TANSZÉKVEZETŐ

DIPLOMATERVEZÉSI FELADAT

Danyi Dávid
Villamosmérnök hallgató részére

Marker alapú helymeghatározás képfeldolgozással

Az információfeldolgozási kapacitás nagymértékű növekedésével párhuzamosan egyre szélesebb körben váltak alkalmazhatóvá (valós időben) képfeldolgozási, gépi látás alapú eljárások. Az ideálistól eltérően a valós képek feldolgozását számos hatás nehezíti, úgy mint zaj, optikai torzítások, megvilágítás- és színbeli különbségek, stb.

A feladat témája aktuális, az egyre gyakrabban alkalmazott mobil robotikai, kiterjesztett és virtuális valóság alkalmazásoknál minden esetben felmerül a „hol vagyok?" kérdés, azaz a nézponti paraméterek becslése, lokalizáció. A SLAM (Simultaneous Localization and Mapping) algoritmus számos szenzorforrás által biztosított információval alkalmazható, így kamerával is. Az általános megoldás tájékozódási pontok egymáshoz képesti elhelyezkedését becsli, a mérési bizonytalanságot folyamatosan csökkentve, valamint méri a tájékozódási pontokhoz képest a megfigyelő pozícióját.

Jelen munka is a fenti témához, kapcsolódik, cél megvizsgálni egy olyan mesterséges, passzív marker megvalósíthatóságát, ami bizonyos paraméterekben (azonosíthatósági tartomány, részleges láthatóság) jobbat kíván biztosítani, mint az elterjedt (ARtag, glyph, stb.) megoldások. A marker egy oldalukon nyitott különböző méretű és alakú négyszögeket tartalmaz. A diplomaterv célja megvizsgálni a javasolt marker jellemzőit és használhatóságát.

A hallgató feladatának a következőkre kell kiterjednie:

- Mutasson be pozícióbecslő algoritmusokat, röviden ismertesse ezek működését!
- Hasonlítson össze különböző nézőpontbecslési algoritmust síkban elhelyezkedő pontpárok alapján!
- Készítsen egy marker felismerő megoldást!
- Végezzen méréseket (ideális és valós képeken) a marker által meghatározott pozíció pontosságára vonatkozóan!
- Hasonlítsa össze és értékelje az eredményeket!

Tanszéki konzulens: Kovács Viktor, tanársegéd

Budapest, 2017. február 18.

Dr. Charaf Hassan
egyetemi tanár
tanszékvezető

Budapest Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Automatizálási és Alkalmazott Informatikai Tanszék

H-1117 Budapest, Magyar tudósok krt 2. Q.B.207
Telefon: 463-2870. Fax: 463-2871
www.aut.bme.hu

Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Automation and Applied Informatics

Marker Based Localisation and Pose Estimation Using Image Processing

Thesis

Author
Dávid Danyi

Consultant
Viktor Kovács

April 29, 2018

Contents

HALLGATÓI NYILATKOZAT

Alulírott *Dávid Danyi*, szigorló hallgató kijelentem, hogy ezt a szakdolgozatot meg nem engedett segítség nélkül, saját magam készítettem, csak a megadott forrásokat (szakirodalom, eszközök stb.) használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelmûen, a forrás megadásával megjelöltem.

Hozzájárulok, hogy a jelen munkám alapadatait (szerzõ(k), cím, angol és magyar nyelvû tartalmi kivonat, készítés éve, konzulens(ek) neve) a BME VIK nyilvánosan hozzáférhetõ elektronikus formában, a munka teljes szövegét pedig az egyetem belsõ hálózatán keresztül (vagy autentikált felhasználók számára) közzétegye. Kijelentem, hogy a benyújtott munka és annak elektronikus verziója megegyezik. Dékáni engedéllyel titkosított diplomatervek esetén a dolgozat szövege csak 3 év eltelte után válik hozzáférhetõvé.

Budapest, April 29, 2018

Dávid Danyi
hallgató

Kivonat

A képfeldolgozás nem újkeletű tudományterület, már évtizedek óta folynak kutatások ezen a téren. Sok, ma is használt algoritmust az 1960-as években fejlesztettek ki. Akkoriban főleg a tudósok körében használt, drága eszköznek számított a számítógépes képfeldolgozás. Műholdképek, orvosdiagnosztikai adatok elemzésére, optikai karakterfelismerésre használták. Az olcsó és nagy teljesítményű, általános felhasználású számítógépek terjedésével azonban új lehetőségek nyíltak meg ezen a területen. Lehetővé vált például a valósidejű képfeldolgozó algoritmusok futtatása. Ezen fejlődés nélkül lehetetlen lett volna a 3 dimenziós látórendszerek kifejlesztése. Ezek a rendszerek jóval számításigényesebbek a klasszikus képfeldolgozási problémáknál, de a mai technológiával már ezek a megoldások és elérhetőek az átlagos felhasználók számára. Ennek legfőbb jele a robot navigációs, virtuális- és kiterjesztett valóság alkalmazások széleskörű megjelenése.

Jelen munka a fiduciális markerek alapján történő nézőpont meghatározás alkalmazhatóságát vizsgálja. A nézőpont-meghatározás célja a kamera pozíciójának és orientációjának meghatározása egy ismert markerhez viszonyítva. Ez egy összetett feladat aminek a megoldása több képfeldolgozási- és optimalizációs feladat megoldását igényli. Ez a dolgozat be fogja mutatni a kép alapján történő nézőpont-meghatározás lépéseit.

A munka első részében ismertetésre kerül a nézőpont-meghatározási probléma. Először röviden össze lesz foglalva a P-n-P néven ismert probléma: a nézőpont meghatározás n pontpár alapján, aminek ismert a világkoordinátákban adott helyzete, valamint a képi helyzetük is. Ezt a problémát többen többféleképp megoldották, néhány ilyen megoldás alapvető gondolatmenete összefoglalásra kerül. Ennek a szakasznak a zárásaként összehasonlítom az eljárásokat és kiválasztom a tulajdonságaik alapján a projekt céljára a legideálisabbat.

A pozicionálás pontosságát és robusztusságát nagyban befolyásolhatja az alkalmazott fiduciális marker is. Vannak ugyan már elterjedt markertípusok (ARTag, glyph, stb...), de jelen munkában kísérletet teszek egy új marker tervezésére. Ez az új marker megpróbál jobb eredményt nyújtani bizonyos területek, mint a már elterjedt megoldások. Ezen marker alkalmazhatósága is vizsgálva lesz ebben a dolgozatban.

A dolgozat utolsó nagyobb része a markerek felismerésére használt képfeldolgozási eljárásokról fog szólni. A különböző algoritmusok rövid elméleti áttekintése után egy-egy implementációs javaslat is közlésre kerül. A felismerő eljárások hatékonysága is vizsgálat alá

kerül, ideális és zajos képeken egyaránt. A mérési eredmények alapján javasolni fogok egy, a projekt számára optimális markerfelismerő eljárást.

Abstract

Image processing has been an intensively researched subject for decades. Many algorithms that are used today have been developed in the 1960s. At that time, it was a costly tool mainly used by scientists for satellite imagery, medical imaging, optical character recognition, etc... The advancement of cheap and powerful general purpose computers opened up new possibilities for research and application. Real time image processing became possible. An interesting and even more computationally expensive sub-field of computer vision is 3D reconstruction. With today's (consumer) technology it is possible to map the 3D world based on image processing solutions. Navigational, Augmented and Virtual Reality applications are spreading.

This paper will examine the use of fiducial markers for camera pose estimation using a single camera. The goal of pose estimation is to determine the position and orientation of the camera with respect to a known marker. This is a complex task, which involves multiple image processing steps, as well as solving optimization problems. This work will provide an overview of the steps necessary for estimating the camera pose based on picture of a fiducial marker.

A section of this work will be dedicated to the pose estimation problem. There will be a short summary of the problem of reconstructing the view point based on point pairs in the world coordinate system and image points. Then some algorithms will be summarised that solved that problem. This section will be closed by comparing the benefits and drawbacks of these algorithms and choosing the one that best suits the need of this project.

The choice of the marker also influences accuracy and robustness of the pose estimation solution. There are already some marker types available for use (ARTag, glyph, etc...). This paper also proposes a new marker type which tries to offer better performance than the aforementioned solutions. The applicability of the new marker will be examined in various conditions.

The last major part of this work is about the different possible methods for extracting the markers from the images. In that section there will be short theoretical summaries of the detection methods. After the theory is covered, implementations of the aforementioned methods will be recommended. The performance of the detection algorithms will also be benchmarked on optimal and noisy images. Based on the tests results an optimal method will be selected.

Abbreviations

This is a complete list of the abbreviations used in this paper.

DOF Degrees of freedom

RQIM Random Quad Image Marker

SHT Simple Hough Transformation

RHT Randomised Hough Transformation

PPHT Progressive Probabilistic Hough Transform

LSD Line Segment Detector

LLA Level-Line Angle

GWN Gaussian White Noise

Introduction

Computer vision, and image processing in general, is a computationally intensive area. In the past the use of these algorithms was severely limited by the lack of processing power. Image processing solutions were mostly used for scientific purposes, and the algorithms ran offline: real-time applications were not possible. Satellite photos were analysed, medical imaging solutions were developed at the time. Optical character recognition was also a popular topic for image processing research. A famous scientific example from that time gave the basis for the Hough transformation, which will also be discussed in this work. The transformation was developed to automatically analyse bubble chamber photographs.

With the developing technology, specifically semiconductor manufacturing, more and more possible uses for image processing began to appear. Around the 1970s cheaper computers and dedicated hardware solutions started spreading. This made it possible to create real time image processing applications for some use-cases. One such use-case was television standards conversion.

As general purpose computers became faster and cheaper, they replaced the specialised circuits in almost all areas of application. Nowadays image processing solution to common problems (localisation, mapping, measurement, etc...) is chosen as a solution because it became the cheapest and most versatile alternative. Furthermore, 3D computer vision applications became not only possible, but widespread. 3D scanners, range finders, virtual- and augmented reality solutions have spread from laboratories and research institutions to consumer electronics. Processing power is no longer a bottleneck for most computer vision applications.

Chapter 1

Pose estimation

1.1 Pose Estimation Algorithms

1.1.1 Fast and Globally convergent Pose Estimation

1.1.2 Linear Pose Estimation from Points or Lines

1.1.3 Robust Pose Estimation from a Planar Target

1.2 Comparison

Chapter 2

Markers

One of the goals of this project was to design a fiducial marker with advantageous properties for use in pose estimation. In a typical scenario the marker may be seen from largely varying viewpoints, therefore it has to have some level of scale invariability. If the observer is far from the marker, the smaller details may be lost due to the limited resolution of the camera. If the same observer moves closer to the marker, it may fill the whole field of view and some features may even slip off the image. This leads to another feature the marker needs to have: redundancy. If the observer gets too close to the marker or some obstacle partially blocks the view, the localisation still needs to provide usable results.

The intended use of the markers is spatial localisation and pose estimation. In other words: approximating the observers 3D coordinates (x, y, z) and orientation (ϕ, θ, ψ) with respect to the marker. It is supposed that the observer uses a single camera system for navigation (e.g. smartphone or robotic application with limited resources). This means the marker needs at least 6 degree of freedom.

To sum up the above discussed specifications, a suitable marker would have to:

- have at least 6 DOF
- be (to some degree) scale invariant
- have redundancy

In the following sections will be a recommendation for a marker conforming for the listed specifications. It is based on 3 connected line segments forming a quad with one missing side. The whole marker is built from quads with different side lengths and angles.

2.1 Quad

A marker is put together from quads. Figure 2.1. shows two examples. One side of the quads is left out: they are put together from three joint line segments. The middle segment, with

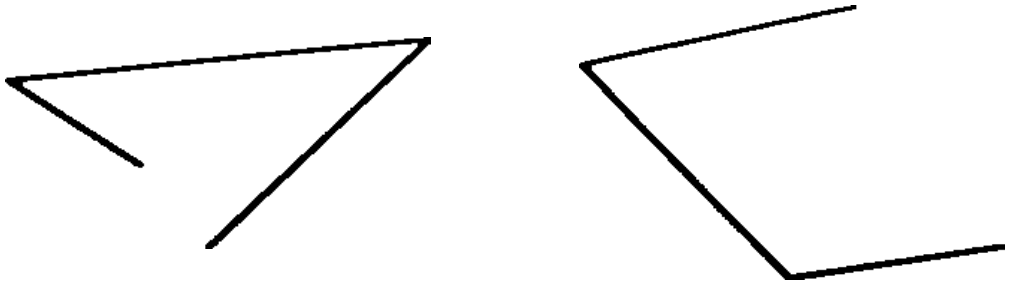

Figure 2.1: *Example for different quads*

two adjoining lines, will be referred to as the 'base' of the quad. The outer segments are going to be called 'arms'.

A quad has 6 degrees of freedom. There are 3 independent distance parameters: the length of the base and the two arm segments. There are also 3 unrelated angle parameters: the angles between each arm and the base, and the orientation of the quad.

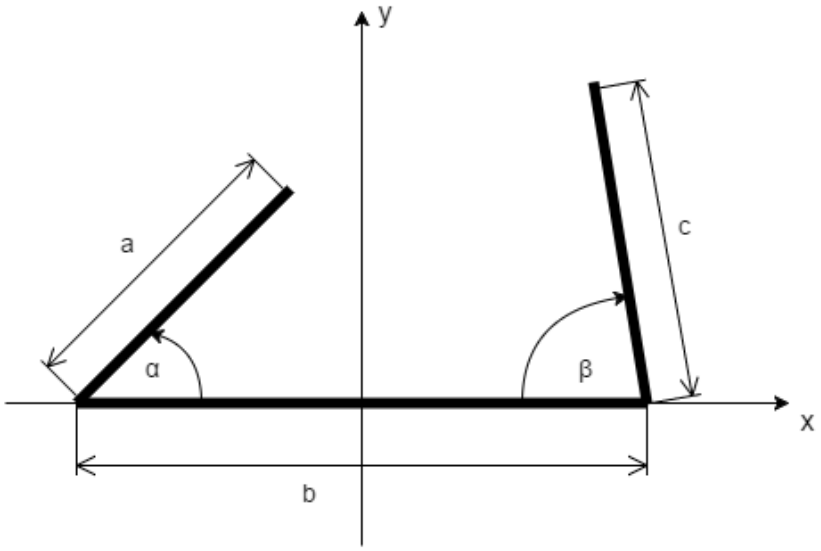

Figure 2.2: *Quad parameters*

Figure 2.2. shows the free parameters of a quad (the orientation is not shown on the image). The following notation is used:

a : The length of one arm

b : The length of the base

c : The length of the other arm

α : The angle between one arm and the base

β : The angle between the other arm and the base

γ : The angle with which the whole quad is rotated

For the sake of simplicity, figure 2.2. does not show the rotation with γ. The quad would be rotated around the origin of it's coordinate system.

The values of the length parameters are given in pixels, although they can be expressed in any unit of distance. The angles can be given is degrees or radians (in the implementation degrees are used for easier human readability).

$$a \in (0, a_{max}] \tag{2.1}$$

$$b \in (0, b_{max}] \tag{2.2}$$

$$c \in (0, c_{max}] \tag{2.3}$$

$$\alpha \in (0, 180^\circ) \tag{2.4}$$

$$\beta \in (0, 180^\circ) \tag{2.5}$$

$$\gamma \in [0, 360^\circ) \tag{2.6}$$

Equations (2.1) through (2.6) specify the range of each parameter. The maximum of the distance parameters are set by the space left on the image for the given marker, there is no theoretical limit for them. There is also no constraint for the resolution of the parameters. From the applications point of view, there are quads with continuous[1] and discrete parameter spaces.

2.1.1 Quad representation

There are several ways to represent quads, each with different advantageous properties. For this work multiple considerations were made in that regard. The most straightforward is to simply store the above mentioned parameters. This is simple and easy for human reading, which is great help in the development process.

A step forward from this is to norm the a, b and c parameter of the quad with the base segment's length. Then the following parameters are used:

s : marker size, the same as the base length

m_a : 'a' multiplier. $m_a = a/b$

m_c : 'c' multiplier. $m_c = c/b$

The α, β, γ angle parameters are not changed. This gives a scale or size parameter for the quad, which is useful for marker generation. These two representations are good for development and marker generation, but not so much for calculations.

[1] That is, only limited by the computational precision

A third option is to store the endpoints of the line segments. It requires the storage of 4 points: two endpoints (E_1, E_2) and two inner points (I_1, I_2). Equations (2.7) through (2.10) define the points' coordinates before rotating with γ, using figure 2.2.'s notation.

$$I_1' = (-\frac{b}{2}, 0) \tag{2.7}$$

$$I_2' = (\frac{b}{2}, 0) \tag{2.8}$$

$$E_1' = (-\frac{b}{2} + b * cos(\alpha), a * sin(\alpha)) \tag{2.9}$$

$$E_2' = (\frac{b}{2} - c * cos(\beta), c * sin(\beta)) \tag{2.10}$$

$$Rot(\gamma) = \begin{pmatrix} cos(\gamma) & -sin(\gamma) \\ sin(\gamma) & cos(\gamma) \end{pmatrix} \tag{2.11}$$

The point E_1, E_2, I_1, I_2 can be obtained from E_1', E_2', I_1', I_2' by a multiplication with the rotational matrix $Rot(\gamma)$.

That method is redundant for storage: it uses 8 parameters instead of 6. However this poses no practical problem in the scope of the project. The one outstanding benefit of this method is it's efficiency in calculations. Because it is based on points in a Euclidean space, linear algebraic methods (matrix multiplications) can be used for calculating the projective transformations.

In this project the second and the third options are used. The first, naive method is omitted because it has no considerable advantage over the other two. The second method, using the size, multiplier and angle parameters is used in marker generation. The third is used during the calculations and the recognition process.

2.2 Marker

Quads are 6 DOF shapes: in theory it would be enough to use only one of them for localisation and pose estimation. However that method would have very low error tolerance and questionable accuracy even in a best case scenario. To comply with the specifications written in the beginning of this chapter, the markers are put together from multiple quads. By placing quads with different orientations and sizes the error tolerance and accuracy can be greatly improved.

An intrinsic positive quality of using multiple quads with varying sizes is the scale invariance. As mentioned, even a single quad is sufficient for the task at hand. If the smaller quads become unrecognisable because of the low resolution or too large distance, a successful measurement is still possible. The same is true on the other end of the spectrum: if the observer is too close to the marker and the larger ones leave the field of view, the position and orientation can be calculated from the smaller quads.

Figure 2.3. shows an example for a marker. It is generated with the simple algorithm described in the next section, and is not optimal in many ways. Nonetheless it is functional, even if only a fraction of the quads are registered for the measurement.

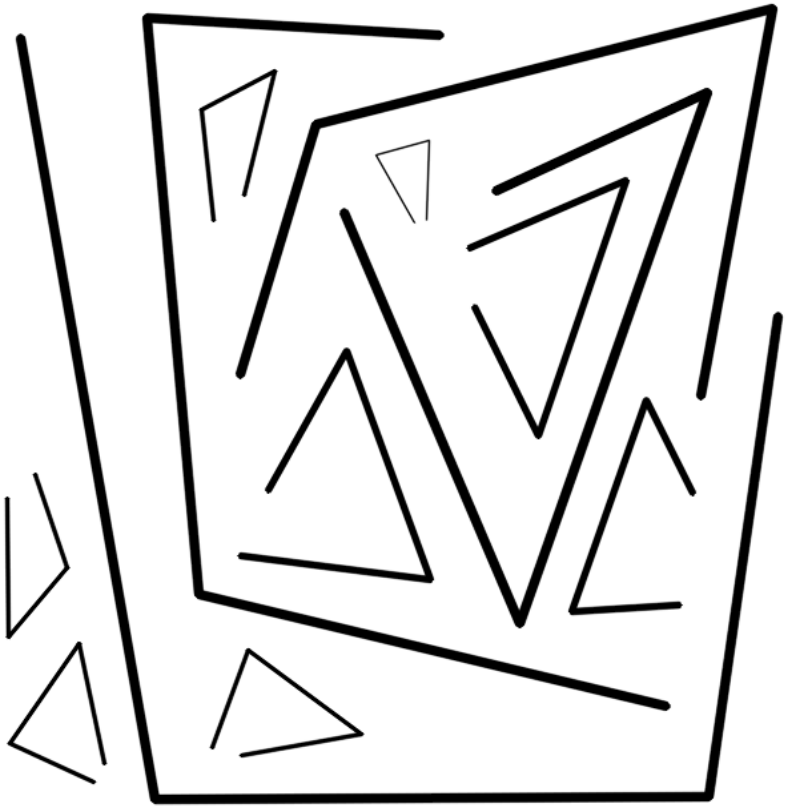

Figure 2.3: *An example for a marker*

The markers are going to be referred to as RQIM, which means Random Quad Image Marker. As the name suggests, the quads are randomly generated and placed on the markers.

Unless otherwise specified, RQIMs use quads with continuous parameter spaces. In this chapter there will also be a small introduction to discrete parameter markers and their potential applications.

2.2.1 Marker generation

In the current state of the project, markers are randomly generated using a simple algorithm. The generator routine receives the number of quads to be used in the current RQIM. The core concept is to create the desired number of random quads and place them on the image.

Let the number of quads to generate be n. First, the quad sizes are picked. There is an upper and a lower limit for them, given in percent of the image size. The generated sizes

follow an exponential distribution:

$$s = e^{-x*f} \tag{2.12}$$

Where s is the quad size, x is random number between 0 and 1 with uniform distribution, and f is a scale factor. Then the n sizes are ordered in descending order.

After the scale factors are picked, the whole quads are generated by the following method. A random quad is created with the first (the largest) scale and placed on the image. Then another quad is created with the next largest size. After every new quad a check is performed whether or not it can be placed on the marker. If it cannot, then a new quad is generated with the same scale factor until it can be placed or the algorithm reached the limit of retries.

With this simple logic n quads are placed on the RQIM and the creation process is finished. Below is the pseudo-code of the algorithm.

```
n_max = number of quads to create
f = scale factor for exponential distribution
lowlim = lower size limit
uplim = upper size limit
n = 0
while n < n_max
        size = exp(-rand() * f)
        if size > lowlim and size < uplim
                store size
                n = n+1
        endif
end
sort(sizes, descending)
n = 0
while n < n_max
        while quad placed or max tries
                quad = create_random_quad(sizes(n))
                if quad can be placed
                        place quad
                        n = n+1
                end
        end
end
return marker
```

This method is not optimal and is based on trial and error, but it gives usable markers for the development process.

2.2.2 Discrete RQIM

There are experiments in progress with discrete parameter space quads. It may be advantageous to quantize the parameter space in order to decrease the error probability in the pose estimation process.

Quads with finite possible states can be stored using much less resources than their continuous counterpart. As an example let us take a look at the following quantisation.

Angles: $15°, 30°, 45°, 60°, 75°, 90°, 105°, 120°$

Multipliers: $0.40, 0.60, 0.80, 1.0, 1.25, 1.50, 1.75, 2.0$

Orientations: $0°, 22.5°, 45°, 67.5° \ldots 270°, 292.5°, 315°, 337.5°$

Sizes: $1, 0.8, 0.6, 0.5, 0.4, 0.3, 0.25, 0.2, 0.1, 0.08, 0.06, 0.05, 0.04, 0.03, 0.025, 0.02$

In this example, there are 8 possible values for the angle parameters, also 8 for the multipliers, 16 for the orientation and also 16 for the sizes. If the possibilities are stored in a lookup table, it is enough for the quad to store the index at which the value is accessible. A quad is defined by two angle parameters, two multipliers, an orientation and a size. The angles (in this case) require at least 3 bits each, the multipliers also. The orientation and the size need 4 bits each. That gives a sum of 20 bits per quad, which is significantly less than the space required to store 6 floating point numbers per quad.

This 20 bit word is also usable as an ID for the quad. It may be possible to code information in these ID-s, so the marker could provide additional information. This information could be related to and used by the localisation process, or be totally unrelated, general data. These possibilities have not yet been extensively researched.

The discrete RQIMs are usually less dense than the continuous ones, due to the limited angle possibilities. This means fewer quads per marker, which leads to decreasing redundancy. An optimum must be found between the number of quads per RQIM and distance between quads in the parameter space.

Chapter 3

Quad detection

In this chapter there will be a summary of the image processing algorithms tried and used for the recognition of the fiducial markers. The input of this recognition step is the image taken by the camera, and the output is a list of quads belonging to the marker visible on the image. As a common preprocessing step for all quad detection methods segmentation is performed on the input image: quad-like blobs are extracted and separately passed to the quad detector logic. This step of the *marker recognition process* takes the segmented input image and initialises quad structures based on the observed picture. The quad structures are then passed to the next processing step: the pose estimation logic.

The process here diverges depending on which quad detection algorithm is used. They all need differently conditioned input images for optimal performance. From a computer vision point of view the task is to detect joint line segments. This is a well researched task in image processing, there are many well tried algorithms for it. For example, the problem can be solved by detecting lines and finding their intersection, or detecting corners and figuring out how they are connected, etc... The detection routines not necessarily have the same output format[1], so conversion may also be needed.

Three separate quad detection techniques and their variants were profiled in this experiment.

- Hough-transformation
- Corner detection
- Line Segment Detector[4]

The first one uses the Hough-transformation for line detection. There are many variants of the transformation: Standard Hough Transform, Probabilistic Hough Transform, Multiscale Hough Transform, etc... The 2 most commonly used are the standard- and the

[1]Some return line segments defined by their endpoint, others use the polar representation of a line etc.

probabilistic variants. The OpenCV framework offers implementations for them, both were tested in the experiment.

The second detector is based on corner recognition. There are more variants of this method to try out, too. The corner metrics of a feature can be calculated differently with (Harris metric, eigenvalues, etc.) varying results. It is also needed for the solution to be scale invariant, which also can be achieved in a number of ways.

The third alternative is the Line Segment Detector algorithm described in [4]. It is a robust and fast algorithm for detecting line segments on an image. The OpenCV framework provides an implementation of it as well.

A typical marker shot with partial visibility is shown figure 3.1.. In this chapter will be a

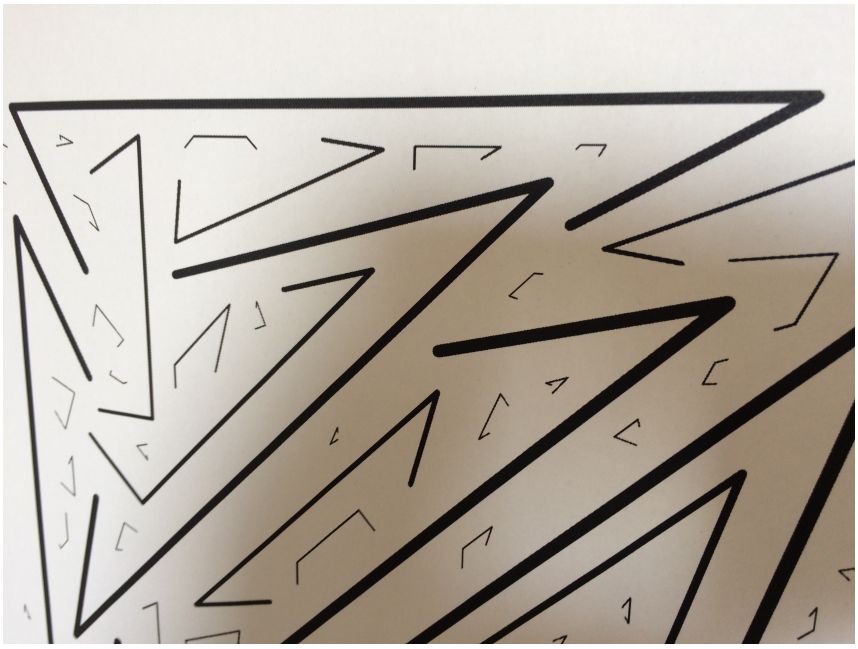

Figure 3.1: *Partially visible marker (taken with commercial smartphone)*

short summary of the algorithms used for testing and performance comparison.

3.1 Theoretical Overview

Before going over how the image processing algorithms were applied to achieve quad detection, a short theoretical overview of the used algorithms will be presented. To solve the problem at hand (i.e. to detect quad instances on an image) multiple well known image processing algorithms were used. For line detection two variants of the Hough transform (Standard[3] and Probabilistic[8]) and a fundamentally different algorithm, the **LSD** was used. As mentioned before, not only solutions based on line detection were tried during the course of this work; corner detection methods were also tried. The Harris corner detector[5] and it's improved version the Shi-Tomasi detector[9] were compared.

Although the implementation of the aforementioned algorithms were provided by the OpenCV framework, it was far from unnecessary to understand how each algorithm works. They show their optimal performance on differently conditioned inputs. For example, corner detection works well on "raw" images, while the Hough-transform based solutions need edge images of skeletons to perform. It was important to know the limitations of each solution. All in all, the understanding of the inner workings of the algorithms used was helpful in choosing the "right tool for the job".

In this section there will be the theoretical overview of the above mentioned algorithms, with some historical context. Their comparative advantages for this project will also be highlighted.

3.1.1 Hough transformation

One of the most commonly used methods for line detection on images is the Hough transform. Over it's long history many publications have been made about it's applications, performance and improvements.

Originally it was developed by Paul Hough in 1959 and later patented in 1962[6]. It was intended to be used for machine analysis of bubble chamber photographs. In it's modern form (with the $\theta - \rho$ parametrisation) was introduced in 1972 by Duda and Hart[3]. The transformation became popular in the image processing community after Ballard's article[1] about generalising the algorithm for detection of arbitrary shapes. There were many optimised and improved variants of the transformation, however the basic concept remained the same. In 1990 a publication[10] introduced the Randomized Hough Transform, which was a fundamentally new approach to the algorithm with notable merits. As opposed to the one-to-many mapping of the simple Hough transform, the randomised version uses a convergent many-to-one mapping when creating the parameter space.

In this work the Standard Hough Transform and one of it's optimised versions, the Progressive Probabilistic Hough Transform will be used. The PPHT, although being probabilistic, doesn't belong to the class of randomised Hough transforms. It uses the same one-to-many mapping as the SHT. The OpenCV framework provides implementations for the SHT and the PPHT, which is one of the main reason why they were chosen for this project.

After this short historical overview the theory of the transformations will be discussed.

Standard Hough Transform

The transformation is used to find instances of a model on digital images. The models are usually simple geometric shapes like lines, circles or ellipses. The curves are described by their parameters, e.g. slope and intercept for a line, centre point and radius for a circle etc.. Every non-zero pixel[2] votes for the features it could be part of. The number of votes is

[2]The transformation works on binary images

stored for every possible parameter combination. Then a threshold is applied to the stored votes, and the remaining parameters are accepted as model instances.

At first Hough described the algorithm to lines, but later the method would be generalised to any analytic[3] curve or shape. This theoretical overview is based on the example of line detection. The process is the same for every analytic curve, the only difference is the parameter space's dimension. The original patent[6] used the slope-intercept representation of lines.

$$y = m * x + b \tag{3.1}$$

In this case, the *parameter space* is 2 dimensional and it's axes are m and b. Every point in the parameter space represent an image space line. With this representation every non-zero pixel in the image space transforms into a line in the parameter space. For a given (x_0, y_0) pair (3.2) gives the line in the parameter space.

$$b = -x_0 * m + y_0 \tag{3.2}$$

Collinear points in the image show up in the parameter space as intersecting lines. The more lines intersect in a given (m_0, b_0), the more likely it is the image contains the $y = m_0 * x + b_0$ line. The problem with this parametrisation is that the parameter space is unbounded along both axes. Both intersect and slope can have values in the range of $(-\infty, \infty)$. Duda and Hart[3] proposed an alternative parametrisation, which turned out to be better for application. They used the *normal parametrisation* of a line, shown in (3.3).

$$\rho = x * cos(\theta) + y * sin(\theta) \tag{3.3}$$

In (3.3) ρ means the distance of the line from the image plane's origin. θ is angle of the normal vector of the line. If the *normal parametrisation* is used the parameter space

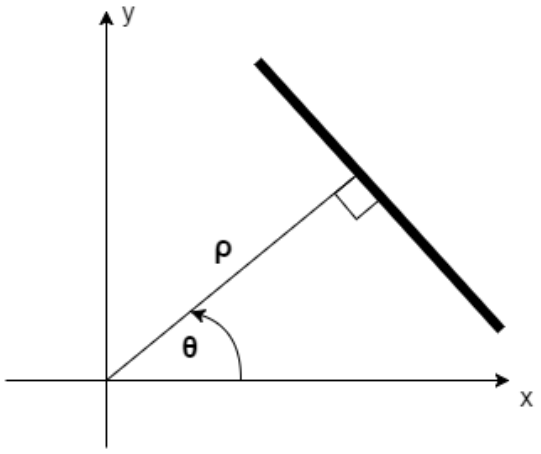

Figure 3.2: *Normal line parameters*

becomes finite in both dimensions. θ is in the range of $(0, 2\pi)$, ρ is bounded by the image

[3]The Generalised Hough Transform even extends to arbitrary shapes

size. In this case the image points define sinusoid curves in the parameter plane, and the line detection is done by searching for their intersections.

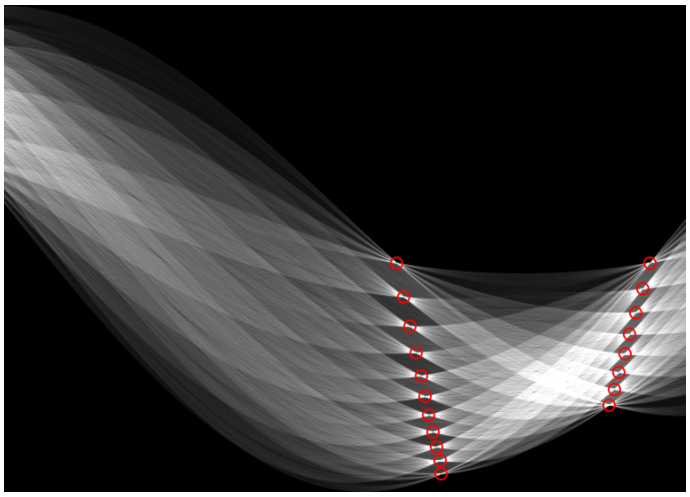

Figure 3.3: *Hough-transform of a chessboard pattern*

As mentioned before, the line detection is based on a voting scheme. The parameter space (in this case a 2 dimensional plane) is divided into *bins*. ρ and θ are quantised in the desired resolution. The discrete (ρ, θ) pairs define the bins. Every bin has accumulator. When a given (ρ_i, θ_i) pair gets a vote it's corresponding accumulator is incremented by 1. The **SHT** (Standard Hough Transform) uses one-to-many divergent mapping. This means that every non-zero pixel votes for every possible parameter pair it could belong to. The above mentioned sinusoid is calculated with the desired resolution for the pixel, and the corresponding accumulators are updated.

When the accumulation phase is completed for the whole image, the local maxima of the accumulators are found. Usually a threshold is applied in order to reduce noise and eliminate too short line segments. The radius of the non-maxima suppression also has impact on the results of the line fitting, it must be chosen carefully. After this step the parameters for the most likely line candidates are available.

As the **SHT** does not provide the endpoints of the line, they must be found by examining the original binary image. This can be done by simple checking every pixel along the line with the given parameters and deciding whether or not it is part of the feature. If it is desired, lines with gaps can also be accepted with this method. For more accurate fitting, a Least Squares approximation can also be applied to the pixels belonging to the line.

Progressive Probabilistic Hough Transform

The progressive probabilistic Hough transform is an optimised version of the SHT described in [8]. Probabilistic Hough transform variants were developed to overcome the comparatively high computational cost of the standard transform. The core concept is the same for most probabilistic versions of the Hough transform: not every none-zero point votes, only

a randomly selected subset. These algorithms have to find a balance between minimising the proportion of image points that are used for voting while maintaining the accuracy of the detection process.

The original probabilistic Hough transform[7] solved this issue by introducing a tunable parameter p for the fraction of points to be used. First, a p fraction of the non-zero points are selected, than the SHT is performed on the selected subset. p can be low, the authors of [7] presented successful experiments with $p = 2\%$. However, the results of the algorithm are greatly sensitive to the sampling rate. The authors analysed the problem on the special case of a single line immersed in noise and tried to formulate a solution for determining the p parameter. They succeeded, but the practical applicability is severely limited[8]: it requires *a priori* knowledge of the number of points belonging to the line. There was another approach to calculate the number of necessary votes[2]. It was shown that the probabilistic Hough transform can be formulated as the Monte Carlo approximation of the SHT, thus it is possible to deduce the desired error rate using the theory of Monte Carlo evaluation. Nevertheless, the core problem remained the same: *a priori* information was necessary for determining the sampling rate parameter. Usually there is only very limited information available, so conservative approximation is needed. This leads to the calculation of more votes then necessary, thus reducing the main advantage of the probabilistic method.

The progressive probabilistic Hough transform solves the above issue by "exploiting the difference in the fraction of votes needed to reliably detect lines (features) with different number of supporting points"[8]. This way for long lines only a small fraction of the line's points have to vote for the line to be registered. For shorter lines this proportion is of course higher. For lines with supporting points close to the votes generated by background noise a full transform must be performed.

The authors of [8] proposed the following algorithm to achieve the aforementioned goal. At each iteration a random non-zero image point is selected for voting to the possible model instances it could belong to. After each vote, the question "could the count be due to random noise?"[8] is evaluated. This requires a single comparison per bin update, with a threshold value changing by each vote cast. When a model instance (line) is detected, the supporting points retract their votes. The other points belonging to the same line are removed from the voting process. The pseudo-code representation below is directly quoted from [8].

```
1. Check input image, if it is empty then finish
2. Update the accumulator with a single pixel randomly selected from the
   input image
3. Remove pixel from input image
4. Check if the highest peak in the accumulator that was modified by the
   new pixel is higher than threshold l. If not then goto 1.
5. Look along a corridor specified by the peak in the accumulator, and find
   the longest segment of pixels either continuous or exhibiting a gap not
   exceeding a given threshold.
6. Remove the pixels in the segment from the input image
7. Unvote from the accumulator all the pixels from the line that have
   previously voted.
```

```
8. If the line segment is longer than the minimum length add it into the
   output list.
9. goto 1.
```

This algorithm has some considerable advantages of the standard and other, previous probabilistic variants of the Hough transform. It eliminates the need of *a priori* knowledge necessary for the tuning of probabilistic transforms while it remains much faster than the SHT. It should detect every instance of a model detectable by the SHT, at the latest when the voting finishes with the same number of voted pixels as for the standard transform. Another positive property of the algorithm is that features are detected as soon as the accumulator allows a decision: it is not necessary for all supporting points to vote. The algorithm can also be terminated at any time and still provide some useful output[4].

Originally this transformation method was developed to speed up the Hough transform, while not being considerably more inaccurate. However, an unexpected result was observed by the authors. The PPHT outperformed the SHT in accuracy as well as speed. In sample images consisting of randomly positioned equal length lines, the PPHT produced less false negatives (missed line segments) and less false positives (incorrectly detected lines). This effect is due to the fact that PPHT clears out the votes of the detected lines as soon as they are found. This reduces the clutter in the accumulator, resulting in more accurate results, while also being more computationally efficient.

It also worth noting that the PPHT could, in theory, use every enhancement that were developed for the SHT. For example, the image gradient of the line segments could be used to reduce the number of pixels selected for voting. However, this aspect was not researched in the boundaries of this project.

3.1.2 Line Segment Detector

A fundamentally different approach to line detection was described in [4]. The algorithm is named **LSD** - for Line Segment Detector - by it's creators. It is which, unlike the SHT, detects line segments with subpixel accuracy by default. The runtime of the process is linear in the pixel count of the processed image. It also has fairly good noise suppression. Another attractive property of the algorithm is that it doesn't have any parameters that require tuning by the user. Every one of it's parameters are automatically tuned "under the hood". Because of these advantageous properties was it considered for use in this project. The implementation used was provided by the OpenCV framework. In this section will be a short summary of the theory behind this algorithm.

The LSD takes as input a grayscale image and provides a list of line segments as output. The line detection is based on the image gradient. As a first step, a gradient field is generated

[4]However this aspect is not really important for this project

from the input image. The gradient is taken using a $2x2$ window, see (3.4).

$$\begin{aligned} g_x &= \frac{i(x+1,y) + i(x+1,y+1) - i(x,y) - i(x,y+1)}{2}, \\ g_y &= \frac{i(x,y+1) + i(x+1,y+1) - i(x,y) - i(x+1,y)}{2} \end{aligned} \tag{3.4}$$

Where $i(x,y)$ is the intensity of the grayscale image at (x,y) point. The magnitude of the gradient is calculated by (3.5).

$$G(x,y) = \sqrt{g_x^2(x,y) + g_y^2(x,y)} \tag{3.5}$$

The algorithm uses the angle of the gradient, which will be referred to as LLA (level-line angle), and is calculated by (3.6).

$$\arctan\left(\frac{g_x(x,y)}{-g_y(x,y)}\right) \tag{3.6}$$

The gradient obtained with (3.4) is the image gradient at the point $(x+0.5, y+0.5)$. This half pixel offset is later added to the endpoints of the detected line segments.

Using the gradient information, a level-line field is constructed. Figure 3.4. shows an example of the visualised level-line field. The next step of the algorithm is the segmentation of this field. It happens based on the level-line angle (the gradient angle defined in (3.6)). The pixels that have the same LLA within a given threshold are grouped together. These segments are referred to as *line support regions*, see figure 3.4. for illustration. The segmentation is done with a region growing process.

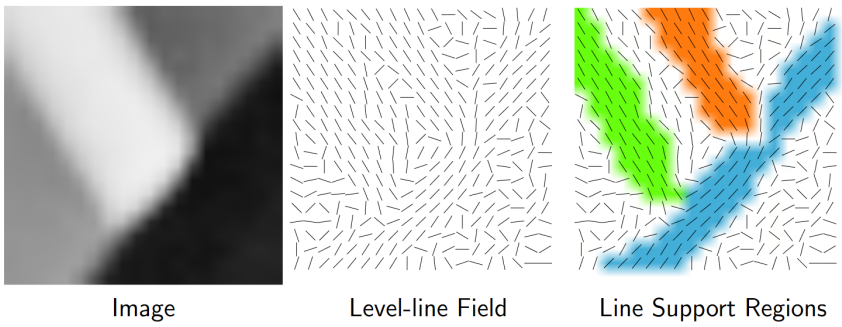

Figure 3.4: *Illustration of the level-line field[4]*

Each *line support region* is a candidate for a *line segment*[4]. The line segments are represented with a rectangle. The main direction of the rectangle is determined by the principal inertial axis of the *line support region*. The size of the rectangle is chosen in a way to cover the whole *line support region*.

The pixels in the rectangle that have LLA close to the angle of the rectangle are called *aligned points*[4]. Figure 3.5. shows an example for the rectangular representation and the *aligned points*. The *aligned points* are used in the validation step of the algorithm.

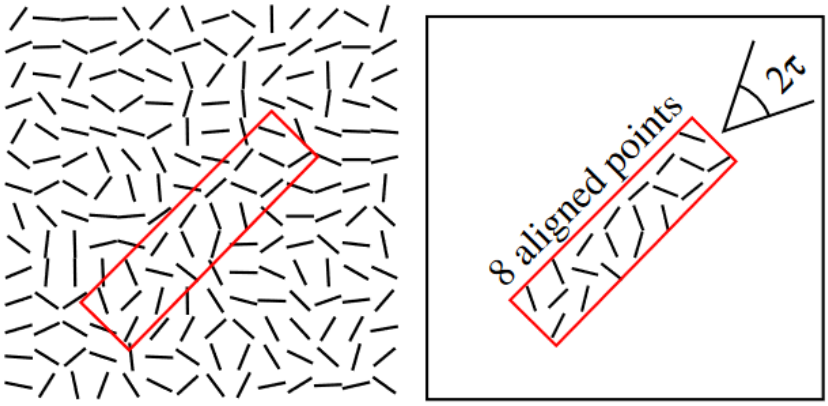

Figure 3.5: *Illustration of the aligned points[4]*

The LSD algorithm uses an *a contrario* validation method. The idea behind that method is checking if it is probable that the current supporting points are caused by random noise. To achieve this, the authors of [4] created a noise model of the **level-line field**. A *line segment* becomes validated if the expected number of it's occurrences on the noise model is low[5].

The algorithm detects the sharp transients in the image gradient. Technically, it detects edges. A line on the image produces two line segments as output, for it's two light-dark transition. The line segments detected by LSD are directional: the order of the endpoints of a line segments depend on the direction of the light-dark transition.

After this short summary of the algorithm[6], some of it's more interesting details will be described.

First off, the algorithm has a preprocessing step. Before calculating the image gradient, the input image is downscaled to 80% along both axes[7]. This is done to cope with aliasing and quantisation artefacts present in most images, for example the staircase effect. The alternative to this subsampling would be the blurring of the image, however that would have some unfavourable side effects. Blurring would affect the statistics of the *a contrario* model. Some structures would be detected in a blurred white noise image. With a correct down-sampling the white noise statistics can be preserved. The choice of scale factor was an optimum between filtering out noise and keeping valuable data.

Another interesting feature of the algorithm is the order in which the possible lines are processed. LSD is a greedy algorithm, it tries to process the most significant edges first. Pixels with higher gradient magnitude correspond to more contrasted edges. In order to process the pixels with the highest contrast first, some ordering is needed. However, most sorting algorithm require $O(n \log(n))$ operations. To avoid this, LSD uses a pseudo ordering that can be done in linear time. The interval between zero and the highest gradient

[5] i.e. It is unlikely to be caused by random noise
[6] The description of the algorithm in pseudo-code form can be found in [4]
[7] or to 64% of it's area

magnitude in the image is divided into 1024 equal bins. Then each pixel is assigned to the bin corresponding to it's gradient magnitude. The processing (region growing) is done first on the pixels selected from the bin containing the largest magnitudes. 1024 levels are enough to almost strictly order the gradients generated from a grayscale image with 256 possible intensities.

To avoid unnecessary processing, a threshold is also applied to the gradient magnitudes. Pixels with a low gradient represent flat regions or slowly changing intensities. These pixels are marked and are not taking part in the later processing steps. This threshold also helps reduce the effects of quantisation noise.

The rectangular approximation of the line segment happens after the segmentation of the level-line field. The rectangle is calculated based on the gradient magnitudes of the pixels belonging to a segment. The gradient magnitude is viewed as the "mass"[4] of the pixel, and the centre of the rectangle is the mass centre point of the segment. The coordinates of the centre point are calculated by the formula given in (3.7)

$$\begin{aligned} c_x &= \frac{\sum_{j \in Region} G(j) * x(j)}{\sum_{j \in Region} G(j)} \\ c_y &= \frac{\sum_{j \in Region} G(j) * y(j)}{\sum_{j \in Region} G(j)} \end{aligned} \tag{3.7}$$

Where $G(j)$ is the gradient magnitude of pixel j, calculated by (3.5). $x(j)$ and $y(j)$ represent the x and y coordinate of point j, respectively. The angle of the main rectangle is defined to be the principal inertial axis of the segment. It can be calculated from the eigenvector of associated with the smallest eigenvalue of the matrix of (3.8).[4]

$$M = \begin{bmatrix} m^{xx} & m^{xy} \\ m^{xy} & m^{yy} \end{bmatrix} \tag{3.8}$$

Where $m^{xx}, \ldots$ is defined below.

$$\begin{aligned} m^{xx} &= \frac{\sum_{j \in Region} G(j) * (x(j) - c_x)^2}{\sum_{j \in Region} G(j)} \\ m^{yy} &= \frac{\sum_{j \in Region} G(j) * (y(j) - c_y)^2}{\sum_{j \in Region} G(j)} \\ m^{xy} &= \frac{\sum_{j \in Region} G(j) * (x(j) - c_x)(y(j) - c_y)}{\sum_{j \in Region} G(j)} \end{aligned} \tag{3.9}$$

This is the short overview of the LSD algorithm, with some of it's more interesting nuances highlighted. The full description is available in [4].

3.1.3 Corner Detection

Detecting quads not necessarily means the detection of line segments. Along with the above described methods based on line detection, a corner detecting algorithm was also benchmarked. The concept of this detection method is as follows. Detect the corners and end-points of a quad with some corner detection algorithm. Checks which detected pairs are connected with lines (or edges). Based on the connected pairs and their ordering, a quad can be reconstructed.

The OpenCV framework provides implementations for some popular corner detection algorithms. Specifically the Harris detector and the Shi-Thomas detector are covered. In this section will be a short theoretical summary of corner detection in general, and some specifics of the above mentioned solutions.

The basic idea of most corner detection algorithm is the following. Considering a local window on an image, corner regions show large change in average intensity if the window is shifted by a small amount in any direction. The mathematical formulation of the following idea is shown in (3.10). $E_{x,y}$ is the change in intensity produced by shifting the window by (x, y). $I_{u,v}$ is the intensity of the image at the point (u, v), and $w_{u,v}$ specifies the image window. In the simplest case, the image window is rectangular and it is unity in a specified region and zero otherwise.

$$E_{x,y} = \sum_{u,v} w_{u,v} |I_{x+u,y+v} - I_{u,v}|^2 \tag{3.10}$$

A naive approach of corner detection is to use (3.10) as it is. The local maxima of the minimum of (3.10) above a certain threshold can be used for a metric. With this method, three cases have to be considered:

1. **Flat region:** The windowed image region has almost constant intensity. In that case, all shifts will show small change.
2. **Edge region:** The windowed image region has an edge in it. In that case shift in one direction will result in large change, but shifts in other directions will show low change in intensity.
3. **Corner region:** If the windowed region contains a corner, all shifts will show large change in the intensity.

The shifts can be chosen in a couple of ways: 90° shifts, 45° shifts in 8 or 4 directions, etc... Actually, this detection method is analysed in [5], and it is the base of the Harris detector.

However, the the above described corner detector suffers from a number of problems[5]. Firstly, it provides an anisotropic response, as only a discrete set of shifts are used. To

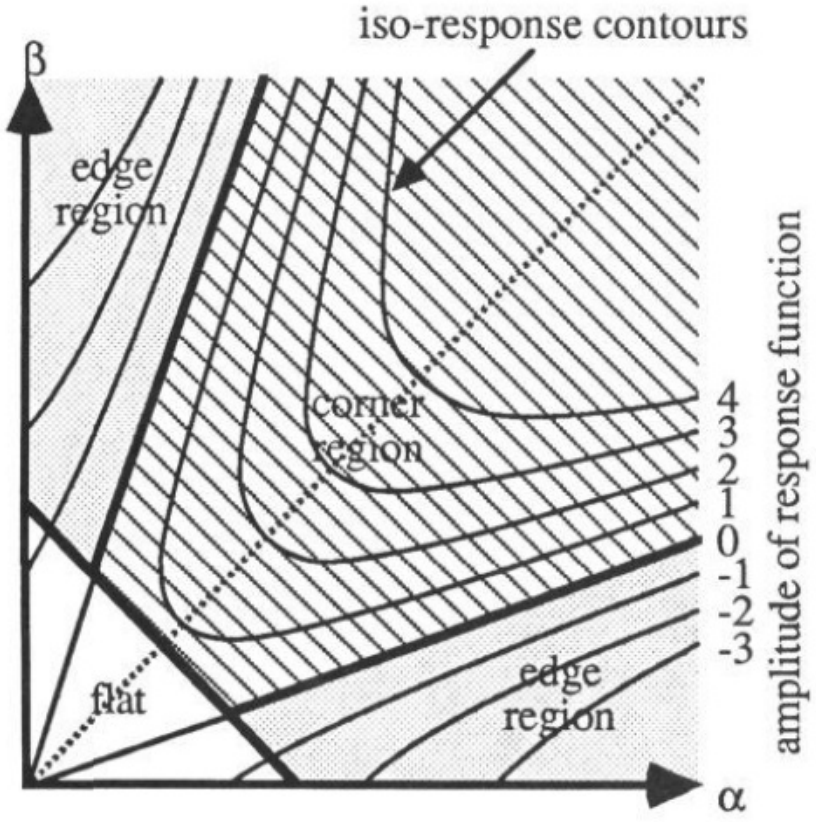

Figure 3.6: *Image point classification based on Harris measure[5]*

address this issue, the Harris detector uses an analytic expansion of (3.10) around origin. See (3.11).

$$E_{x,y} = \sum_{u,v} w_{u,v}(I_{x+u,y+v} - I_{u,v})^2 = \sum_{u,v} w_{u,v}(x\frac{\partial I}{\partial x} + y\frac{\partial I}{\partial y} + O(x^2, y^2))^2 \tag{3.11}$$

Another problem is that the above detection method's response is noisy[5] because of the rectangular and binary image window. The Harris detector resolves this issue by using a circular and smooth (for example Gaussian) window:

$$w_{u,v} = e^{-\frac{u^2+v^2}{2\sigma^2}} \tag{3.12}$$

The third issue Harris found with the use of (3.10) as a corner metric is that it responds too readily t edges[5]. This is because only the minimum of E is taken into account when deciding whether a sampled window contains a corner or not. To address this issue, a reformulation of the corner measure was proposed that takes the variation of E with the direction of the change into consideration. For small changes, E, the average change in intensity generated by a shift with (x, y) can be written as:

$$E(x, y) = (x, y)M(x, y)^T \tag{3.13}$$

Where M is composed of the image gradients in the window. (3.15) shows M using the

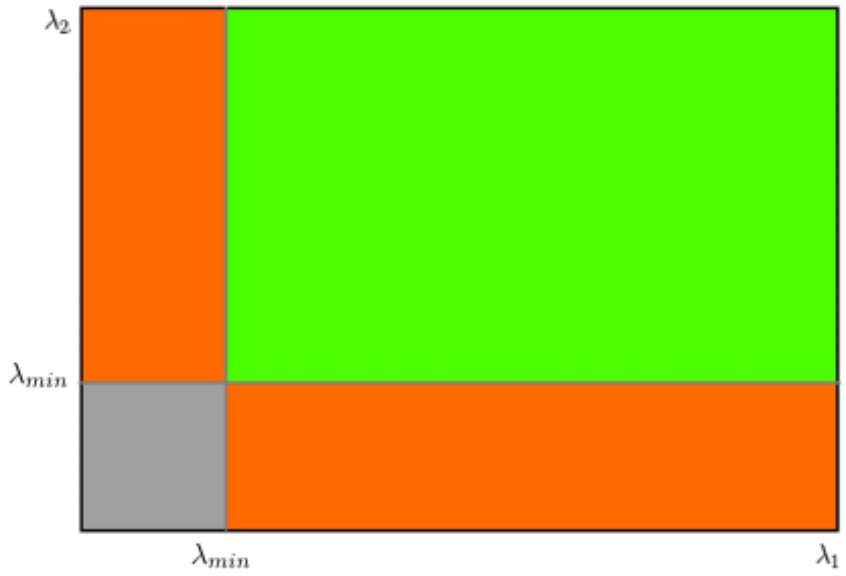

Figure 3.7: *Image point classification based on Shi-Tomasi measure. Image source: OpenCV documentation*

notation of (3.14)

$$I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y} \tag{3.14}$$

$$\begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{3.15}$$

Note that M describes the shape of the local autocorrelation function's shape at the origin. To describe E, [5] uses the eigenvalues of the matrix M, as it provides a rotationally invariant description. With this new method, the above mentioned cases (flat region, edge, corner) can be expressed as follows.

1. **Flat region:** Both eigenvalues are small.
2. **Edge region:** One eigenvalue of M is small, the other is comparatively large.
3. **Corner region:** Both eigenvalues are large.

Figure 3.6. shows the above defined regions with respect to the eigenvalues (α and β are the eigenvalues of M). The Harris detector also uses a metric for the "quality" of the detected edges or corners. This is noted with R and is defined below.

$$R = \det(M) - k(Trace(M))^2 \tag{3.16}$$

k is a tunable parameter, its value is usually in the range of $0.01 - 0.05$. The higher R is, the more likely that a corner is present in the sampled window. In figure 3.6. the curves mark the regions in the $\alpha - \beta$ space that have the same R value.

This is the short theoretical summary of the Harris detector. A detailed mathematical derivation of the formulae can be found in [5].

The other corner detection method provided by the OpenCV framework is the Shi-Tomasi detector[9]. This detector uses the same concept as the Harris detector, however uses a different measure to classify features as corners. The scoring function is defined in (3.17). λ_1, λ_2 is the same as α, β for the Harris detector.

$$R = \min(\lambda_1, \lambda_2) \tag{3.17}$$

A feature is classified as a corner if R is greater than a threshold λ_{min}. Figure 3.7. shows this relation in the $\lambda_1 - \lambda_2$ space. The Shi-Tomasi measure, while being simpler and requiring less computational power, shows better performance in images[9].

3.2 Application for Quad Detection

3.2.1 Conditioning

Before running the line detection algorithms some conditioning steps are done in order to improve their effectiveness. These processes are not uniform - every fitting method needs it's own.

The Hough transformation traditionally works best on thin lines. The easiest way is to generate an edge image with high-pass filtering. The OpenCV framework offers a wide variety of features for this task. The best result are obtained by using the Canny edge detector.

The skeletoning detector does not need a conditioning step, as it performs the band thinning on it's own.

The methods based on image gradients and corner detection both require some level of smoothing on the picture. In case of the corner detection the smoothing is useful for removing the false positive matches caused by the jagged edges, or in case os JPEG images the artefacts caused by the compression. The gradient detector simply gives a more spread-out and easier to analyse result on a smoothly changing gradient than it would on strict edge. The smoothing is also implemented using the OpenCV framework, which provides easy access to Gaussian filtering. The OpenCV implementation is based on convolution with a configurable Gaussian kernel. The kernel size and the deviation in x and y direction can be set. The Gaussian kernel and the convolution itself is handled in the framework.

3.3 Performance comparison

3.3.1 Test method

In this project multiple quad detection solutions are prototyped. The best is needed to be selected. It is not a trivial choice, as there are many criteria to be satisfied. The algorithm

needs to be accurate, robust, relatively fast, work on noisy images, etc... Thorough testing is necessary to select the best algorithm for the task. The tests should be reproducible and statistically relevant. In this section will be a description of the testing method used in this project.

As a first step, a data source is necessary for the algorithms under test. For this, randomly generated quads were used. The random generator was constructed in a way that allowed some control over the generated data: the *wuad size* and the range of the other parameters could be set. To provide data for extensive testing, a large data set was generated. Every algorithm received random quads with sizes ranging from 1% to 75% of the image size, in 1% increments. From each size 1000 quads were generated and provided to the detectors. These values were chosen to cover a significant portion of the parameter space and to provide statistically significant results.

The generated quads were rendered by OpenCV and the generated images were the inputs for the detection algorithms. The testing was done with images containing only a single quad. There are many reasons for this. First, it is important to limit the test scope. In this step the quad detection was benchmarked. not the segmentation logic. In the real use-case, the input image containing a marker is segmented, and the segments are fed to the quad detector separately. The images were rendered with different levels of additive Gaussian White Noise. The test series were run first on optimal images, and after that rerun with more and more added noise. This was done to test the robustness. Although the GWN covers only a small portion of the noise in a real image, these test did provide some insight.

The accuracy of the detection was analysed based on many different error measure. Those are defined in the next section. The experiments measured the expected value and the standard deviation of error of the algorithms. These were plotted at the end of the test cycle.

The test method is summarised by the following pseudo-code program.

```
quad_groups = empty_list
for size in [0.01:0.01:0.75]:
        group = generate_quads(count=1000, size=size)
        quad_groups.append(group)

error_series = empty_list
for group in quad_groups:
        pairs = empty_list
        for quad in group:
                img = render(quad)
                img = add_noise(img, noise_level)
                detected_quad = detector.detect_quad(img)
                pairs.append(quad, detected_quad)
        group_error = calculate_error(pairs)
        error_series.append(group_error)
display_result(error_series)
```

3.3.2 Error measure

In order to compare the performance of the quad detection algorithms some kind of scoring function is needed. For this the use of detection error seems intuitive. As a quad has 6 independent parameters, and these are stored in 2 different representations, defining an error measure is not straightforward. Because of this, several scoring functions were defined during development, and many of those are actually used for comparing the algorithms.

As described in the previous section, the calculation of detection error is based on rendering a known quad and running the detection algorithms on it. After the detection is done and it is successful[8], some kind os measure is necessary to calculate the "distance" of the detected and the original quad. Below will be the definitions the error measures used within this project to compare the quad detection algorithms. The notations used in the formulae are listed here.

- Q^o, Q^d: Original quad, detected quad
- Q_p: Parameter p of quad Q, where p can be any of the followings: $\{s, m_a, m_b, \alpha, \beta, \gamma\}$.
- C^o, C^d: Corner set of the original and the detected quad
- C^o_i, C^d_i: The i. corner of the original and the detected quad
- $C_{i,x}$: x coordinate of the i. corner of a quad
- P^o, P^d: the parameter space of the original and the detected quad. Using Section 2.1's notation, it can be defined as $P = \{s, m_a, m_b, \alpha, \beta, \gamma\}$

As mentioned above, there are two different quad representations used[9]. One stores the coordinates of the corners, the other the quad parameters. The most intuitive way for error calculation is based on the corner representation. We can calculate the distance between the detected and the original corner. (3.18), (3.19) and (3.20) define 3 scoring functions based on this idea.

$$E_{c,abs} = \sum_{1}^{4} \sqrt{(C^o_{i,x} - C^d_{i,x})^2 + (C^o_{i,y} - C^d_{i,y})^2} \tag{3.18}$$

The first possibility is to calculate the sum of the distance between the detected and the original corners. This naive approach has many drawbacks. The main concern is that it does not take into account the *size* of the quad. The cumulation of error from every corner also distorts the results. In reality, if every corner has some amount of noise in it's position, the quad parameter representation can still be quite close to the original. However, if only

[8] A quad instance is returned, which is not guaranteed
[9] for details, see Section 2.1

one corner has a larger amount of error, the distortion will be much higher. This error measure reports the same amount for both cases.

$$E_{c,avg} = \frac{1}{4} E_{c,abs} \tag{3.19}$$

The above points are also true is the average of the absolute displacements are used.

Better results can be achieved by using relative coordinate error. The formula used by this work can be seen in (3.20). The problem of the error depending on the quad *size* is solved by this. As seen on the formula, the coordinate error is compared to the coordinates of the original quad corners.

$$E_{c,rel} = \frac{1}{4} \sum_{1}^{4} \frac{\sqrt{(C^o_{i,x} - C^d_{i,x})^2 + (C^o_{i,y} - C^d_{i,y})^2}}{\sqrt{(C^o_{i,x})^2 + (C^o_{i,y})^2}} \tag{3.20}$$

(3.20) was chosen for the comparison of algorithms based on the accuracy of corners detected.

The following scoring functions are based on the other quad representation. This has considerable advantages compared to the corner based representation, because a much clearer picture of the distribution of error factors is obtained. The quad parameters can be classified into 3 sets. The is the quad size, which is an absolute length, measured in pixels. Another category is formed of the angle parameters: the angle between the *base* and each *arm*, and the orientation (which is basically the angle between the *base* and the image frame). The third is the multiplier parameter. These are not absolute lengths, they are calculated based on the base length. Below will be proposed error measures for the three categories.

For the angle parameters, as a first approach an absolute error measure was defined. As (3.21) shows, the 2 angle errors are summed. As an alternative, their average also can be used. Contrary to the coordinate errors, here the absolute angle error does not depend on the size of the quad. These scoring functions can be useful if the absolute magnitude of the error is of interest.

$$E_{a,abs} = |Q^o_\alpha - Q^d_\alpha| + |Q^o_\beta - Q^d_\beta| \tag{3.21}$$

$$E_{a,avg} = \frac{1}{2} E_{a,abs} \tag{3.22}$$

Of course, relative error can also be defined for the angles, too. See (3.23). This was used for comparison, because the relative values given as percentages were easier to evaluate.

$$E_{a,rel} = \frac{1}{2} \left(\frac{|Q^o_\alpha - Q^d_\alpha|}{|Q^o_\alpha|} + \frac{|Q^o_\beta - Q^d_\beta|}{|Q^o_\beta|} \right) \tag{3.23}$$

The multipliers are very similar to the angles in terms of error metrics. They describe a relative parameter, so the quad scale is not an issue here. Nonetheless, both absolute and

relative error measures were defined for completeness. Similarly to the angles, both the sum os absolute errors and their average can be meaningful, depending on what is the goal of analysis.

$$E_{m,abs} = |Q^o_{ma} - Q^d_{ma}| + |Q^o_{mb} - Q^d_{mb}| \tag{3.24}$$

$$E_{m,avg} = \frac{1}{2} E_{m,abs} \tag{3.25}$$

For readability, the relative measure was chosen as a basis for comparison. The values are normed with the respective parameter of the original (generated, thus it's parameters are known to an arbitrary level of precision) quad.

$$E_{m,rel} = \frac{1}{2}\left(\frac{|Q^o_{ma} - Q^d_{ma}|}{|Q^o_{ma}|} + \frac{|Q^o_{mb} - Q^d_{mb}|}{|Q^o_{mb}|} \right) \tag{3.26}$$

Orientation is handled differently from the other angle parameters. This is due to two reasons. First, it's conceptually different. It describes a rotation as opposed to α and β that describe angles between lines segments. The second reason is mainly empirical: it turned out to be less error-prune than the other two.

$$E_{o,abs} = |Q^o_{\gamma} - Q^d_{\gamma}| \tag{3.27}$$

$$E_{o,rel} = \frac{|Q^o_{\gamma} - Q^d_{\gamma}|}{|Q^o_{\gamma}|} \tag{3.28}$$

Similarly to the previously discussed categories, the absolute and relative errors are defined. Equations (3.27) and (3.28) show the definitions. For comparison, as before, the relative error was used.

The *base length* or *size* is the only absolute length parameter in this quad representation. It makes sense to handle it separately, because many other parameters (orientation, arm multipliers) depend on it. The scoring functions are defined below, similarly to the previous ones.

$$E_{o,abs} = |Q^o_{s} - Q^d_{s}| \tag{3.29}$$

$$E_{o,rel} = \frac{|Q^o_{s} - Q^d_{s}|}{|Q^o_{s}|} \tag{3.30}$$

For consistencies sake, also the relative error was used here for comparison.

The above defined error functions provide useful information if the distribution of error between the quad parameters is interesting. However, this parametric quad representation lacks a scoring function that would provide information on the error magnitude as a whole. To resolve this, one more error measure is proposed and used by this work. The independent parameters of a quad can be viewed as coordinates in a 6 dimensional space. The analogy stands, as the parameter space is a subset of $\mathbb{R}^6$. A distance in that 6-dimensional space

can be defined, see (3.31).

$$E_{sum} = \sqrt{\sum_{p \in P^o, q \in P^d} (p - q)^2} \tag{3.31}$$

This gives useful information about the absolute magnitude of error in the parametric wuad representation. Of course, the relative version of the above error measure can also be defined.

$$E_{sum,rel} = \frac{\sqrt{\sum_{p \in P^o, q \in P^d} (p - q)^2}}{\sqrt{\sum_{p \in P^o} p^2}} \tag{3.32}$$

With this, a complete set of error measurement functions are defined.

It is reasonable to use both quad representations, thus both sets of error measures. The corner representation is intuitive. Also, as it is stated in the previous chapters of this work, the pose estimation algorithms use corresponding point pairs. So the most relevant error component seems to be the one present in the quad corner coordinates, as it directly influences the accuracy of the calculated camera pose.

However, the two representations are equivalent. It is possible to convert between the two without loosing useful information. If discrete markers are used, with the parametric representation it is possible to further refine the detection results. This can be done by replacing the detected quad with the closest known discrete one[10]. Currently this aspect of the marker is not implemented, it can be a subject of future improvements.

[10]This will be elaborated later on. This sentence is just to give a general idea and is not technically correct.

Chapter 4

Marker Recognition

The preprocessing steps used for preparing the images for the line fitters (segmentation, thresholding, filtering etc.) will also be discussed.

The input is the raw image taken[1] by the observer. The first problem is finding the RQIM on the picture. When the marker area is located, it is necessary to discard the only partially visible and/or unrecognisable quads. At this point there is an image or set of images containing potentially good quads.

4.1 Preprocessing

The preprocessing is done in two stages. The first is the segmentation, when quad-like blobs are found. The second step is the preparation of the aforementioned blobs for the line fitting algorithms' needs.

4.2 Segmentation

The segmentation process is carried out on images roughly like the one shown in figure 3.1.. First the photos are converted to binary format by applying a threshold. The image is inverted in the process, because it makes more sense for the objects to be marked with non-zero elements than vice-versa. The threshold's value is determined using Otsu's method, which maximises the inter-class variance of the clusters[2]. The implementation is provided by the OpenCV framework.

Afterwards, the binary image is conditioned with a *close* morphology operator. The closing removes the gaps from the large connected areas (possible quads) and removes the *salt and pepper*-like noise. In the current implementation the kernel size of the morphology operator

[1] In the development phase rendered pictures were used for better repeatability
[2] Foreground and background

is constant, however it could be beneficial to calculate it from the global or local image parameters[3].

The segmentation is based on finding continuous contours on the binary image. The OpenCV framework provides great functionality for this. The implementation is based on calculating the 8-neighbour chain code for the binary blobs on the image. The functions returns a list of list of points for the borders os each distinct contour.

The next step is the filtering of the found blobs. First the surely partial quads are discarded. This is done by calculating the bounding box of the contours, and if one of it's sides are touching the image border, the blob is marked as partial. With this approach it is possible that some fully visible quads that only touch the image border with one of their corner are lost. This problem can be easily fixed by checking the neighbourhood of the contact point, but this is not yet implemented.

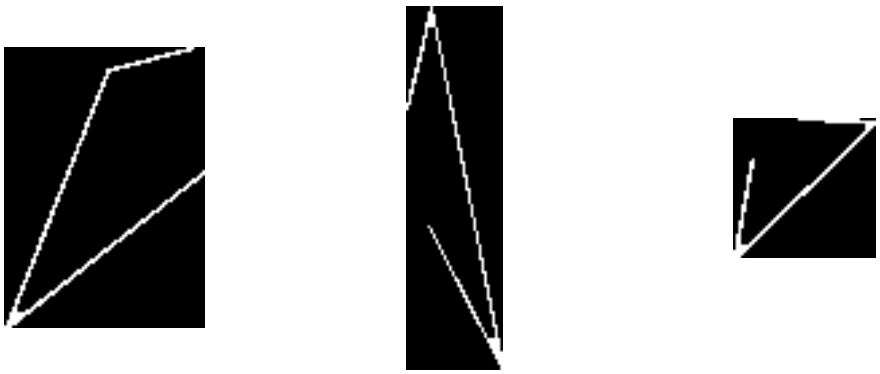

Figure 4.1: *Quad candidates after segmentation*

The next blob-filtering step is to filter out the false-positive contours. These false hits can be caused by the light conditions or the scene around the marker. For this purpose a simple metric is used to measure how likely a blob is to be a quad. This metric is the ratio of a blob's area and circumference. By experimentation this ratio for quads is found to be in the range of 10 and 50. The contours with ratios outside these limits are discarded.

The segmentation processes output is available as a singe image with colour-coded[4] blobs or as a list of separate images each containing a quad candidate. Figure 4.1. shows an example for the output of the segmentation process.

[3] e.g. image size, area of the connected region, etc.
[4] Gray level, to be exact

Bibliography

[1] D.H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111 – 122, 1981.

[2] James R Bergen and Haim Shvaytser (Schweitzer). A probabilistic algorithm for computing hough transforms. *Journal of Algorithms*, 12(4):639 – 656, 1991.

[3] R. O. Duda and P. E. Hart. Use of the hough transformation to detect lines and curves in pictures,. *Comm. ACM*, pages 11–15, January 1972.

[4] Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. Lsd: A fast line segment detector with a false detection control. 32:722–32, 04 2010.

[5] C. Harris and M. Stephens. A combined corner and edge detector, 1988.

[6] P.V.C. Hough. Method and means for recognizing complex patterns. *U.S. Patent 3,069,654*, Dec. 1962.

[7] N. Kiryati, Y. Eldar, and A.M. Bruckstein. A probabilistic hough transform. *Pattern Recognition*, 24(4):303 – 316, 1991.

[8] J. Matas, C. Galambos, and J. Kittler. Robust detection of lines using the progressive probabilistic hough transform. *Computer Vision and Image Understanding*, 78(1):119 – 137, 2000.

[9] Jianbo Shi and Carlo Tomasi. Good features to track. pages 593–600, 1994.

[10] Lei Xu, Erkki Oja, and Pekka Kultanen. A new curve detection method: Randomized hough transform (rht). *Pattern Recognition Letters*, 11(5):331 – 338, 1990.