50 trees, Decision Tree (DT), Logistic Regression (LR) and Support Vector Machine (SVM). Given our binary classification problem (defect vs. non-defects) we balanced the data set and performed 5-fold cross validation. We plotted the ROC curve for our four different models (see Fig. 8) and evaluated their performances. The RF classifier provided the best performance with an accuracy of 83.16% and an AUC of 91.14% (see Table 2).

B. SINGLE FEATURE PERFORMANCE

Based on our eight defect descriptors (AoO, area, length, width, rotation, contrast, minimum angle and maximum angle) we plotted the ROC-curve per individual feature

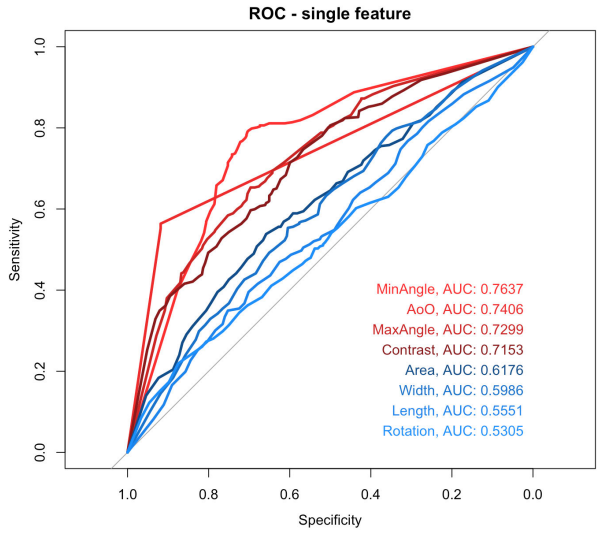

FIGURE 9. ROC curve for single features: AoO, maximum angle, minimum angle, rotation, width, area, and length.

(see Fig. 9). The performance for the single features in chronological order were: minimum angle (AUC = 76.37), AoO (AUC = 74.06), maximum angle (AUC = 72.99), contrast (AUC = 71.53), area (AUC = 61.76), width (AUC = 59.86), length (AUC = 55.51), and rotation (AUC = 53.05). Based on individual features the minimum angle provided the best classification performance and alone rotation produces the worst performance. The AoO, minimum angle and maximum angle are linear depended and thus one feature can be explained as a linear combination of the other two.

C. FEATURE IMPORTANCE AND FEATURE SELECTION

Based on the out-of-bag (OOB) error (i.e. the average prediction error calculated of each training sample using predictions only from the trees not containing the same data as the respective training sample) of the RF classifier we calculated the scaled mean decrease in accuracy (sMDA) (see Fig. 10). The sMDA was computed for all our features as

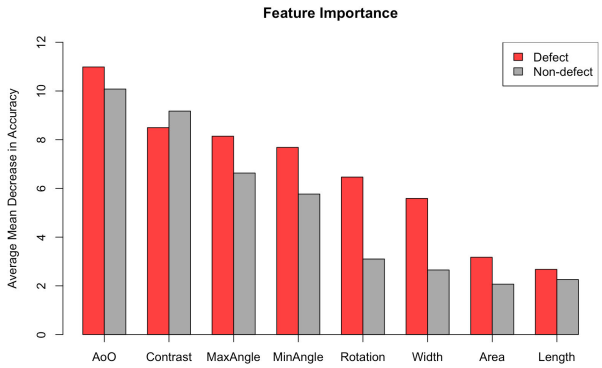

FIGURE 10. Based on the sMDA the feature importance order for defects is; AoO, contrast, maximum angle, minimum angle, rotation, width, area, and length. The feature importance order is very similar for non-defects with only the two least important features being swapped.

sMDA = mean difference between trees/standard deviation of the differences [43]. We trained all trees (50) in the RF based on the balanced data set ($n \approx 800$). We then compared the accuracy of the OOB samples with the accuracy of OOB samples where one feature is permuted [44]. The feature permutation was repeated for all features and was split between the classes *defect* and *non-defect*. For each tree we measured the difference in accuracy with and without feature permutation. We identified that AoO has the largest sMDA for both defects (sMDA = 10.99) and non-defects (sMDA = 10.08) where large sMDA scores represents that the feature had substantial importance and a consistent decrease in accuracy. Based the sMDA (using RF, 50 trees) we concluded the feature importance in the order; AoO, contrast, maximum angle, minimum angle, rotation, width, area, and length.

Based on the computed sMDA feature importance we compared RF classifiers trained on the top n features (see Fig. 11). Using the top three features (AoO, contrast and maximum angle) we observed similar performance (Top 3, AUC = 89.50) to training an RF using the top eight features (Top 8/all features, AUC = 91.14). Using only a RF trained on the top two features we noticed a cutback in performance (Top 2, AUC = 80.58). We concluded that the feature space can be reduced to include only the top three features with a limited reduction in classification performance.

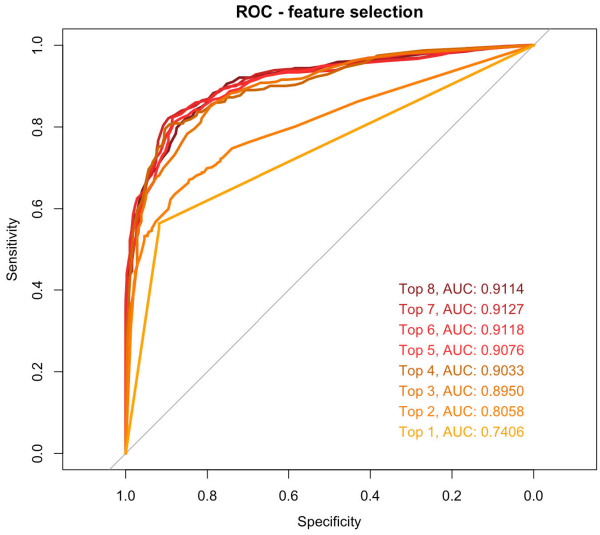

FIGURE 11. ROC curve for feature selection based on the feature importance. Top 8 includes all features. Top 7 includes the top seven features based on their feature importance (sMDA) etc.

D. DATA SPLIT BASED ON REFLECTED LIGHT

The visual appearance, and consequently the visual appearance of our features, varied depended on view angle (see Fig. 1). From our data we observed that polishing was visible only when we had high specular reflection in the images, and thus our non-defects drastically changed appearance when polishing was visible i.e. the non-defects were on average smaller when the surface was diffuse versus

non-defects were longer and had larger areas with higher specular reflection (see Fig. 7). Based on the surface view angle we hypothesize that we could split the data set in two sets; diffuse and specular. We defined the specular data set based on the appearance of individual polishing strokes and an increase in the overall brightness of the image for a 5° angle. The diffuse data set constitute the remaining data where the polishing is not clearly visible. We trained a RF classifier on the full data set and compared performance to a RF classifier trained on the diffuse data and the specular data separately. Our data sets were balanced and performance was measured using 5-fold cross validation. We transfer the already obtained information about AoO across the different data sets to keep the effect of AoO constant. We get the best performance from the diffuse data set with an accuracy = 84.38% and AUC = 91.64% closely followed by the full data set with an accuracy = 83.16% and AUC = 91.14% (see Fig. 12 and Table 3). From the specular data set we see a decrease in performance (accuracy = 80.31% and AUC = 85.96%). Since data splits are not of equal size they are not directly comparable and thus the performance reduction for the specular data set may be due to the reduction in data. Our hypothesis is that the features (especially for non-defects) are different between our two new data splits: specular and diffuse. Based on calculated sMDA scores from the 5-fold cross validated models we compute the ranking (from 1-8, 1 having the highest sMDA and 8 the lowest sMDA) for all features. We average the ranking across the different cross validated models for each data set *diffuse, specular* and *full* (see Table 4). For the full and diffuse data set we recognise similar ranking with ascending order of AoO (ranking between R: 1.0-1.6), contrast (R: 1.6-3.4) and maximum angle (R: 1.8-3.2). For the specular data see that the ranking order have highest rank for maximum angle (R: 1.8-2.0) minimum angle (R: 1.6-4.0), and rotation (R: 2.4-3.2). This means that different features become important when training a classifier on the specular data set.

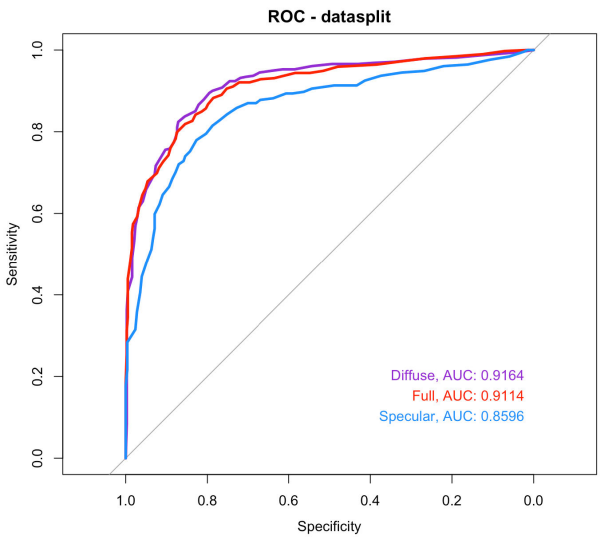

FIGURE 12. ROC curves for the different datasplits: diffuse, specular, and full data set. The full (AUC = 91.14) and diffuse (AUC = 91.64) data set have similar performance whereas the specular data set have decreased performance (AUC = 85.96).

TABLE 3. Data Splits: The diffuse data split provides the best performance with an accuracy of 84% and an AUC of 92%.

	Diffuse	Full	Specular
Accuracy (%)	**84.38**	83.16	80.31
AUC (%)	**91.64**	91.14	85.96

TABLE 4. Feature ranking based on the calculated sMDA scores from the 5-fold cross validated models we calculate the average ranking between 1-8. The text is formatted to present the ***best rank***, second best rank, ***third best rank*** and the remaining data.

RANKING	AoO	C	Max	Min	R	W	A	L
Diffuse								
defect	***1.4***	*3.4*	**1.8**	*3.4*	5.6	6.2	7.2	7.0
non-defect	***1.4***	**1.8**	*3.0*	4.6	7.2	6.8	6.2	5.0
total	***1.6***	***1.6***	**3.2**	*3.6*	7.4	6.2	6.6	5.8
Specular								
defect	**2.8**	5.0	***2.2***	*4.0*	**2.8**	5.8	7.0	6.4
non-defect	4.0	5.0	***1.8***	**2.0**	*2.4*	7.2	7.8	5.8
total	3.4	5.4	**2.0**	***1.6***	*3.2*	6.8	7.6	6.0
Full								
defect	***1.0***	**2.6**	*3.0*	3.6	5.0	5.8	7.4	7.6
non-defect	***1.4***	**1.6**	*3.2*	3.8	6.0	6.2	7.2	5.0
total	***1.4***	**2.4**	3.2	*3.0*	5.0	6.2	7.2	7.6

E. FEATURE INTERPRETATION

To better understand and explain the features we simplified the model to a decision tree using only the top three features; AoO, contrast and maximum angle. The DT work as an example and provide an explanation of the features involved in defect detection on brushed aluminium surfaces. We used the classifier on a balanced a data set with a 80/20 for train/test split (see Fig. 13). In the decision tree and AoO of 1.5° yielded an 87% defect probability. We identified that an AoO above or equal to 1.5° yielded a probability of a defect of 87% with 32% of all data going to this node (node 2, Fig. 13). A contrast above or equal to 0.0057 produced a 92% probability of a defect and thus classified 29% of the data as defects. The remaining approximately 20% (node 10-1%, node 24-5%, node 100-3%, node 202-7%, node 28-4%) of defects are classified from a mixed contribution from the features AoO, contrast, and maximum angle. The nodes displaying thresholds with the maximum angle (where maximum angle = 0 is equal to a surface orthogonal to the viewing direction) imply that images when the surface rotation is 30-40 degrees have influence on the predictions determining defects from non-defects.

IX. DISCUSSION

Based on our data of brushed aluminium surfaces defects had an average AoO of 3.56° (SD: ±4.52) (see Table 1). Non-defects had a lower average AoO of 1.11° (SD. ±0.49). Based on the decision tree trained on the top three features an

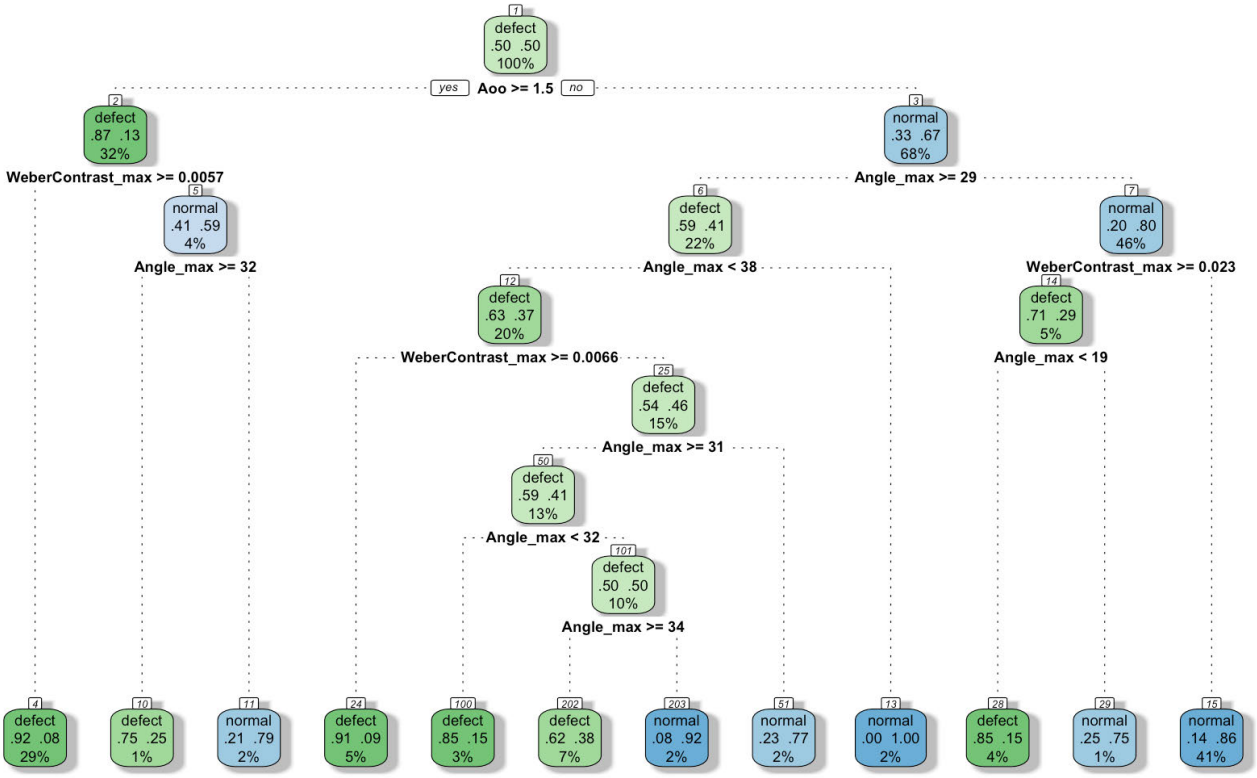

FIGURE 13. Using only the top three features (AoO, contrast and maximum angle) we use a Decision Tree to provide an explanation of the effects of the features involved in defect detection.

AoO of 1.5° yielded a 87% defect probability in node 2 with 32% of the data going through this node (see Figure 13). This means that defects had a higher AoO than non-defect or in other words, defect were visible for a larger total subtended angle. Designers of a semi-automated system in this case would need a resolution in rotation of 1° or higher when capturing data in order to use AoO as a defect descriptor. Especially highly reflective materials with brush strokes requires units to be inspected from various view angles but previous approaches regularly captured data from only a single viewpoint [14], [15], [17], [19], [26] and generally lacked multi-view assessment of defects taking advantage of the spatio-temporal domain [20]. Capturing single viewpoint means that defects only visible in certain view angles will not guarantee being captured and the number of false positives for a single frame can be numerous (see Figure 7). All features (AoO, contrast, etc.) can be used for defect detection (see Figure 9). Using these defect descriptors in similar defect detection tasks such as on transparent materials (e.g. glass) or non-polished materials (e.g. plastic) might improve classification results and should be investigated in future research. However, limiting the scope to scratches could be a reason why we see good results based on ROC-curve and accuracy, while we see mediocre results in a previous study applying AoO and deep learning for the detection task on similar data with the same brushed material (precision-recall curve of 67%) [20]. Defect detection on other materials have provided better result (accuracy of 89% for detection of defects on highly reflective ring components [27] compared to our accuracy of 84%), but the considered surfaces differ in visual appearance of both defect and material properties. We argue that an automated system assisting human assessors in low-volume premium production (where *all* defect candidates were thoroughly evaluated) should afford (capturing images, setting quality thresholds etc.) identifying all true positives, even though this entail a large number of false positives. In conclusion, leveraging the current practice of assessors rotating objects we investigated a spatio-temporal image capturing setup. Our pipeline construct defect descriptors such as AoO, which were less depended on single-image appearance but instead was a measure of the subtended angle during which a defect was visible.

Our simple pipeline has limitations and tracking was not optimized for fast processing. Our tracking pipeline provided unique IDs to all defect candidates and updated the IDs based on area intersection from the previous frame. However, sometimes we loose tracking, several BLOBs overlaps, or several BLOBs constitute only one defect candidate but our approach force assigning unique IDs to all new BLOBs (see Fig. 6). Since our pipeline did not track defects consistently, whenever we lost track of a defect candidate or assigned IDs to ambiguous defect candidates (see Fig. 6) this influenced our measured AoO. This results in creating more defect candidates with smaller AoO since the defects are not connected. While performance was high (our trained RF classifier had a performance of AUC = 91.14) we expect even better results

if tracking is further improved since we from our data judge the true positives (defects) to have longer AoO than false positives (non-defects). We expect even better results since improved tracking will make the difference between true positives and false positives larger.

Defects had higher values for the measured defect descriptors: length, width, area, rotation and contrast (see Table 1). On average defect were bigger than non-defects and defects were visible for longer than non-defect (potentially due to the fact the defects were bigger). The spatial contrast was lower for defects than for non-defects and shows that when the defect width is shorter the estimated spatial frequency becomes higher. The minimum and maximum angle represented the rotation of the surface (i.e. surface normal) in comparison to the illumination and view angle, where we identified an average of 21° rotation (min.: 20.03, max.: 22.60) for defects with an averaged standard deviation of ±13°. We recommend placing the light source at a 45° angle. Based on our data we recommend placing a light source at a 45° angle compared to the view direction. Line defects such as scratches are often captured at an object surface rotation of a 21° (±13°) whereas non-defects is mostly visible at a 17° rotation (±11°). The illumination caused varying appearance of defects and non-defect (see Fig. 7) and consequently a high variance in the common defect descriptors such as length, width and contrast (see Table 1, standard deviation). Our analysis suggest that we can benefit from using spatio-temporal data. Based on the analysis of our image series we identified a difference in feature importance between specular and diffuse data (see Table 4). For images with diffuse reflection the AoO, contrast and maximum angle had the best ranking based on sMDA and thus the highest importance when classifying defects. The full and diffuse data set had similar performance (Full: AUC = 91.14 versus Diffuse: AUC = 91.64) but performance decrease for the specular data split (Full: AUC = 85.96). We had a minor increase in performance (AUC: +0.5) when using only the diffuse data set. This could argue for improving performance by excluding images that had high specular reflection since the specular reflective images contain many false positives in the form of long polishing strokes. This could prove effective when detecting common defect types as scratches. However, the total measured AoO will be shorter since the defects will not be tracked in the excluded frames. In other words, by excluding the specular data we can expect fewer false positives but we loose information in the form of; a) images potentially including rare defect types (e.g. stripes or crazing) only reflecting light in frames with high specular reflection thus not being detected at all and b) we reduce our best predictor AoO.

X. CONCLUSION

We provide evidence for the value of spatio-temporal defect descriptors as AoO (the total subtended angle during which defects are visible) in classification of defects on low-volume premium products, specifically for line defects on brushed aluminium surfaces. In this setup, we constructed AoO using a bespoke tracking pipeline using edge enhancement and defect tracking, which allowed for extracting other defect descriptors such as contrast, area, width, length, and rotation. AoO is found when the setup mimics the human visual inspection and was applied in our context of defect detection on brushed aluminium surfaces. Out of four different models, Random Forest had the highest AUC of 91%. We could reduce the number of features to the top three (AoO, contrast and maximum angle) with a performance AUC of 90%. For defect detection on brushed and reflective surfaces we recommend capturing different angles in the order of single degrees rotation or lower dependent on granularity of polishing. Capturing spatio-temporal data allowed for high classification performance since AoO, as a defect descriptor, depends less on a defect's appearance from a single view angle and can therefore be combined with other defect descriptors to make classification more robust.

ACKNOWLEDGMENT

Special thanks to Bang & Olufsen.

REFERENCES

[1] Y.-L. Deng, S.-P. Xu, and W.-W. Lai, "A novel imaging-enhancement-based inspection method for transparent aesthetic defects in a polymeric polarizer," *Polym. Test.*, vol. 61, pp. 333–340, Aug. 2017.

[2] A. J. Hansen, H. Knoche, and T. B. Moeslund, "Getting crevices, cracks, and grooves in line: Anomaly categorization for AQC judgment models," in *Proc. 10th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, May 2018, pp. 1–3.

[3] Z. Gosar and D. P. Gruber, "In-line quality inspection of freeform plastic high gloss surfaces aided by multi-axial robotic systems," in *Proc. Int. Electrotech. Comput. Sci. Conf.*, 2017, pp. 445–448.

[4] R. W. Fleming, "Visual perception of materials and their properties," *Vis. Res.*, vol. 94, pp. 62–75, Jan. 2014.

[5] T. Levitt, "Production-line approach to service," *Harvard Bus. Rev.*, vol. 50, no. 5, pp. 41–52, 1972.

[6] European Committee for Standardization, *Geometrical Product Specification (GPS)—Surface Imperfections—Terms, Definitions and Parameters*, Standard EN ISO 8785, 1999

[7] M. Pointer, "CIE TC1-65—A framework for the measurement of visual appearance," Paris, France, CIE Publication 175-2006, 2006, pp. 1–100, vol. 5.

[8] *Standard Test Method for Specular Gloss*, Standard ASTM D523-14, ASTM International, West Conshohocken, PA, USA, 2014, vol. 06.01.

[9] International Organization for Standardization, *ISO 2813: Paints and Varnishes–Determination of Gloss Value at 20 Degrees, 60 Degrees and 85 Degrees*, Standard International Organization for Standardization 2813, 2014.

[10] C. Eugène, "Measurement of 'total visual appearance': A CIE challenge of soft metrology," in *Proc. 12th IMEKO TC1 TC7 Joint Symp. Man, Sci. Meas.*, 2008, pp. 61–65.

[11] L. Zsíros, A. Suplicz, G. Romhány, T. Tábi, and J. G. Kovács, "Development of a novel color inhomogeneity test method for injection molded parts," *Polym. Test.*, vol. 37, pp. 112–116, Aug. 2014.

[12] J.-L. Maire, M. Pillet, and N. Baudet, "Measurement of the perceived quality of a product: Characterization of aesthetic anomalies," *Int. J. Metrol. Qual. Eng.*, vol. 4, no. 2, pp. 63–69, 2013.

[13] P. Gospodnetić and F. Hirschenberger, "Detection and visibility estimation of surface defects under various illumination angles using bidirectional reflectance distribution function and local binary pattern," in *Proc. Croatian Comput. Vis. Workshop*, Dec. 2016, pp. 9–14.

[14] D.-M. Tsai, M.-C. Chen, W.-C. Li, and W.-Y. Chiu, "A fast regularity measure for surface defect detection," *Mach. Vis. Appl.*, vol. 23, no. 5, pp. 869–886, Sep. 2012.

[15] S. Chen, B. Lin, X. Han, and X. Liang, "Automated inspection of engineering ceramic grinding surface damage based on image recognition," *Int. J. Adv. Manuf. Technol.*, vol. 66, nos. 1–4, pp. 431–443, Apr. 2013.
[16] J. J. Liu and J. F. MacGregor, "Estimation and monitoring of product aesthetics: Application to manufacturing of 'engineered stone' countertops," *Mach. Vis. Appl.*, vol. 16, no. 6, p. 374, 2006.
[17] D. P. Gruber, J. Macher, D. Haba, G. R. Berger, G. Pacher, and W. Friesenbichler, "Measurement of the visual perceptibility of sink marks on injection molding parts by a new fast processing model," *Polym. Test.*, vol. 33, pp. 7–12, Feb. 2014.
[18] L. A. O. Martins, F. L. C. Padua, and P. E. M. Almeida, "Automatic detection of surface defects on rolled steel using computer vision and artificial neural networks," in *Proc. 36th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Nov. 2010, pp. 1081–1086.
[19] C. Mera, M. Orozco-Alzate, J. Branch, and D. Mery, "Automatic visual inspection: An approach with multi-instance learning," *Comput. Ind.*, vol. 83, pp. 46–54, Dec. 2016.
[20] A. J. Hansen, M. P. Phillipsen, H. Knoche, and T. B. Moeslund, "Machine vision for aesthetic quality control of reflective surfaces," in *Proc. Int. Conf. Artif. Intell. Comput. Vis.*, 2021, pp. 389–401.
[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Apr. 2015, doi: 10.1007/s11263-015-0816-y.
[22] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, Oct. 2019, doi: 10.1007/s11263-019-01247-4.
[23] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," 2020, *arXiv:2001.05566*. [Online]. Available: http://arxiv.org/abs/2001.05566
[24] P. Dendorfer, A. Osep, A. Milan, K. Schindler, D. Cremers, I. Reid, S. Roth, and L. Leal-Taixé, "MOTChallenge: A benchmark for single-camera multiple target tracking," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 845–881, Dec. 2020, doi: 10.1007/s11263-020-01393-0.
[25] G. Rosati, G. Boschetti, A. Biondi, and A. Rossi, "Real-time defect detection on highly reflective curved surfaces," *Opt. Lasers Eng.*, vol. 47, nos. 3–4, pp. 379–384, Mar. 2009.
[26] Z. Zhang, B. Li, W. Zhang, R. Lu, S. Wada, and Y. Zhang, "Real-time penetration state monitoring using convolutional neural network for laser welding of tailor rolled blanks," *J. Manuf. Syst.*, vol. 54, pp. 348–360, Jan. 2020.
[27] R. A. Boby, P. S. Sonakar, M. Singaperumal, and B. Ramamoorthy, "Identification of defects on highly reflective ring components and analysis using machine vision," *Int. J. Adv. Manuf. Technol.*, vol. 52, nos. 1–4, pp. 217–233, Jan. 2011.
[28] C. Hu and Y. Wang, "An efficient convolutional neural network model based on object-level attention mechanism for casting defect detection on radiography images," *IEEE Trans. Ind. Electron.*, vol. 67, no. 12, pp. 10922–10930, Dec. 2020.
[29] J. Günther, P. M. Pilarski, G. Helfrich, H. Shen, and K. Diepold, "Intelligent laser welding through representation, prediction, and control learning: An architecture with deep neural networks and reinforcement learning," *Mechatronics*, vol. 34, pp. 1–11, Mar. 2016.
[30] C. Hofmann, F. Particke, M. Hiller, and J. Thielecke, "Object detection, classification and localization by infrastructural stereo cameras," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (VISAPP)*, 2019, pp. 808–815.
[31] T.-H. Sun, F.-C. Tien, F.-C. Tien, and R.-J. Kuo, "Automated thermal fuse inspection using machine vision and artificial neural networks," *J. Intell. Manuf.*, vol. 27, no. 3, pp. 639–651, Jun. 2016.
[32] X. Xie, "A review of recent advances in surface defect detection using texture analysis techniques," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 7, no. 3, pp. 1–22, 2008.
[33] T. Czimmermann, G. Ciuti, M. Milazzo, M. Chiurazzi, S. Roccella, C. M. Oddo, and P. Dario, "Visual-based defect detection and classification approaches for industrial applications—A SURVEY," *Sensors*, vol. 20, no. 5, p. 1459, Mar. 2020.
[34] H. El-Mesery, H. Mao, and A. Abomohra, "Applications of non-destructive technologies for agricultural and food products quality inspection," *Sensors*, vol. 19, no. 4, p. 846, Feb. 2019.
[35] E. Verna, G. Genta, M. Galetto, and F. Franceschini, "Planning offline inspection strategies in low-volume manufacturing processes," *Qual. Eng.*, vol. 32, pp. 1–16, Oct. 2020.
[36] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec Ad—A comprehensive real-world dataset for unsupervised anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9592–9600.
[37] C. G. Healey and A. P. Sawant, "On the limits of resolution and visual angle in visualization," *ACM Trans. Appl. Perception*, vol. 9, no. 4, pp. 1–21, Oct. 2012.
[38] M. D. Fairchild, *Color Appearance Models*. Hoboken, NJ, USA: Wiley, 2013.
[39] S. Appelle, "Perception and discrimination as a function of stimulus orientation: The 'oblique effect' in man and animals," *Psychol. Bull.*, vol. 78, no. 4, p. 266, 1972.
[40] W. Li and G. Westheimer, "Human discrimination of the implicit orientation of simple symmetrical patterns," *Vis. Res.*, vol. 37, no. 5, pp. 565–572, Mar. 1997.
[41] G. Bradski, "The OpenCV library," *Dr Dobb's J. Softw. Tools*, vol. 25, pp. 120–125, 2000.
[42] V. Bewick, L. Cheek, and J. Ball, "Statistics review 13: Receiver operating characteristic curves," *Crit. Care*, vol. 8, no. 6, pp. 1–5, 2004.
[43] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[44] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/a:1010933404324.

ANNE JUHLER HANSEN received the M.Sc. degree in medialogy with a specialization in computer graphics from Aalborg University, in 2016, where she is currently pursuing the Ph.D. degree in aesthetic quality control with the Visual Analysis and Perception Laboratory. From 2016 to 2017, she worked as a Research Assistant with the Department of Architecture, Design and Media Technology, Aalborg University. Her research interests include working with perceptual models, aesthetic quality control using machine vision, and combining these in an industrial context.

THOMAS B. MOESLUND received the Ph.D. degree from Aalborg University, Denmark, in 2003. He is currently the Head of the Visual Analysis and Perception Laboratory, the Head of Section for Media Technology, and the Head of AI for the People Center at Aalborg University. He has published more than 300 books, journals articles, and conference papers, and he has been cited 13 000 times. His overall research interests include building intelligent systems that make sense out of data with a special focus on computer vision and AI. His H-index is 43. His awards include the Most Cited Paper Award, in 2009, the Teacher of the Year Award, in 2010, the Innovation Award, in 2013, and the seven best paper awards.

HENDRIK KNOCHE is currently an Associate Professor with the Department of Architecture, Design and Media Technology, University of Aalborg, Denmark. He has worked on interaction design and user experience both on desktop and mobile devices for more than 15 years both in industry and academia in various domains including games, mobile social networking, automated content adaptation, perceived video, and aesthetic quality. His larger research interests include human-centered design, mediated experiences, and ICT for inclusion and development along with methods for modelling users, prototyping, evaluating applications, and their user experiences in real life contexts.