

UNIVERSITY POLITEHNICA OF BUCHAREST
FACULTY OF AUTOMATIC CONTROL AND COMPUTERS
COMPUTER SCIENCE AND ENGINEERING DEPARTMENT



DIPLOMA PROJECT

Automatic Tire-markings Detection & Recognition

Robert-Mihai Lică

Thesis advisor:

Conf. dr. ing. Iuliu Vasilescu

Conf. dr. ing. Daniel Rosner

BUCHAREST

2022

UNIVERSITATEA POLITEHNICA DIN BUCUREŞTI
FACULTATEA DE AUTOMATICĂ ŞI CALCULATOARE
DEPARTAMENTUL DE CALCULATOARE



PROIECT DE DIPLOMĂ

Detectarea și Recunoașterea Automată a Marcajelor de pe
Cauciucuri

Robert-Mihai Lică

Coordonator științific:

Conf. dr. ing. Iuliu Vasilescu

Conf. dr. ing. Daniel Rosner

BUCUREŞTI

2022

CONTENTS

1	Introduction	1
1.1	Motivation	1
1.2	Problem Statement	2
1.3	Overview of the Solution	3
1.4	The Document's Structure	4
2	State of the Art	5
3	The Proposed Solution	7
3.1	Tire Unwrapping	7
3.1.1	Circle Detection	7
3.1.2	Convert to Polar Coordinates	13
3.2	Text Detection	14
3.2.1	Segmentation	15
3.2.2	Segment Binarization	16
3.2.3	Filtering Artifacts	19
3.2.4	Voting	21
3.3	Text Recognition	23
4	Implementation Details	25

ABSTRACT

This work shall provide an image processing pipeline for automatic detection and recognition of tire-markings from images. Because tires can be one of the most expensive consumables of a truck, they are of interest to fleet managers who want to monitor their placement, parameters and possible illegal swap. My solution shall provide an easy way to extract from an image the serial code and/or certification number in order to aid an individual to almost uniquely identify a tire. The system can identify a wheel in an image and unwraps it. Then, it detects the text regions on the tire by using a band-pass filter on the frequency domain of the image, and lastly Microsoft Cognitive Services' OCR is employed to perform the character recognition.

SINOPSIS

Această lucreare o să ofere o bandă de procesare a imaginilor pentru detectarea automată și recunoașterea marcajelor de pe marginea unui cauciuc. Deoarece cauciucurile pot fi unul dintre cele mai scumpe consumabile ale unui tir, este în interesul managerilor de flote să monitorizeze utilizarea cauciucurilor, parametrii lor și eventuale schimbări ilegale. Soluția mea va oferi o metodă ușoară de a extrage din imagini codul de serie și/sau numărul certificării pentru a identifica aproape unic un cauciuc. Sistemul poate să identifice o roată într-o imagine și să o desfășoare. Apoi, aplică un filtru bandă pe domeniul de frecvență a imaginii și OCR-ul de la "Microsoft Cognitive Services" este utilizat pentru a efectua recunoașterea caracterelor.

1 INTRODUCTION

1.1 Motivation

In an ideal world, tires would not be that expensive, especially in the truck market, and their theft or illegal swapping wouldn't be that profitable. Because we do not pay much attention when it comes to our tires, aside from: in the winter to swap them with the winter ones and in summer put the summer ones, if somebody was to change our tires for other similar looking ones (just the same color is enough) the majority of the people would not even notice. You would have to keep track of what tires you are using, probably mark down their serial number and certification number (if present) and from time to time check that the ones on your car or trucks have still the same information.

The tire-markings consists of embossed letters on the side of tires and are put by manufacturers to represent the characteristics of the tire, show its certification and to distinguish between different production batches. A common serial code found on a tire is the DOT code (Figure 0a), the acronym meaning "The Department of Transportation". It is a marking that is mandated by the United States of America to be present on all the tires that are commercialized in the country and is as close as possible to a serial number. Because having different production lines is expensive, manufacturers print this DOT code also on tires that are sold in other regions of the world. This code is usually composed of: DOT marking (Figure 0a – i), tire manufacturer or manufacturing plant code (Figure 0a – ii), the size code (Figure 0a – iii), tire manufacturer (Figure 0a – iv) and finally the week and the year the tire was produced in (Figure 0a – v). Unfortunately, some of these groups of letters are sometimes missing. Anyway, this code is not enough to uniquely identify a tire as the most unique part is the date which has weekly increments only.

Another code present on the tire is the E-mark used in Europe (Figure 0b) to mark the certification which the respective tire follows and complies with. It consists of a circle with letter 'e' or 'E' followed by a number (representing the country who issued the approval that the tire meets the certification) inside a circle and next to the circle is a code or 2 lines of code representing the certification itself that the tire complies with.

The problem of tire theft and illegal swapping is more predominant in the truck market as there tires are more expensive and wear out faster. For a truck fleet owner this is a big problem because swapped lower quality tires can be a road hazard and in case of failure they can have catastrophic results. The solution at the moment is a labor intensive, slow and prone to human error or ill will: a person, that can be bribed, has to manually write down the

information after buying a new tire and at some intervals of time it has to manually check again that the tire on the truck is still the same one, with the same specifications, and that nobody illegally swapped it for an older or lower quality one. In the case of tire theft, there isn't much to do as there is no way to determine where the tires end up utilized. If we were to apply the same solution to tire theft – as we apply to tire illegal swapping – we would need people that would inspect at the tires passing on a road, mark down the information and enter it in a database. If the respective tire was declared stolen, there would be at least a starting point in the theft investigation.

But this human inspection approach doesn't scale when it comes to hundreds of millions of vehicles. So, this is a great task for automation. It could be replicated indefinitely and the limiting factor would only be the hardware required. An automatic system for collecting information from the side of tires and reporting to a database that the serial number was seen in a particular location at a certain moment would help in identifying stolen tires that are put again in use on the roads. Or can check that the tires on a truck have the serial numbers that they should have and were not swapped.

1.2 Problem Statement

To further create a system that is able to almost uniquely recognize a tire, firstly it is needed to extract the markings from the image of a wheel. These are composed of embossed letters on the side of the tire. They are black in color just as the background, so the contrast is not great. The tire-markings that are quite unique and of interest to use are the DOT serial code (Figure 0a) and E-mark certification number (Figure 0b). The manufacturer name(Figure 0c), maximum load (Figure 0d), construction materials, the ISO metric tire codes (Figure 0e) and other supplementary markings only contain information to the physical characteristics of the tire itself and would be nice to obtain, but not necessary.

At the moment, such recognition systems require extra equipment to create special environment conditions [3] or are computationally intensive and require processing on an external server [1]. My solution is to create a more robust image processing pipeline, that would not require supplementary incident lights for better letter contrast with the background. It would also be best to move as much as possible of the processing in situ to be as independents as possible from a central server.

TODO: Figure 0 a – DOT code with i, ii, . . . , v under each segment of the code — 0 b – E-mark — 0 c – manufacturer's name — 0d – maximum load — 0e – ISO metric tire codes

1.3 Overview of the Solution

My solution is to create a more robust image processing pipeline, that would not require supplementary incident lights for better letter contrast with the background, that could be run in situ, without the need of an external server for processing, in the detriment of precision (TODO: give exact numbers preferably) in text recognition. In my image processing pipeline I will be feeding photos of car tires captured with a Cannon EOS 1300D with a resolution of 5184 by 3456 pixels. The images were taken from approximately (TODO: masoara de la ce distanta fac poze in medie) cm and at least the hole tire was always in the shot (Figure 1). With this resolution and distance from the captured object, the characters composing the serial number have around 60 by 20 pixels in size, so not very much. I divided my pipeline in 3 big steps that tackle the problem sequentially: tire unwrapping, text detection, and text recognition.

a) Tire Unwrapping

Consists of the process of determining where in the image the tire is situated, detecting its outer edge, inner edge and the center of the tire, and converting the circular shape of the tire in a rectangular one (unwrapping). In Figure 1 we can see the image of a tire that in Figure 2a has its circles detected and in Figure 2b is the unwrapped result of this step. In the unwrapped version it can be seen that the tire is not perfectly straight because of the perspective of the captured image. If the camera was not perpendicular on the tire's plane and in line with the wheel's axle, the tire has a slight oval shape that is accentuated by the unwrapping as an oval has 2 centers and not only one.

TODO: Figure 1

TODO: Figure 2a and 2b

b) Text Detection

The scope of this step, to determine the regions where text might be present. Because the next step, text recognition, can be one of the most computationally intensive parts and the image's pixel count is still high after the unwrapping (TODO: get the average size of the unwrapped images), we want to reduce the space where we try to recognize characters.

I opted this step to not use machine learning and pre-trained models and instead to have a deterministic approach by using a combination of processing techniques and greedy components detection. My approach was to pass the image in the frequency domain in order to remove the high frequencies from the image – as those tend to represent noise – and the low frequencies – who usually represent the uniform background color of the tire, leaving the markings behind. This was difficult to accomplish as the tire-markings – being embossed

letters – are not very prominent compared to the background of the tire (Figure 4a) without controlling lighting conditions. I needed through multiple tests to come up with a series of heuristics to propose the areas that contain text. At this step I accepted to allow a higher number of false positives in the regions than a higher number of false negatives (who would have meant to miss on some text regions). It will be the task of the next step to deal with the falsely voted regions of text. In the end, I output a binary image (Figure 3) with the same size as the unwrapped one. The white pixels represent supposed text area that the next step should attempt to recognize.

TODO: Figure 3

c) Text Recognition

This is the step where the information is practically extracted from the image (so far I have only prepared the image or selected the regions of interest). Taking into consideration the hardships in extracting the letters' regions in the previous step, I opted to use a pre-trained OCR algorithm in order to be more robust and creating an OCR from scratch was not the scope of this project. The OCR used is part of the solutions suite of Microsoft Cognitive Services [5]. Applied on the found regions of interest, it gave (TODO: metrics).

1.4 The Document's Structure

In this section I presented the problem as a hole and then focused on the problem of automatically recognizing tire-markings. I showed what a solution to this problem would require and my approach to solving it.

The next section – 2. State of the Art – (TODO: introduce reference) will present other work in the field, what their approaches were, their setups and results and how I am bringing new contribution to the field.

In section 3. The Proposed Solution (TODO: introduce reference) I will go in depth in the image processing pipeline. I will present each step, their sub-steps and each action performed to. I will provide information on each action in a consistent pattern: the output, how it's done – the algorithm in pseudo-code and its explanation – and optionally other approaches that didn't work and motivated me to chose this action in the end.

In section 4. Implementation Details (TODO: introduce reference) I will present and explain the code, the libraries used and how the data is managed. When is the case, I will motivate the choice of an implementation over another one.

TODO: talk also about evaluation

2 STATE OF THE ART

In designing my solution I've taken into account other work performed in the field of automatically recognizing tire-markings. The difficulty of the task is twofold. On one hand, the markings on tires' sides are rarely printed because they would fade in time or under the effects of the elements. Instead, the approach employed by manufacturers is to have embossed letters on the side of the tire that would fade slower than their painted counter parts. They are not foolproof either as they can also fade in time as the letters dull and become indistinguishable from the normal side of the tire, but are usually enough to outlive the tire grooves that dull the first because of the tire's usage. One characteristic of these embossed letters is that they are practically rubber on rubber, so black on black (Figure 4a). Their visibility is not great, even with the human eye can be hard to distinguish the letters in some lighting conditions. For the human eye, it can be beneficial to have an incident source of light that would make the letters cast a shadow on the side of the tire like can be seen in Figure 4b and a person could accomplish this with a flashlight that he positions at an optimal angle.

TODO: Figure 4 (a and b)

This was also the approach of Wajahat Kazmi et alia in their work [3]. They had a 2 camera setup that each would capture half of the wheel and a supplementary source of "Strobe light incident at steep angles with respect to the plane of the sidewall" to quote them. This arrangement would help with the resolution of the images and character detection in the later stages of their pipeline. As a first step, they also performed an unwrapping of the tire using Circular Hough Transform [8]. After, they focused on detecting only the DOT code by using their crafted features in order to keep a low memory footprint (the average sizes of their images were 500x2800 pixels) and extracted a Histogram of Oriented Gradients. This output they would feed in a Convolutional Neural Network based Multi-Layered Perceptron and would have as output regions of the image where text is present. The training was done with a synthetic data-set. Then, they would localize in the proposed areas the DOT code by using a deep neural network trained on a synthetic data-set of DOT foregrounds and different tire backgrounds. The same model was used also for character recognition and was trained on a 700,000 synthetic data-set of characters on black background to mimic the low contrast appearance of the embossed markings. They claim their pipeline obtained an accuracy of 80% in images that are considered acceptable to human standards, but make a note that an objective benchmark is not available. The lower quality images that would pose difficulties even to a human to recognize the text obtained an accuracy ranging from 73% to until 14%.
TODO: sa spun aici de ce eficienta am eu si ca data-set-ul il pot face public ca lumea sa faca banchmark pe el.

While the past paper's goal was for an industrial system that would have controlled conditions when performing the tire-markings recognition, there is also work in the field, by Anton Katanaev et alia [1] for a consumer solution that would try to extract the tire specifications in order to help with ordering new ones. Their goal is to obtain the "ISO metric tire codes" that specify the type of tire, the width, aspect ratio, construction, diameter, load index and speed rating. Their first step was to collect a data-set of tire images by using internet scrappers and then filter through them using a classification model based off ResNet64. Compared to the previous work, Anton Katanaev et alia found the Circular Hough Transform unsuitable because of parameter tuning and opted for a segmentation approach for detecting the tire from the background. After this, they focused on image preprocessing to combat the different illumination in the images collected in the data-set and also perform circle correction for when the tires were appearing with an oval shape because of the camera angle. To reduce the space in which to perform the character recognition, the team also employed a step in which regions of text are searched in the image. They compared 14 pre-trained text detection models and chose the best performing one to provide supposed regions of text. After this, the proposed regions are fed into the best performing text recognition model they've tried: SEG OCR. This model obtained a character error rate of 0.16 on the original images, performing worst on the ones where the contrast was adjusted. As a final step to combat errors that might appear in the character recognition part and to identify a tire by its properties, they used a database of more than 15,000 tire characteristics for correcting incomplete or erroneous data. Because of the many learning models used, the processing part of their application is in cloud (they leverage the computational power of AWS Cloud). On the user's end is a smartphone application to submit the captured tire image to the server to start processing and provide suitable tire candidates.

In our Problem Statement we stated the desire of a system that would not require complex setups, a controlled light environment and that would be nice to be as much as possible self contained (to not rely completely on a server for processing). This is why in my approach I preferred mathematical algorithms over learning models to accomplish the task and provide a robust system that can perform in natural lighting conditions. TODO: sa introduc chestii despre acuratete dupa ce o calculez

3 THE PROPOSED SOLUTION

The solution is an image processing pipeline that would extract from the side picture of a tire the markings written on it. These are present in the form of embossed letters on the tire's exterior walls. By desiring to not have a complex setup and not require supplementary light sources, the system was tested on images taken only with ambient light. This proved to increase the problem's difficulty quite considerably when it came to detecting the regions of text, because the contrast between the embossed letters and the background image got even smaller. Multiple approaches were considered in section 3.2 Text Detection and in the end one proved fruitful in delivering acceptable results.

The system's steps are described in the following sections:

3.1 Tire Unwrapping

As a first step, I wanted to extract only the tire from the input image. This would help because the information I am interested in is located on it. I would also like to unwrap the tire from its disk shape (a circle with a smaller circle as a hole inside of it) into a rectangular one in order to have the writing from left to right rather than in all directions like if it was in a circle.

To better explain how this was accomplished, this step was split in multiple sub-steps:

3.1.1 Circle Detection

I detect where in the image are the circles representing the outer and inner borders of the tire and get their center's coordinates. The two detected circles might not have a common center because of the imperfections in taking the images, but I account and try to correct for that through a series of heuristics.

Because this is the beginning of the pipeline, I introduced a step to equalization the images and have some independence from the lighting conditions. Then I detected the first circles in the image using Hough Circle Transform and filtered its output through a series of heuristics.

a) Equalization

Accomplishes:

Lighting conditions are not controlled and this might result in low contrast in some areas of the image. Here we ensure a constant contrast across the entire image and that the entire histogram is used. This increases the robustness of the overall system.

Reasoning:

The pixel in a black and white image is a representation of how bright that particular part of the image should be (0 is for no light and 255 is for maximum brightness on a 8 bit sensor). If we take a photo in broad daylight, the values of all the pixels could all be over 200 lets say. The image would appear as being washed-out and lacking details (Figure 5a). If the photo is taken in a dark environment and all pixel values are under 100, the image appears again quite dark and lacking detail (Figure 5b). A solution is to use the space that remained unused in pixel values and stretch the existing pixel values to also cover those, thus increasing the contrast of the image and its level of detail (Figure 5c). This action is called equalization and can be used to enhance the contrast in an image. Furthermore, it spreads the pixel values across their entire value range so that a under-light and an over-light picture's histogram would look similar. The histogram is a plot that has on the X axis all possible pixel values and on the Y axis the number of times that value appears in an image.

TODO: Figure 5 (a, b, c)

A deviation from this standard equalization is Adaptive Histogram Equalization (AHE) that is not applied on the entire image and just on a small portion of it. When calculating the new value for a pixel, only a neighborhood around it is taken into consideration. Thus method while useful in bringing more contrast in to an image that has lighter and darker regions, it also increases the noise in the image. To counteract this, Contrast Limited AHE (CLAHE) was created. The addition is that it has a clipping upper limit and redistributes the pixels that appear too often in the image into another ranges.

Algorithm:

CLAHE is described in pseudo-code in Algorithm 1 in accordance to the implementation found in [12] and the explanation in the paper [4].

On line 1 we split the image in multiple cells of size k by k because CLAHE is an adaptive algorithm. Then, for each of the cells we calculate their histogram to find out how many different pixel values they contain (N_{gray}). To perform the clipping, we need to calculate a few things:

$N_{avg} \leftarrow (k \times k) \div N_{gray}$, that is the number of pixels in the cell divided by how many pixel values are present in the cell;

Algorithm 1 CLAHE

Require: img be a $X \times X$ matrix; $cl \in [0, 1]$; k

Ensure: $k \mid X$

- 1: $S \leftarrow (img \text{ split into } k \times k \text{ cells})$
 - 2: **for** s in S **do**
 - 3: $H_s \leftarrow histogram(s)$
 - 4: $H_{res}, N_{remain} \leftarrow clip(H_s, cl)$
 - 5: $H_{res} \leftarrow redistribute(H_{res}, N_{remain})$
 - 6: $s \leftarrow adjust_cell(s, H_{res})$
 - 7: **end for**
 - 8: $img \leftarrow \text{interpolate cells } S$
-

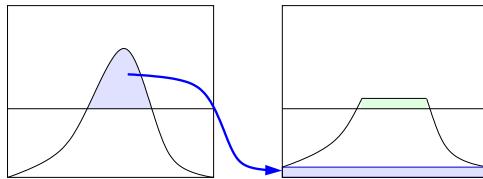


Figure 1: the histogram after first clipping, source: TODO wikipedia

$N_{CL} \leftarrow cl \times N_{avg}$, the clipping limit calculated based on input parameter cl ;

$N_{\Sigma_clip} \leftarrow \sum_{i=0}^{N_{gray}} \max(0, H_s[i] - N_{CL})$, is the number of pixels that need to be clipped, $H_s[i]$ is the pixel count for a certain pixel value;

$N_{avg_gray} \leftarrow N_{\Sigma_clip} \div N_{gray}$, is how many pixels in average we will modify in addition to have a certain value.

Then, we iterate through all the pixel values in the histogram and create a new one. If the pixel count is \geq than N_{CL} , we will clip that pixel value. Else, we will add to the pixel count N_{avg_gray} and if we exceed the clipping limit, we floor to it and take note of this in the form of N_{remain} . After this step, the histogram would look like in figure 1.

The limit we excede with will be redistributed after this on line 5. Here, we go through the pixel values in the histogram again and at a simmilarly random step we add one more pixel to the count that overflowed earlier. The step is computed by the formula: $step \leftarrow N_{gray} \div N_{remain}$.

Before starting to compute the next cell, we adjust the current one according to the new histogram we just computed: H_{res} .

After all the cells were modified and equalized, we can stitch them back together. But because we had a cell by cell approach to calculating the histogram, border pixels might have very different values. This is why an extra step of bilinear interpolation is performed on the border pixels of each of the cells.

Results:

After equalization is applied, if the input image's histogram looked like TODO: Figure, now it looks like TODO: Figure.

b) Hough Circles Transform (HCT)

Accomplishes:

It detects circular shapes present in an image. The ones we are interested in are the outer and inner rims of the tire.

Reasoning:

The reasoning for why we want to detect the tire in the image is twofold. Firstly, the text searching part is a computational expensive one and we want to reduce the search space from the whole 3456 by 3456 pixels to usually a forth of that. The second reason is that the text is present in the image in a circular form, while the vast majority of optical character recognition (OCR) software is performing best on text that is aligned along one dimension only. So, by detecting the tire's circular shape in the image, we can transform the tire into its unwrapped version (that also has a smaller pixel count).

Algorithm:

Algorithm 2 Hough Circles Transform

Require: $img; r_{min}; r_{max}; dp; canny_params$

- 1: $blur \leftarrow gaussian_blur(img)$
- 2: $gray \leftarrow color2gray(blur)$
- 3: $edges \leftarrow canny(gray, canny_params)$
- 4: **for** $r = r_{min} : +1 : r_{max}$ **do**
- 5: $acc[i] \leftarrow matrix(shape(img) \div dp)$
- 6: **for each pixel** $p > 0$ **in** $edges$ **do**
- 7: $acc[i] \leftarrow accumulate(acc[i], p, r)$
- 8: **end for**
- 9: **end for**
- 10: $center, radius \leftarrow extract(acc)$
- 11: **return** $(center, radius)$

Hough Circles Transform is particularization of the classical Hough Transform [8] and its scope is to extract features from images. The original algorithm is searching for a circle with a known radius, but as can be seen in Algorithm 2, it can be generalized to an interval of radii.



Figure 2: Canny output

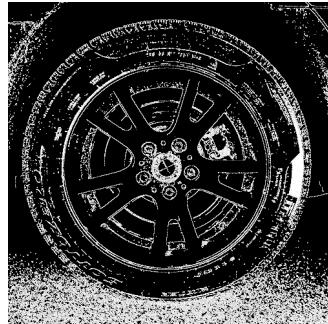


Figure 3: HCT accumulator



Figure 4: HCT result

The first 2 steps in the algorithm are to apply Gaussian Blur [11] (to remove the noise that follows a Gaussian distribution) and convert the color image into gray-scale.

Then it uses Canny Edge Detector [9] to obtain a new image that contains only the outlines where the algorithm considered a boundary between 2 objects exists. An example can be seen in Figure 2, where each white pixel represents an edge.

On line 4, it can be seen the *loop* that will probe for the radius interval that was requested.

For each radius that the algorithm will probe, will create an accumulator matrix that is dp times smaller than the original image. The purpose of this matrix will be explained shortly.

The central part of the algorithm is the for loop on line 6. It goes through all the pixels that resulted in being edges and finds their corresponding position in the smaller $acc[i]$ matrix. Then, all around that point in the accumulator matrix at distance r , it increments the corresponding cell. 3 iterations of this can be seen in Figure 3.

Because the core algorithm probes for a specific radius, multiple accumulator matrices are used. So, the last step of the algorithm (line 10) is to parse all the accumulators and find the most incremented cell (the red dot in Figure 3). That cell is considered the center of the circle and the radius is corresponding to the accumulator matrix it was found in.

Results:

Because tires have different dimensions and an outer and inner rim, I have to search for multiple circles in the image. So, I probed for multiple radii and the algorithm found multiple circles in the image, as can be seen in Figure 4. This will be the input for the next step, where through a series of heuristics the circle count will be reduced to only the most concentric ones and hopefully only 2.

c) Heuristics

Accomplishes:

Reduces the number of detected circles in the previous step and remains only with 2 concentric circles representing the tire's inner and outer rims. If this step fails to remain with only 2 circles, then the image is discarded.

Reasoning:

There are many types of tires in the wild with different shapes and sizes, so it was better in the previous step to detect more circles in the image than to look for only 2 circles. When looking only for 2 circles using the Hough Circular Transform, the car's wheel arc can be falsely detected as a circle or the wheel cap can have a pattern on it that would falsely trick the circle detection algorithm. Because of this reason, I preferred to detect many false positives and at this stage filter through them and remain with only the most concentric circles that have at least a certain radius difference. Another possible problem would be that the tire is not a perfect circle because of the contact point with the ground or because when taking the image, the camera sensor was not in line with the wheel's axle.

Algorithm:

The first heuristic I apply is to reduce the circles that are very similar one to another. I start from the most prominent circle found by the previous step. If the conditions $dist(c_1, c_2) < 0.5 \times size(img)$ and $|r_1 - r_2| < 0.37 * size(img)$ are found true, that means the circles are probably the same one and I calculate a mean one. The mean is calculated by taking the average of the centers and the average of the radii. In checking other circles, this one will be used. The values for the thresholds were found experimentally.

The second heuristic's goal is to create 2 concentric circles. At first I was taking the biggest and smallest circles and using the average of their center's coordinates to calculate a new center. This gave unsatisfactory results as the computed center would be quite far from the true one and in the unwrapping phase, it would not make the tire a continuous horizontal strip. The improvement I ended up finding was that the smaller radius circle would have the center closer to the wheel's true center, so it would be better to use the inner circle as the base rather than to calculate a new center. The reason to why the smaller circle has a better center is that the inner circle usually represents the inner rim of the tire that is a perfect circle, while the outer rim is deformed at the contact point between the ground and the wheel. That deformity can throw off the center calculated at the previous step. If more than 2 circles are present, this step is skipped and the image will be discarded as we couldn't detect the inner and outer parts of the tire. An improvement that could be made is that if more circles are found, to select the ones that have the closest centers and a radius bigger than a certain value.

Results:

The resulted concentric circles are an approximation of what the true circle might be,

but it is close enough so that the tire can be unwrapped. The circle might not follow perfectly the rim of the tire because of the distortion of the wheel on the ground, or the distortion was introduced when taking the picture if the camera's sensor wasn't horizontal and in line with the wheel's axle. If we couldn't remain with 2 concentric circles at this step, the image is discarded. Otherwise, it is passed to the next step.

3.1.2 Convert to Polar Coordinates

Accomplishes:

This step unwraps the tire and outputs it as a continuous horizontal strip as can be seen in Figure [TODO]. This can be done because we have detected the supposed center of the tire and its radius.

Reasoning:

The horizontal strip will help in text detection and recognition, as it is easier to work with text going in only one direction.

Algorithm:

A normal image representation is in the Cartesian domain. In this domain, a pixel is characterized as having 2 coordinates: x and y that we are familiar with. So a pixel is defined by 2 distances and the 2 distances have an unique pixel tied to them. Another domain to represent images in, is the Polar domain. In this domain, a pixel is characterized also with 2 coordinates: *angle* and *radius*. The entire image is constructed around a center point and if information in the Cartesian domain was arranged in as the height of the image is the y and the width is the x , here they are the *angle* and *radius* respectively.

Because of this correlation, we can convert from one domain to another. We are interested now in converting from a Cartesian system into a Polar one as the tire has a circular shape. We can use the found circle center to make it the origin of the new Polar domain and start taking points from the image starting at the small radius until the big one. For each pixel in the image, now with the found circle's center as the image's one, we can find its new x and y coordinate and from them, we can calculate its representation as *angle* and *radius* using the following formulas.

$$radius = \sqrt{x^2 + y^2}$$

$$angle = \arctan(y/x)$$

There is a problem though. Closer we are to the new selected center, less points are present on the surrounding circle than on a bigger one. The resulting image after the Polar transformation will have as it's width the pixel count of the pixels that were on the most outer

circle's perimeter. The points closer to the center must be stretched out in order to fill in the gaps and this is solved using interpolation. New computed pixels are used to fill in the gaps.

Results:

Multiple interpolation levels were used to observe if it affects in any way the output, but the letters on the tire are far away from the center of the polar transform that even a bilinear interpolation would not affect them. Now, I have the unwrapped tire as a new image and it is ready to be processed further in order to obtain the regions of text.

3.2 Text Detection

Now that the initial image was processed and just the tire region was kept and unwrapped to have a rectangular form with the text spanning from left to right, the text detection phase can begin. The goal now is to obtain a bitmap of the rectangular tire with regions where text is present. This phase is useful because it will further reduce the space where we have to perform text recognition, and because the goal is to reduce, false-positives are permitted in an acceptable limit. If we can reduce the image search space by at least 80%, this step is considered successful and the text recognition can begin. Otherwise the image is discarded.

Before starting with the pipeline's steps, it must be mentioned the unsuccessful initial attempts at detecting the text in the image. The main culprit was the low contrast between the embossed letters (that make up the tire-markings) and their background. Because the letters are not a distinctive feature as can be seen in Figure [TODO: ref la figura din chap 1, e una pe acolo poate chiar 2], if they appear in the output of a processing algorithm, they are always accompanied by noise or other features on the tire that do not interest us.

For the beginning, I tried with Sobel image gradient [TODO: cite] to obtain a gradient of change in the image, but it did not look useful as can be seen in Figure [TODO:].

Then I tried further by applying Canny Edge Detector [9] but in order to obtain usable results, I had to fine-tune its parameters and these were extremely influenced by the color of the tire, the lighting conditions and noise. A somewhat usable output after fine-tuning can be seen in Figure [TODO:]. Noise was a real issue because if too much blurring was applied to remove the noise the letters would lose shape.

I also tried some thresholding techniques. OTSU Thresholding [TODO: cite] proved useful because it is a global algorithm and different materials on the wheel reflect light in different amounts so a single threshold is not practical. I tried also the option of adaptive Thresholding and it proved useful as can be seen in Figure [TODO:], but it suffers also from parameter tuning, just like the Canny Edge Detector and because the system must be robust this is not acceptable.

Before resorting to Machine Learning models to solve this step (a thing that I avoided in order to keep the processing pipeline simple and not computationally intensive), a solution appeared. Filtering certain frequencies in the frequency domain of the image showed that only the letters can be left prominent in a segment of the image. From this stepping stone, the following steps were devised in order to obtain a bitmap with supposedly text regions

3.2.1 Segmentation

Accomplishes:

Splits the image in smaller cells in order to perform operations on each one of them. These cells are overlapped a certain percentage to be able to vote on them regions of interest.

Reasoning:

While searching a way to obtain the text region in an image, frequency filtering proved useful and not needing parameter tuning. The less ideal part is that the last step is OTSU thresholding that is a global algorithm. While not performing useful if we were to process the entire image at the same time, it proved very helpful in binarizing smaller areas of the big image. So, I decided to split the original image in overlapping cells.

The cells overlap because in the end a voting process is employed to remove regions that were considered text in just a few of the segments.

Algorithm:

I need to calculate the corners of cells that have a certain overlap between them. This overlap is present on the x and y axes. Firstly, I calculate the number of cells that will be needed on each of the axes with the formula:

$$N = \frac{L - o}{no}$$

where L is the number of pixels, o is the size of the overlap region and no is the region of a cell that is not overlapped.

Results:

Now the big picture was split in multiple smaller overlapping cells that will be further processed individually.

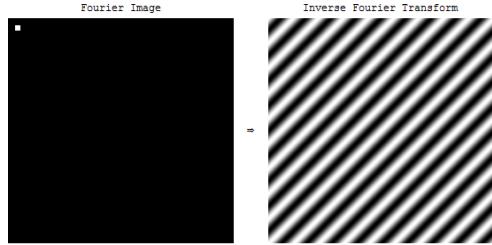


Figure 5: Frequency domain representation

3.2.2 Segment Binarization

a) Image Frequency Filtering

Accomplishes:

This is a band-pass filter applied to the frequency domain of the segment. The remaining frequencies will consist of letter shapes (the tire-markings) and some artifacts.

Reasoning:

Because the background of the tire is fairly uniform in blackness and noise, these can be considered low frequency color variations and high frequency noise respectively. By eliminating certain bands from the frequency domain of the image, we are left with the letter shapes that could be processed further to enhance them and filter out any artifacts that remained.

Algorithm:

Algorithm 3 Band-pass Filtering

Require: img

- 1: $f_rep \leftarrow FFT(img)$
 - 2: $f_rep \leftarrow bandpass_filter(f_rep, low, high)$
 - 3: $img \leftarrow IFFT(f_rep)$
-

Any analog signal is composed from multiple signals of different frequencies. These frequencies could be extracted using Fourier Transform [TODO: cite] or reconstruct the signal using its inverse. Because it is a computationally expensive algorithm, in practice Fast Fourier Transform [TODO: cite] is used because it usually gives acceptable results.

This step can be described by Algorithm 3. But firstly, it must be made clear what passing an image into the frequency domain means. The easiest is to take a one dimensional example (a line) and then generalize for a two dimensional case (an image).

On the line, each pixel value could be considered to represent the amplitude of a signal. The signal progresses in time to one direction of the line. This signal can be decomposed

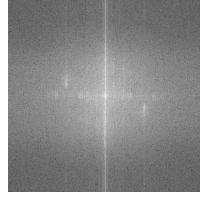


Figure 6: Frequency domain

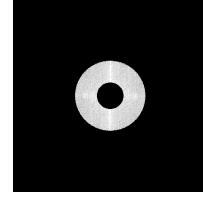


Figure 7: Frequency domain

in its core frequencies, to the left the low ones and the high ones to the right. The same thinking could be also applied to a 2D image, just that we get a 2D frequency image with real and imaginary components. For example, Figure 5 represents what happens if we have only a frequency present, what the resulting image would look like. The distance from the corner represents the frequency, while the position of the dot is the angle the bands will appear after applying the Inverse Fourier Transform.

To be easier to visualize the frequencies, the image is usually shifted so that the 0 frequency is in the middle of the image rather than in the corners like in Figure 6. This is useful also because it makes it easier to apply a band-pass filter. To apply this, we just have to blackout the middle of the image and the outer parts. In the end we are left with a disk shaped frequency representation of the image (Figure 7). Experimentally I got usable results by keeping a disk with the small radius 14% and the big radius 36% of the found frequencies in the image.

After applying the Inverse Fast Fourier Transform, we can observe that the low (the background) and the high (the noise) frequencies were removed. I successfully removed the background and the noise and still remained with the general letter outline in the image as can be seen in Figure [TODO: la fel ca cel de la results].

The huge benefit of this method is that it can be applied to more images using the same band-pass filter, without changing the cutoffs.

Results:

Outputs an image like in Figure [TODO:], where letter shapes could be observed. There are also artifacts remaining that will be dealt with at a later stage.

b) Black Hat Morphological Operation

Accomplishes:

Enhances the darker regions representing the letter contours that are surrounded by brighter regions.

Reasoning:

It can be seen that in the resulted image at the previous step (Figure [TODO: ca mai sus]), the letters have an outline in dark pixels surrounded by bright regions. It can be more clearly seen in Figure [TODO: zoom pe litera]. By extracting this outline of the letters, we will be able to have the letters more prominent in the image than other features.

Algorithm:

A morphological operation is based on mathematical morphology [2]. The main components to it are an object A (our image) and a structuring element B . Normally, these operations are applied on binary images but variations exist that let them to be used also on gray-scale ones. Firstly, a path must be created for our structuring element (B) to traverse the edge of the object (A). This path if formally defined as a function that maps points from our image to a real line.

$$f : E \rightarrow R$$

The Black Hat Morphological operation is defined mathematically as:

$$T_b(f) = f \bullet b - f$$

where b is a structuring element and \bullet means the closing operation that is defined as a dilation followed by an erosion of a set A by a structuring element B :

$$A \bullet B = (A \oplus B) \ominus B$$

Results:

The result of this step can be seen in Figure [TODO:]. The figure has the brightness increased for viewing, but even if it's a darker image, it would not be a problem for a thresholding algorithm that can select his own threshold. This is why, in the next step we apply OTSU thresholding.

c) OTSU Binarization

Accomplishes:

Computes automatically a threshold for the image and applies it, outputting a binary image. Because of the processing previously done, the output consists of tire-markings and some artifacts.

Reasoning:

When trying different methods to make the regions of interest (the tire-markings) remain in the image and remove the parts I am not interested in, OTSU binarization happened to do just that. Before the voting could begin, I anyway would have had to binarize the image

so this is even more helpful. The fact that OTSU computes automatically its threshold is also very helpful as it prevents the need to experimentally fine tune the algorithm and perform overfitting. The output would have some artifacts that I removed as much as possible through a set of heuristics later.

Algorithm:

The novelty in the OTSU Thresholding (Algorithm 4) is that it computes the threshold by minimizing the intra-class variance (the variance between the white and black pixels that would result).

For each possible threshold value computes the binary image. After which the algorithm computes on the respective image the variance in white and black pixels and returns the weighted variance of the two. This can be seen on line 4 of the algorithm. We store the resulting image with the threshold and the weighted variance in 2 vectors (I and WV) that are self incrementing.

As a last step (line 6) we select the threshold that minimizes the weighted variance and retuns the corresponding computed binary image with its corresponding threshold.

Algorithm 4 OTSU Thresholding

Require: img

```

1: for  $t$  in possible thresholds do
2:    $img_{bin} \leftarrow threshold(img, t)$ 
3:    $I[i] \leftarrow (img_{bin}, t)$ 
4:    $WV[i] \leftarrow weighted\_variance(img_{bin})$ 
5: end for
6:  $it \leftarrow pos\_of\_min(WV)$ 
7: return  $I[it]$ 
```

Results:

It can be seen in Figure [TODO:] that the letters or text remains almost as a connected component while the artifacts span across the entire segment or are small dots. This is very helpful as we can filter for the components that have a certain size and shape.

3.2.3 Filtering Artifacts

Accomplishes:

The artifacts that can be seen in Figure [TODO: cea de mai sus] are removed as much as possible, while retaining the letters' shapes in the binary map. This process is a combination of heuristics and morphological operations.

Reasoning:

Because the output of this step will be used in voting the regions of interest the text recognition will take place on, we would like to discard as many artifacts as possible. A big majority of them can be identified and removed because of their shape and size. The regions of text we are interested in are elongated and not very tall, while the letters have a common supple shape.

Algorithm:

Algorithm 5 Filtering Artifacts

Require: *img*

- 1: *img* \leftarrow *dilate*(*img*, *se*, 1)
 - 2: *CCs* \leftarrow *contours*(*img*)
 - 3: *img* \leftarrow *filter*(*img*, *CCs*, *heu*₁)
 - 4: *img* \leftarrow *dilate*(*img*, *se*, 4)
 - 5: *CCs* \leftarrow *contours*(*img*)
 - 6: *img* \leftarrow *filter*(*img*, *CCs*, *heu*₂)
 - 7: **return** *img*
-

The pseudo-code for the filtering process can be seen in Algorithm 5.

It starts off with a dilation, which is a morphological operation [2] which results in the thickening of the white connected components. It can be described mathematically as:

$$A \oplus B = \{z \in E | (B^s)_z \cap A \neq \emptyset\}$$

where *A* is our image with sets in it, *B* is a structuring element that is applied across the entirety of *A* and *B*^s is the symmetric of *B*. This is used in order to unite any artifacts and letter parts that might be disconnected because of the initial lighting conditions. For now I apply a single iteration.

The operation on line 2 looks in the image and finds all the connected components (*CCs*).

Then on line 3 we apply the first series of heuristics. These were determined experimentally and through the observation of the sizes of letters. On each component, a bounding box is placed. Any of the following rules (*heu*₁) would force the component to be discarded: $h < 10p$ | $w < 10p$ (even if this was a letter, it would be too small to be recognizable), $h/w > 1.5$ | $h/w < 0.33$ or $h > 0.33 * h_{seg}$ & $w > 0.66 * w_{seg}$. (*h* - height of the component, *w* - width of the component, *h*_{seg} - height of the segment, *w*_{seg} - width of the segment).

After this first series of filtering noise, I've decided to apply again dilation operation but for 4 iterations in order to thicken up the remaining sets. I do not want necessarily to close

the spaces between them, as I would have chosen to use the closing morphological operation if I wanted that.

I search again for the connected components (*CCs*) in the image as they might have changed quite drastically from the last time.

And now, as a last step (line 6), I apply another filtration with other rules. In my experiments, I observed that some more artifacts could be eliminated at this step so I chose to do just that. The rules for removing these artifacts (heu_2) are: $h < 26p \mid w < 10p$ (even if this was a letter, it would be too small to be recognizable), $h/w > 1.5 \mid h/w < 0.33$ or $h > 0.33 * h_{seg} \& w > 0.66 * w_{seg}$. Practically, almost the same rules as in the first round of filtration, just that now the height of the component must be bigger as we have applied dilation a few times already.

Results:

Even if the filtering process gives some false negatives and lets some artifacts to pass, it would be way worse to have false positives and start removing from the tire-markings themselves. I am satisfied with the results that can be seen in Figure [TODO: poate aici sa lipesc mai multe poze ca sa arate pe mai multe cazuri]. This bitmap will be used for voting in the accumulator that has the same size with the unwrapped image.

Finding connected components is a quite expensive operation, so further improvement could be done on this step in order to reduce the number of searches for connected components to speed up the pipeline.

3.2.4 Voting

Now that the segment was processed (revealed the zone where text is present and removed as much as possible of the artifacts present in the image), the white zones that remained will have their corresponding place in the accumulator matrix (that is the same size with the unwrapped image) increased by one. This is the action that I called voting. Because of the overlap between segments this voting step can be used to further reduce the number of false positive regions of interest.

After all the segments were processed, it results a voting map like the one in Figure [TODO: sa i fac enhance ca sa fie mai vizibila pe foaie] (it was enhanced to be more visible), where the brighter the region, the more votes it got. But this voting accumulator can be further processed in order to reduce the zones that will remain as regions of interest for the text recognition part of the pipeline.

The filtration could be described by the pseudo-code in Algorithm 6. I start off with a threshold operation, where I discard any region that did not get at least 3 votes. This step is done to reduce the artifacts that appeared only in some segments. The thresholding



Figure 8: Text Regions on Unwrapped Tire

operation also makes the pixels that passed the check to have the maximum value, resulting in the accumulator becoming a binary image.

Algorithm 6 Cleaning Accumulator Matrix

Require: acc

- 1: $acc \leftarrow threshold(acc, 3)$
 - 2: $CCs \leftarrow contours(acc)$
 - 3: $acc \leftarrow filter(acc, CCs, heu_3)$
 - 4: $acc \leftarrow dilate(acc, se, 4)$
 - 5: $CCs \leftarrow contours(acc)$
 - 6: $acc \leftarrow filter(acc, CCs, heu_4)$
 - 7: **return** acc
-

Then, on line 3 I am filtering through the connected components found in the accumulator matrix (the operation on line 2). The component is discarded if its bounding box has $h < 40p \mid w < 10p$ (heu_3) as it would be too small to be even a letter. At this step the height limit was increased compared to heuristic hre_2 because the components that are voted by multiple segments tend to thicken up. This might be because of the frequency filtering leaves different aspects of a feature (tire-marking or artifact) visible as it is very dependable on the content of the segment.

On the filtered accumulator matrix, 4 iterations of the morphological operation dilation were applied to connect the letters. Other artifacts that are still present would not connect with anything else if they are far away enough from other connected components.

Now, on the newly diluted accumulator matrix and the new connected components, we apply another filtration using the heuristics heu_4 that consist of: heuristic hre_3 , $h > 200p$ (the text region is too tall so it is a region that we are not interested in) or $h > 100p \ \& \ h/w > 0.85$ (a region too tall that doesn't have a prominent enough shape).

The result of this step is a bitmap that when applied over the wrapped tire shows regions of text (Figure 8). This bitmap will be used in the text recognition step to indicate the regions of interest in the image and reduce the space on which the algorithm will run on.

3.3 Text Recognition

Accomplishes:

Now that we have reduced our unwrapped image to a few regions of interest, we can use an OCR to start extracting the text. This will be the output of our system and it will define our performance.

Reasoning:

It could be considered that the reason is stated in the Motivation section of this work, but to reiterate, we want to extract the serial DOT code and the E-mark certification in order to almost uniquely identify a tire. I chose to use an OCR for this step because they are the most common approach to text recognition and the availability of ready made solutions only increased in the past years [10]. There is also a wide variety to the OCR solutions that are offered as services, accessible through APIs.

Algorithm:

The OCR I used is a service from Microsoft Cognitive Solutions [5] and it uses "deep-learning-based models and works with text on a variety of surfaces and backgrounds" (source: [7]). So, it is not specially tailored to our case (tire images). This is a good thing as it provides versatility at the expense of performance and openness of the system.

The Read API (that is the OCR's name) [6] is accessed through API calls to a configured endpoint. For each region of interest extracted in the Text Detection 3.2 section, a request is sent to the endpoint which returns its guess on the text in the image. While not being able to perform all the processing on only one local machine (as we had to rely on the computation power and OCR solution hosted on a remote server), the impact this has on performance is reduced because we submit only the regions of interest (which have a smaller size than the original unwrapped tire image).

Results:

The results from this step are in the form of annotations that are coupled with the original detected region of interest. The recognized text returned by the OCR will be categorized in the following categories: DOT serial code, E-Mark certification and miscellaneous.

a) TEMPLATE

Accomplishes:

Reasoning:

Algorithm:

Results:

4 IMPLEMENTATION DETAILS

BIBLIOGRAPHY

- [1] Katanaev Anton, Ukhmylov Yaroslav, Chekmareva Alfiya, and Khalili Roman. How to identify vehicle tires using deep learning visual models. <https://blog.griddynamics.com/how-to-identify-vehicle-tires-using-deep-learning-visual-models/>. Last accessed: June 2022.
- [2] Robert M. Haralick, Stanley R. Sternberg, and Xinhua Zhuang. Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(4):532–550, 1987.
- [3] Wajahat Kazmi, Ian Nabney, George Vogiatzis, Peter Rose, and Alex Codd. An efficient industrial system for vehicle tyre (tire) detection and text recognition using deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 22(2):1264–1275, 2021.
- [4] J. MA, X. Fan, S.X. Yang, X. Zhang, and X. Zhu. Contrast limited adaptive histogram equalization based fusion for underwater image enhancement. *Preprints 2017*, 2017030086, 2017.
- [5] Microsoft. Microsoft cognitive services. <https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/>. Last accessed: June 2022.
- [6] Microsoft. Microsoft cognitive services read api. <https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/overview-ocr>. Last accessed: June 2022.
- [7] Microsoft. What is computer vision? <https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/overview>. Last accessed: June 2022.
- [8] OpenCV-Docs. Hough circle transform. https://docs.opencv.org/3.4/d4/d70/tutorial_hough_circle.html. Last accessed: June 2022.
- [9] Adrian Rosebrock. Opencv edge detection (cv2.canny). <https://pyimagesearch.com/2021/05/12/opencv-edge-detection-cv2-canny/>. Last accessed: March 2022.
- [10] Han Ted and Hickman Amanda. Our search for the best ocr tool, and what we found. <https://source.opennews.org/articles/so-many-ocr-options/>. Last accessed: June 2022.
- [11] Wikipedia. Gaussian blur. https://en.wikipedia.org/wiki/Gaussian_blur#cite_ref-ShapiroStockman_1-0. Last accessed: June 2022.

- [12] Karel Zuiderveld. C ansi implementation of contrast limited adaptive histogram equalization. <http://www.realtimerendering.com/resources/GraphicsGems/gemsiv/clahc.c>. Last accessed: June 2022.