Tampereen yliopisto

Roni Kilpeläinen

PERCEPTION TOOLS AND THEIR USAGE IN SPORTS

Bachelor of Science Thesis
Faculty of Engineering and Natural Science Thesis
Examiner: Roel Pieters
May 2023

ABSTRACT

Roni Kilpeläinen: Perception tools and their usage in analyzing sports
Bachelor of Science Thesis
Tampere University
Bachelors's Degree Programme in Automation Engineering
May 2023

Computer vision and perception tools are an important step on the developing world of robotics. The goal of computer vision development is to enable machines to understand visual information. When this goal is acquired, there are countless usages in fields such as production, healthcare, and entertainment like sports. This thesis addresses perception tools, computer vision and especially their usage amongst sports.

The thesis can be roughly divided into two parts, the first part is literature research about computer vision and perception tools. The second part of the thesis is a practical example with OpenDR's pose estimator tool, which is used to do a short analysis on two separate golf swings. These swings are then compared to each other and by this one way of using perception tools in sports is showcased.

First part of the literature research addresses the basics of computer vision. The beginning of this part introduces what computer vision and perception tools are. In addition, a rough version of a vision system is presented. After that we move on to how computers process images. Briefly computers segment the picture and then give a value to each part that corresponds into a color. After image processing we switch over to 3D-vision and present three different ways on how it is executed. The methods that were presented are, triangulation, time-of-flight, and structured light. Lastly, we researched on how object tracking is implemented on vision systems. We found out that no one way of executing object tracking is set in stone, and there are available several different algorithms. One common factor in these algorithms was the need of background filtering. Usually this is done by comparing subsequent images and separating objects that are in motion from those that are motionless. At the end of this part project OpenDR is also showcased, which is used in the practical part of the thesis.

The second part of the literature research contains presentations of different systems that are already used in sports. In total 4 systems were presented, each of them focusing on a different part of computer vision. The presented systems were used in hockey, tennis, cricket, and badminton. The purpose of these systems include improvement in training efficiency, added entertainment value, and increasing the accuracy of officiating. Some of the systems are already in commercial use and some were still in development. Although the purpose of these systems is to develop sports in general, there are some difficulties. These difficulties include technical ones like accuracy of the system or price, and impact on the entertainment value by disrupting the flow of the game.

On the last part of the thesis a pose estimator algorithm was used to detect joints and other keypoints from images. These keypoints were used to compare two different golf swings and to detect differences in these swings. These differences were then used to consider on what should be changed on the swing to acquire better results. The algorithm made it easier to detect the differences in the swings and we concluded that with more data such as different angles, in addition to better sport knowledge, more accurate advice could be provided.

Keywords: computer vision, perception tools, sports, vision system

TIIVISTELMÄ

Roni Kilpeläinen: Konenäkö ja sen käyttäminen urheilussa.
Kandidaatintyö
Tampereen yliopisto
Automaatiotekniikka
Toukokuu 2023

Konenäkö ja havainnointityökalut ovat tärkeä osa kehittyvässä robotiikan maailmassa. Konenäön kehityksellä pyritään tietokoneiden ja robottien visuaalisen informaation ymmärrykseen. Kun tavoite saavutetaan, on tälle lukemattomia käyttötarkoituksia niin tuotannossa, terveydenhuollossa ja viihteessä, kuten urheilussa. Tämä työ käsitteleekin konenäköä, havainnointityökaluja ja erityisesti niiden käyttöä urheilussa.

Työ voidaan jakaa karkeasti kahteen osaan. Työn laajempi osa on kirjallisuuskatsaus konenäöstä ja sen käytöstä urheilussa. Toisessa osassa tehdään pienimuotoinen käytännön työ OpenDR:n pose estimator-työkalulla, jonka avulla analysoidaan tiettyä urheilusuoritusta. Tätä suoritusta verrataan ammattilaisen vastaavaan suoritukseen. Näin esitellään yksi tapa hyödyntää konenäköä urheilussa.

Kirjallisuuskatsauksen ensimmäinen osa käsittelee konenäön perusteita. Osion alussa esitellään lyhyesti, mitä konenäkö on. Tämän lisäksi esitellään konenäköjärjestelmien karkea rakenne. Tämän jälkeen siirrytään konenäön toimintaperiaatteisiin. Ensimmäisenä käydään läpi, kuinka tietokoneet käsittelevät kuvia. Tämä toteutetaan syöttämällä tietokoneelle kuva. Tietokoneelle syötetty kuva pilkotaan osiin, jonka jälkeen jokaiselle osalle annetaan tiettyä väriä vastaava arvo. Kuvankäsittelyn jälkeen käydään läpi tietokoneiden 3D-konenäköä ja muutamia tapoja 3D-näön toteuttamiseen. Työssä esitellään yhteensä kolme eri tapaa. Seuraavaksi työssä käsitellään kuinka järjestelmät toteuttavat jonkun tietyn objektin seurannan videokuvasta. Havaittiin että objektin seurantaan ei ole tiettyä tapaa. On olemassa useita eri algoritmeja, joilla pyritään vahvistamaan objektinseurannassa tiettyjä osa-alueita, kuten havainnointinopeutta. Tällöin muut osa-alueet kuten tarkkuus voivat kuitenkin kärsiä. Objektinseurannassa yhtenä pääkohtana havaittiin taustan erottelu halutusta objektista. Tämä toteutetaan usein vertailemalla peräkkäisiä kuvia ja erottelemalla niistä liikkuvat ja liikkumattomat objektit. Viimeisenä tässä osiossa esiteltiin vielä OpenDR-projektia, jota hyödynnettiin käytännön osiossa.

Kirjallisuuskatsauksen toisessa osiossa esitellään useita eri urheilussa käytettäviä järjestelmiä. Konenäköjärjestelmiä esiteltiin yhteensä neljä. Jokainen järjestelmä keskittyy konenäön eri osa-alueeseen. Esitellyt järjestelmät olivat käytössä jääkiekossa, tenniksessä, sulkapallossa ja kriketissä. Järjestelmien tarkoituksiin kuuluivat esimerkiksi viihdearvon lisäys, tuomaroinnin tarkkuuden parantaminen sekä harjoittelun tehokkuuden kehittäminen. Osa esitellyistä järjestelmistä on jo kaupallisessa käytössä, kun osa taas vielä kehitysvaiheessa. Vaikka järjestelmien tarkoitus onkin viedä urheilua eteenpäin, on niissä myös haasteita. Osa näistä haasteista liittyy järjestelmien teknisiin ominaisuuksiin, kuten tarkkuuteen ja hintaan. Osa taas keskittyy niiden vaikutukseen urheilun viihdearvoon, kuten pelin liiallisiin keskeytyksiin.

Työn viimeisessä osiossa käytettiin konenäköalgoritmia tunnistamaan kuvasta ihmisen niveliä ja muita olennaisia osia asennon arviointiin. Tämän avulla vertailtiin kahta golflyöntiä ja pyrittiin tunnistamaan näistä eroja. Näiden erojen avulla pohdittiin, kuinka lyöntiä täytyisi muuttaa, jotta päästäisiin parempiin lopputuloksiin. Algoritmin avulla lyöntien erot olivat helpommat huomata, ja tultiin siihen tulokseen, että suuremmalla datamäärällä ja paremmilla lajitiedoilla lyönnin kehitykseen voitaisiin antaa paljon enemmän hyödyllistä tietoa.

Avainsanat: konenäkö, havainnointi, urheilu, konenäköjärjestelmä

PREFACE

This topic was chosen due to personal interest in robotics and sports. Computer Vision and Perception Tools are an important part on a functional system both in robotics and in a sports environment. Computer vision as a topic was not familiar to me beforehand, but the practical approach kept me interested in the topic for the whole process. I want to thank my friend Tuomas Vuorenmaa for being an example on the practical part of the thesis and my instructor Roel Pieters for introducing me to the OpenDR toolkit as well as allowing a more practical approach to the thesis.

Tampere, 27 May 2023

Roni Kilpeläinen

CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

2D	Two dimensional
3D	Three dimensional
FTOC:	Algorithm called fast tracking based on object center
AI	Artificial Intelligence, used to describe intelligence displayed by machines
LIDAR	Light detection and ranging, a method for determining ranges
ZED	A 3D vision camera
a	acceleration
F	force
m	mass

1. INTRODUCTION

Robotics is a rapidly growing branch of technology due to its diverse utilization in a variety of different fields. There are many advantages in robotics, some of them are decrease in manual labor and automatization of dangerous tasks. Developing robotics and making them more efficient is a very important task on the journey of making self-reliant machines. Although robots possess many advantages, they also have limitations. These limitations become evident when robots interact with the real world and cannot properly adapt into unpredictable environments. The solution is in developing a system that allows robots to perceive their environment and react based on that.

Perception tools are widely used in robotics to enable machines to perceive and interact with their environment. These tools include sensors as cameras and LIDARs which provide robots with data about their surroundings. By analyzing this data, robots can make informed decisions and perform tasks with greater efficiency and accuracy.[1]

Scientists have been trying to develop ways for machines to understand visual data for over 60 years. During these same times AI also emerged as an academic field of study and since then AI, Deep Learning, and perception tools have often been used and improved simultaneously. Some early topics of computer vision research were edge detection and different object modeling algorithms. Early researchers of computer vision thought that solving the visual input problem would be an easy step along the path of solving more difficult robotics problems. Since then, we now know the problem is much more difficult than that. [2, s.29]

In recent years, there has been a growing interest in applying software and perception tools to the field of sports analysis. By using sensors and cameras to capture data on athlete and object movement, sports analysts and coaches can gain insight into how athletes can improve their performance and optimize their training routines. Additionally wearable sensors can also improve athletes’ safety by alerting when an injury is likely and allowing them to make preventive actions.[1]

Despite the potential benefits of using perception tools in sports analysis, there is a need for further research on their effectiveness and practical applications. This study aims to address this gap by examining the usage of perception tools in analyzing sports, with a focus on their benefits and limitations. Specifically, this study will explore the following questions: What are the basic principles of perception tools? What are the different perception tools used in sports analysis? How are these tools applied in practice? What are the benefits and limitations of using perception tools in sports analysis?

In addition of perception tools in general, this study specifically explores the OpenDR toolkit provided by several different institutions such as Aristotle University of Thessaloniki and Tampere University. As an example, OpenDR toolkit is used to analyze a golf swing and provide information about how the movement could be optimized.

The structure of the study is as follows. Chapter 2 focuses on reviewing the basics of computer vision and perception tools. Chapter 2 also includes a broad review of OpenDR. In chapter 3 it is explored what tools are used in sports and sports analysis and how are these tools applied. In chapter 4, the pose estimator algorithm from the OpenDR toolkit is used in analyzing a sports movement. That analyzing is done by comparing a movement of an amateur to a professional.

2. BASICS OF PERCEPTION TOOLS

In this chapter we will provide an overview of the fundamental principles of computer vision and perception tools. Computer vision is a field that focuses on developing technologies and techniques to enable machines to perceive information from images and video data. Perception tools are a key component of computer vision, as they enable machines to collect the information and data on their surroundings.
We will also examine the advantages and limitations of these fundamentals. While they have the potential to change industries they also come with various challenges. Furthermore, we will discuss what OpenDR toolkit is and how it is addressing some of the challenges in computer vision.
Overall, this chapter will provide a comprehensive introduction to computer vision and perception tools, setting the stage for an exploration of the OpenDR project and its possible uses.

2.1 Computer vision

While humans possess the abilities to sense with vision, smell, hearing, touch and taste, machines do not possess these senses. For machines to be more agile and allowing them to work more independently we have had to and still need to develop ways for them do acquire these senses. Computer vision strives to improve the vision sense on machines to be more human like. As Wiley describes computer vision [3] “Computer Vision is about how we can automate image or video understanding on machines.”

A functional vision system needs several different components. Firstly, you need a radiation source. If an object does not radiate and does not illuminate its surroundings, it is impossible to be observed. Therefore, the wanted object must radiate. To capture the radiation from the source we need some capturing device like a camera. In many environments the captured signal is light. Light can be captured with a simple optical lens.[4] In certain environments, the desired object may not emit light, making necessary to use alternate methods. For instance, in X-ray imaging, the object of interest may be contained withing another object that does emit light.

Other components of a vision system include a sensor, processing unit and actors. Sensors are used to convert the received radiation into a suitable signal for further processing. Processing unit processes and stores the incoming data and Actors react to the visual observation. By connecting these components, we get a basic vision system and its chain of steps as seen in figure 1.

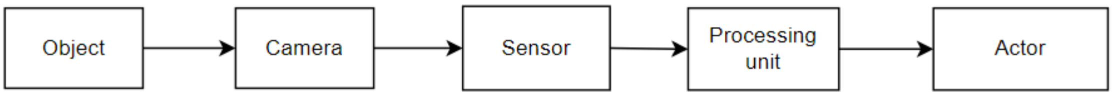


Figure 1. *Chain of steps in a basic vision system. Based on source [4 p.3].*

2.2 Image understanding

While we as humans find it easy to understand images and visual information. It's a much more complex task for computers. Unlike us, computers can't see images, they must break them down into many smaller pieces and assign a value to each one to understand the differences between them. To a computer, an image is just an array of numbers that corresponds to colors (Figure 2). If we want to acquire accurate and reliable results from robots, the arrays must be much bigger and more complex. For example, to be equivalent to the human eye there would have to be around 127 million elements in the array. In addition, the more we add colors, the more complex and heavier the computation process gets. [4]

67	67	66	68	66	67	64	65	65	63	63	69	61	64	63	66	61	60
69	68	63	68	65	62	65	61	50	26	32	65	61	67	64	65	66	63
72	71	70	87	67	60	28	21	17	18	13	15	20	59	61	65	66	64
75	73	76	78	67	26	20	19	16	18	16	13	18	21	50	61	69	70
74	75	78	74	39	31	31	30	46	37	69	66	64	43	18	63	69	60
73	75	77	64	41	20	18	22	63	92	99	88	78	73	39	40	59	65
74	75	71	42	19	12	14	28	79	102	107	96	87	79	57	29	68	66
75	75	66	43	12	11	16	62	87	84	84	108	83	84	59	39	70	66
76	74	49	42	37	10	34	78	90	99	68	94	97	51	40	69	72	65
76	63	40	57	123	88	60	83	95	88	80	71	67	69	32	67	73	73
78	50	32	33	90	121	66	86	100	116	87	85	80	74	71	56	58	48
80	40	33	16	63	107	57	86	103	113	113	104	94	86	77	48	47	45
88	41	35	10	15	94	67	96	98	91	86	105	81	77	71	35	45	47
87	51	35	15	15	17	51	92	104	101	72	74	87	100	27	31	44	46
86	42	47	11	13	16	71	76	89	95	116	91	67	87	12	25	43	51
96	67	20	12	17	17	86	89	90	101	96	89	62	13	11	19	40	51
99	88	19	15	15	18	32	107	99	86	95	92	26	13	13	16	49	52
99	77	16	14	14	16	35	115	111	109	91	79	17	16	13	46	48	51

Figure 2. *Different versions of an image. An array of numbers (left) which are the values of grey scales in the low- resolution image of a face (top right). Based on source [4].*

While some tasks must be only a fraction as accurate as human eye, robots that utilize computer vision and interact with humans must be reliable. So, robots must be able to process video footage in real time. Also, the robot some information that robots perceive must be colorized in order to separate different entities. That adds another layer of complexity in our vision systems and makes it no easy task.

2.3 3D Imaging

Non-contact depth sensing is an important part on a vision system. Functional vision systems must be able to detect and react to objects in their environment to work properly without constant monitoring. There are many different depth sensing technologies and here are presented a few of them. Different technologies have different advantages, and the method of choice should be based on them.

Triangulation is the most widely used technique for optical shape measurements [4]. As in its name triangulation utilizes triangles, geometry, and parameters such as distance between the cameras to figure out the distance to the wanted object. There are many different triangulation techniques, but they all have the same basic principle of trigonometry.

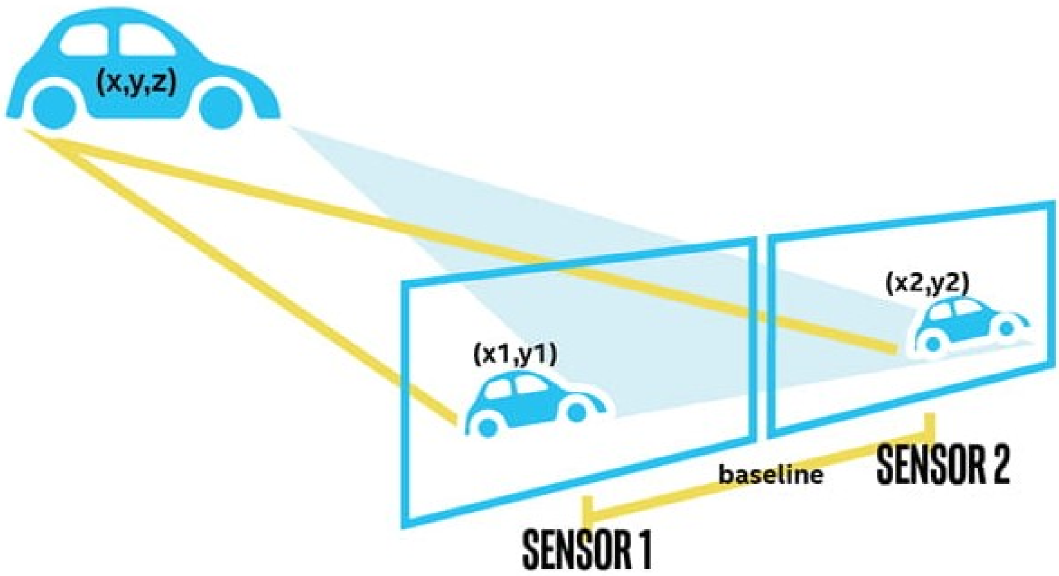


Figure 3. *Example image of triangulation with 2 sensors. [5]*

Time-of-flight method uses speed of light to determine the depth to an object. A light signal is sent to an object and a sensor catches the reflected light. The distance can then be calculated with

$$z = \frac{c\tau}{2} \tag{1}$$

where z is the distance to the object, c is the speed of light and τ is the time-of-flight of the light signal.[4] Time-of-flight is utilized in common sensors such as LIDAR.

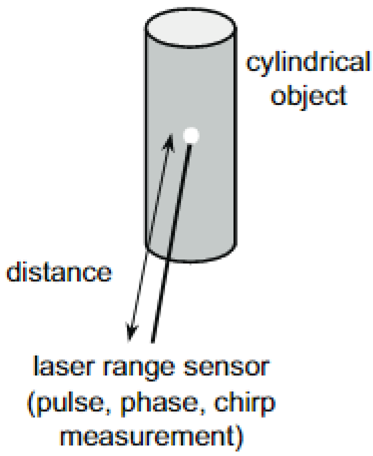


Figure 4. *Principle of a time-of-flight sensor.[4]*

Another way of measuring distance to an object is using structured light. In this technology patterned, usually infrared light is projected into an object. Sensor then observes the pattern and because the projected pattern is known, it provides the information needed to acquire depth.

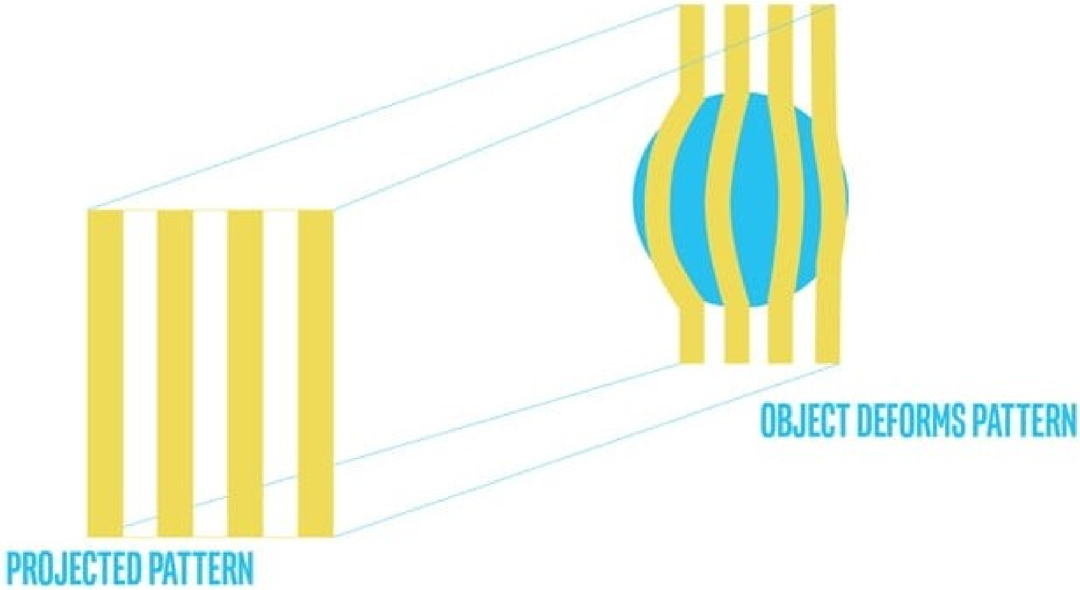


Figure 5. *Illustrative image on usage of structured light. [5].*

2.4 Object tracking

Tracking an object of interest has many different practical applications due to its importance on robots doing task autonomously. Fundamentally it refers to using sensor measurement like a camera to determine its location and path.[6] Like all computer vision applications, also object tracking has many different problems and difficulties. These difficulties include, illumination changes and shadows, difference in weather, background updates, and moving objects that do not interest us, like branches or clouds.

There are a wide variety of techniques and algorithms that try to solve the object tracking tasks in different ways, but it they always try to extract the wanted object from a background. Tracking usually starts by building a background image. There are many different solutions to this, but many utilize consecutive frames and segment out the parts that do not move.

After that you need to detect the object of interest via some detection system. One popular detection system is called Hawkeye, it is used in many sports such as cricket. Hawkeye uses various cameras and calculates the wanted objects 3D position from x

and y co-ordinates from the consecutive frames. With the information acquired it can track the object and make predictions about its future path. [7]

Generally, object tracking does not have a straightforward implementation due to different conditions in environment. The basic principles of image processing and 3D imaging still apply and for object tracking system to work it must utilize those concepts.

2.5 OpenDR

OpenDR is an EU 2020 project that was released under the Horizon 2020 program. OpenDR is built upon Python, and it aims on developing a modular, open, and non-proprietary toolkit for core robotic functionalities. OpenDR focuses on perception and strives for active robot perception in a human-centric environment.[8]

Although Deep Learning has benefited AI tremendously it has some barriers in a practical robotics scenario. These barriers include heavy computational load and Deep Learnings applications on mostly static environments.[8] For Deep Learning and robotics to be applied in sports, the systems must be mobile, and computation must be quick. While this project may not be directly related to sports, its research has potential applications in the field, such as 3D multi object tracking [9], or human action recognition [10].

3. USAGE OF PERCEPTION TOOLS IN SPORTS

Computer vision has the potential to revolutionize sports in numerous ways. By utilizing image processing techniques, computer vision can analyze sporting events. Computer vision can analyze everything from movements of players to the trajectory of an object. This can lead to improved training methods better understanding of game dynamics and adding more entertainment value. In this chapter I will explore how computer vision and perception tools are currently used in sports.

3.1 Ice Hockey

Finnish hockey league Liiga is currently using technology named Wisehockey. Wisehockey is a system that tracks player and puck statistics such as:

- Speed, time on ice and distance traveled on ice
- Puck control
- Faceoffs
- Shot and goal detection
- Shift tracking
- Penalties and powerplays
- Players' +/- statistics
- Momentum

These statistics are calculated real time and provided in their web portal. [11]

In chapter 2.1 five components of functional vision system were introduced. This system utilizes sensors within the puck and on the players to provide the wanted radiation. The locators installed in the arena then capture this signal and transmit it to the arena server. The data is then forwarded to a cloud environment, where algorithms transform it into readable and useful statistics. These statistics are then made available on a web portal to this systems actors. The web portal is accessible to a variety of people including coaches, viewers, and sports bettors, who based on the data can take appropriate actions.[10]

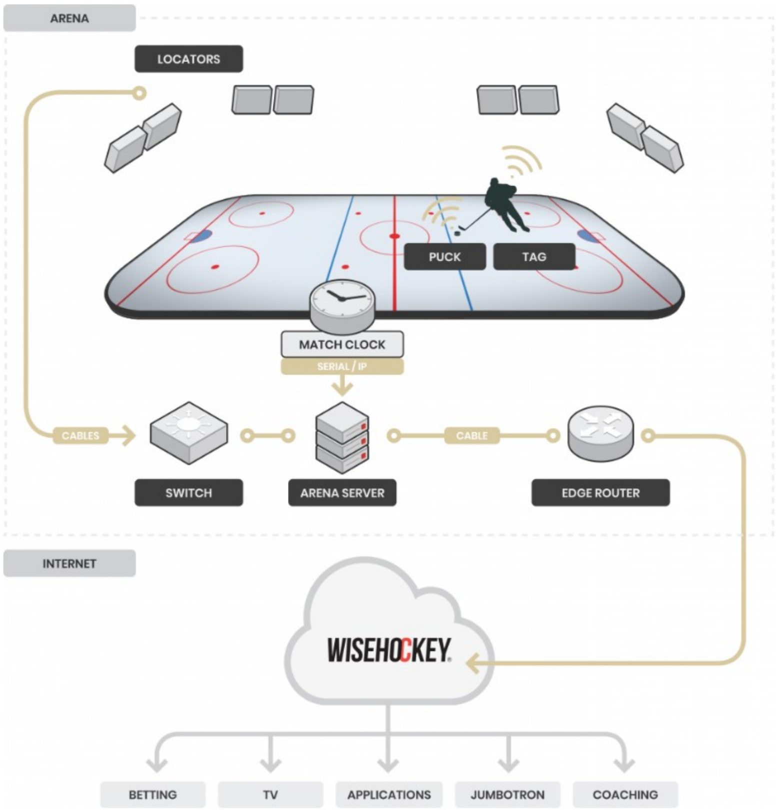


Figure 7. *Wisehockey systems' hardware infrastructure. [10]*

Liiga and a few other hockey leagues use this system for statistics that demonstrates one approach to player tracking in hockey. This system effectively addresses many of the challenges associated with player tracking via the wearable sensors. According to a study [13], one major challenge in player identification is the difficulty of telling apart players from the same team, who have almost identical uniforms and similar physical size. Additionally, when players are wearing protective gear, it can be even harder to differentiate the players form each other. To address this problem, V. Kanav et al. [13] propose a solution that involves identifying players based on their jersey numbers, which are the most prominent features on player jerseys. System like Wisehockey demonstrates that there is some demand for player tracking and should courage more research on how to develop a similar system that only utilizes cameras and visual footage.

3.2 Tennis

Tennis players and coaches can employ perception tools to analyze and enhance their performance. For instance, cameras and sensors can monitor how players and ball move around the court, collecting information on things like shot velocity and movement. Based on the data coaches and players can prepare and make tactical decisions and improve their performance for future matches.

Depending on the wanted analysis one of the first steps in analyzing sport like tennis is accurate player silhouette extraction. While hockey had its own challenges in player detection, tennis has also faced its own set of issues. According to Vito et al. [14] study some of the challenges in tennis silhouette extraction come from wearable sensors being unallowed in official matches, color of the court, and players' uniform and skin color.

Vito et al. [14] propose a way to achieve accurate silhouette extraction with their algorithm called GIVEBACK. Their system consists of 4 cameras, 2 for each side of the court.

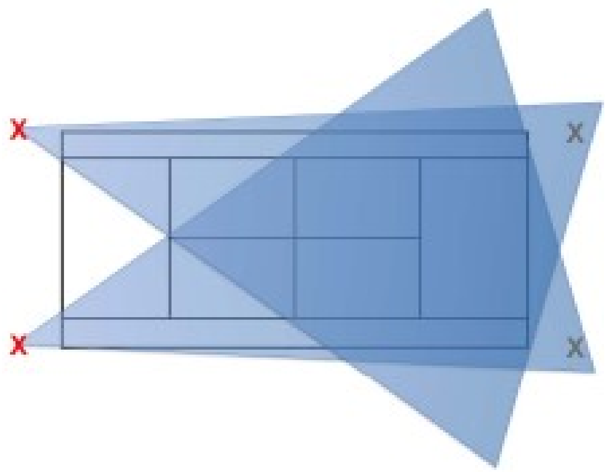


Figure 8. *Position of acquisition hardware on the tennis court.[14]*

Even though the camera views overlap the system is not reliant on multiple views and can function as a single camera system.
In their study Vito et al. divide their algorithm into 3 main building blocks, initialization, processing, and update. As seen in Listing 1. the first step is executed only once, and after that each frame is analyzed and modified to achieve wanted results.

```
Background Initialization
for each frame
    Variance process
    One step frame differencing
    if (background is learned)
        Foreground extraction
    Fine tuning process
    Background Update
    Energy Process
```

Listing 1. *Silhouette extracting algorithm as pseudocode. Based on source [14].*

Figure 9. *Example of synchronized acquisition. [14]*

Even though only player silhouettes (Figure 10.). are not very helpful for everyday tennis players, or even professionals, it is an important step on creating more advanced systems for tennis enthusiasts. These systems could utilize anything from racket angle on impact to some movement patterns that aren’t easily visible to human eye. In addition to making their own training more efficient and effective, systems like this can be used to find weak spots from their opponents, allowing to strategize a better game plan.

It is also important that these tools to become easily accessible. Vito et al. have addressed this, and, in their study, they mention how their method is efficient and operates form raw video footage which encourages its implementation on portable devices like smart cameras [14].

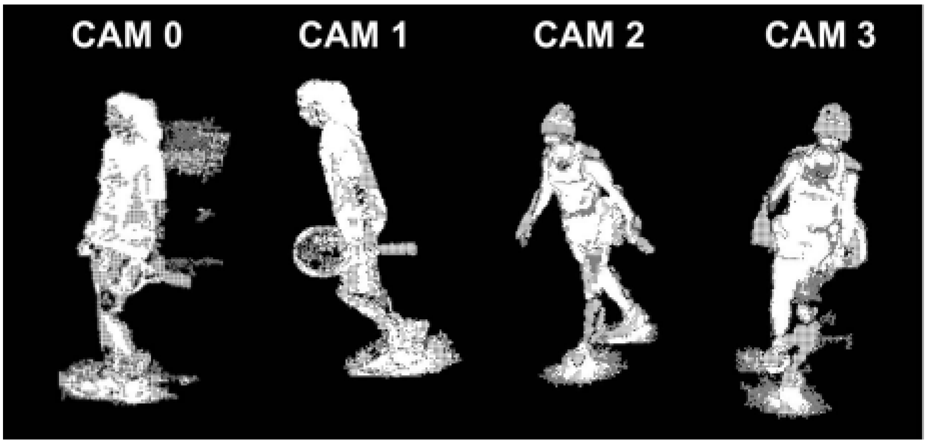


Figure 10. *Example of player silhouettes extracted from four synchronized views. CAM0 and CAM1 refer to Player 1, while CAM2 and CAM3 to Player 2.[14]*

3.3 Badminton

As tracking players offers a variety of possibilities on improving sport experiences, tracking objects also provides many areas of improvement. In their study Chen et al. [15] provide an algorithm for badminton robot to track the shuttlecock in real time for it to make necessary adjustments to hit it back. When this kind of technology is developed properly, it allows players to practice different game scenarios in a controlled environment. This can help players to target their weaknesses and improve without the need for a partner or a coach.

Chen et al. ‘s paper focuses on 2D real-time shuttlecock tracking with their FTOC algorithm. Shuttlecock tracking differs a little from sports like tennis. The fact that shuttlecocks may alter their appearance due to deformation or abrupt motion, and in addition of the high pace of the shuttlecock, tracking them can be challenging.

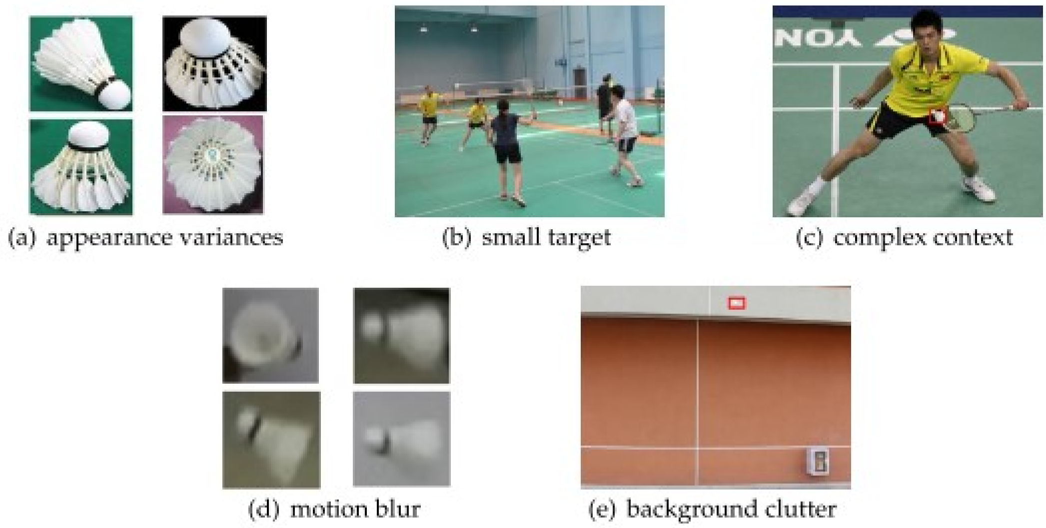


Figure 11. *Challenging scenarios for shuttlecock tracking. [15]*

In their system Chen et al. [15] use a ZED camera for their data acquisition. The images are then sent to their proposed FTOC combined with AdaBoost algorithm. As an input their tracking algorithm takes video stream of flying shuttlecock or a stream directly from ZED. As an output it produces the shuttlecock coordinates in 2 dimensions. (x, y). These coordinate pairs are then used to predict the 3D (x, y, z) spatial coordinates. For this Chen et al. chose the LSM method due to its robustness and efficiency. The robot is then directed to act accordingly based on the transmitted coordinates.

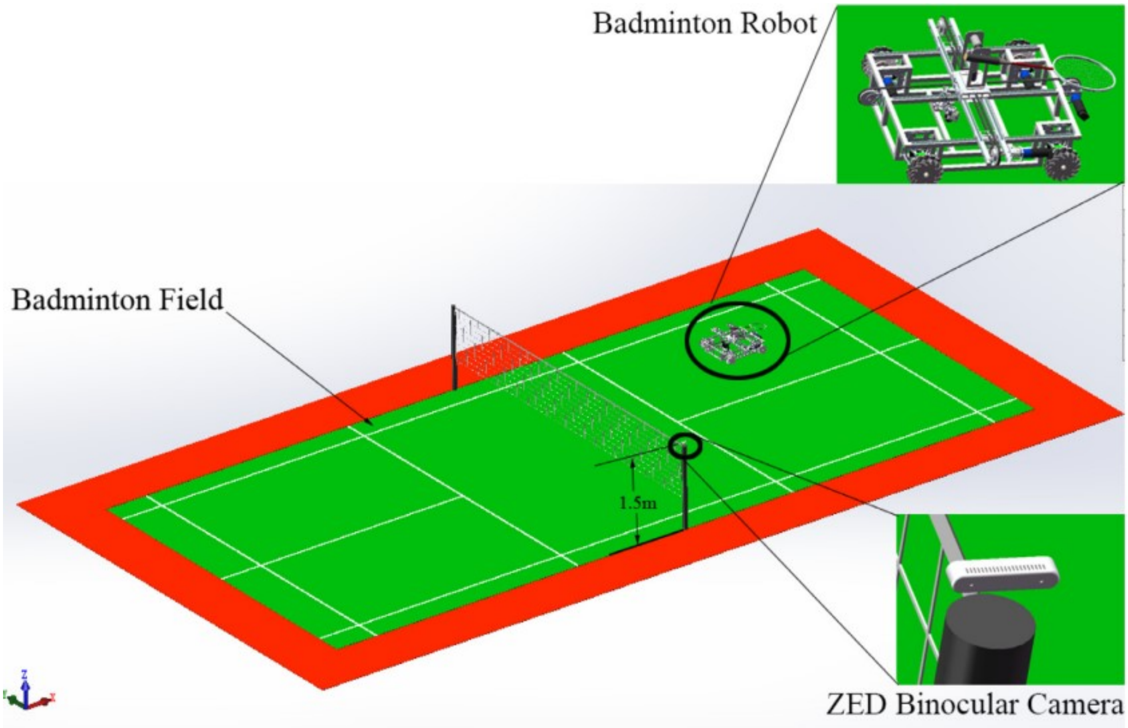


Figure 12. *3D figure of the proposed system, including ZED camera and Badminton Robot designed by GDUT_BT robot team. Based on source [15].*

To be useful, robots and their computation must be fast, efficient, and reliable. especially in a sport like badminton since shuttlecocks can gain speeds up to $493\frac{km}{h}$ [9]. Chen et al. say that in their implementation there are only 0.5 seconds for the system to track and estimate the final locations. The lack of time produces accuracy issues but, in the paper, Chen et al. claim that they will try to solve the problem in their future work. This type of system is currently on the development phase, and it is hoped that in the future, both sports enthusiasts and professionals will be able to benefit from them.

3.4 Hawkeye system

The Hawkeye system was developed by a British technology company in 2001. It was first used in cricket to determine the legitimacy of leg before wicket decisions made by umpires. The system uses multiple high-sped cameras placed around the stadium to capture the ball's movement from different angles. The cameras then feed the data to a computer which processes the information and creates a 3D image of the ball's trajectory. The system can predict if the ball would have hit the stumps if it did not hit the player's leg, helping the umpires make accurate decisions.[7]

The system has had a great impact on sports. It has reduced human error in officiating and ensured that the right decision is made. The Hawkeye has also made sports more fun for fans, as they can see replays of critical moments and make their judgements-. The use of technology has reduced the chance of making biased decisions.

Systems such as Hawkeye also have downsides. They can slow down the pace of the game and interrupt its flow. Additionally, the cost of installing and maintaining the system can be high, making it too expensive for smaller leagues and tournaments. While the system has downsides, the benefits outweigh its limitations. It is a great example of how technology can improve sports and make them more enjoyable for everyone.

4. POSE ESTIMATION IN OPENDR

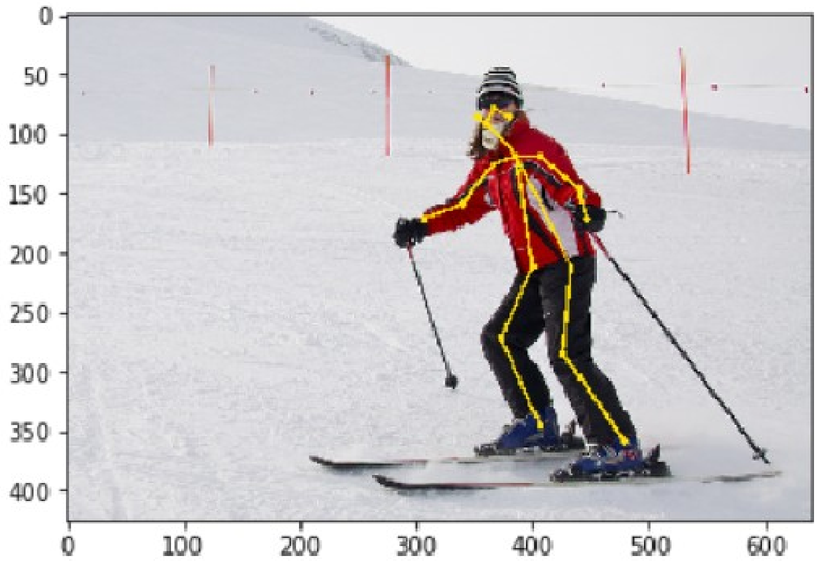
In this chapter I will introduce OpenDR's pose estimator tool and how it can be utilized in analyzing a golf swing. This analyzing will be done to my friend Tuomas and professional with a similar body structure Rory McIlroy.

4.1 Overview on pose estimator

From the pose estimation toolkit, we will be using the Lightweight OpenPose tool. It is a Python based algorithm that essentially tracks the joints and other keypoints from the human body. After the keypoints have been detected it draws lines between the keypoints representing the human skeleton and other features as seen in figure 13. The algorithm also gives coordinates of the keypoints that can be seen on the same figure. In this example we will use the tool on static images, but the tool can also be used on live webcam footage.[16]

Lightweight OpenPose is a deep learning algorithm. In short, a deep learning algorithm like this consists of many processing layers that each have their own purpose of detecting different elements. These layers then work together to recognize objects within the data [17]. Although our example only uses single person recognition from a static image, the Lightweight OpenPose tool is capable of multi-person pose estimation. This is usually done in two ways called top-down, or bottom-up. Top-down method first runs a person detector and after that applies the pose estimation on each person. The disadvantage of this method is the decrease in inference speed when number of people inside the image increases. The bottom-up method that our tool uses, first finds the keypoints and then groups these keypoints by human instances. With this method the inference speed stays almost the same regardless of the number of people in the given image. [16]

A deep learning algorithm must be trained to provide accurate results. In essence training a deep learning model is feeding large amount of data to the model, for it to detect different elements from the data. This training can be done with custom datasets, to acquire more accurate results or use a pre-trained model.[18] Due to time and hardware restraints, this example was done using the pre-trained model of Lightweight OpenPose. Lightweight OpenPose pre-training is done using the COCO [19] dataset, which contains various images such as objects, people, and animals [16].




```
nose: [360, 92]
neck: [386, 119]
r_sho: [366, 126]
r_elb: [340, 159]
r_wri: [307, 172]
l_sho: [406, 119]
l_elb: [439, 146]
l_wri: [446, 172]
r_hip: [400, 212]
r_knee: [373, 272]
r_ank: [406, 345]
l_hip: [433, 212]
l_knee: [426, 285]
l_ank: [479, 358]
r_eye: [353, 86]
l_eye: [366, 79]
r_ear: [-1, -1]
l_ear: [380, 86]
```

Figure 13. *Example image of the keypoint mapping and their coordinates. [20]*

4.2 Swing comparison

As mentioned in this test we will compare the swing of a professional to an amateur's and point out some differences that can be seen from the pictures. The swing is split into four phases and are in pairs on our figure 14. On the top row is the amateur Tuomas and in the bottom row is the professional Rory. As seen from the first pair, or the starting position, you can already see that Tuomas' shoulders are level where Roryss are not. From this angle the second pair of pictures looks fairly similar. Lot of the keypoints are clumped together and behind the head so with this tool the differences can be hard to point out. One difference that can be noted with the tool, is that Rory's shoulder goes past the nose line and therefore his upper body rotates more than Tuomas'. From the third pair of the pictures, we can note that Tuomas' left heel separates from the ground and therefore his left knee gets more bent than Rory's. From the fourth picture, in addition to the previous pictures knee angle, you can see an early rotation of the shoulders and hips in comparison to Rory's swing, while Rory still faces forward Tuomas has already begun his rotation to face the field.

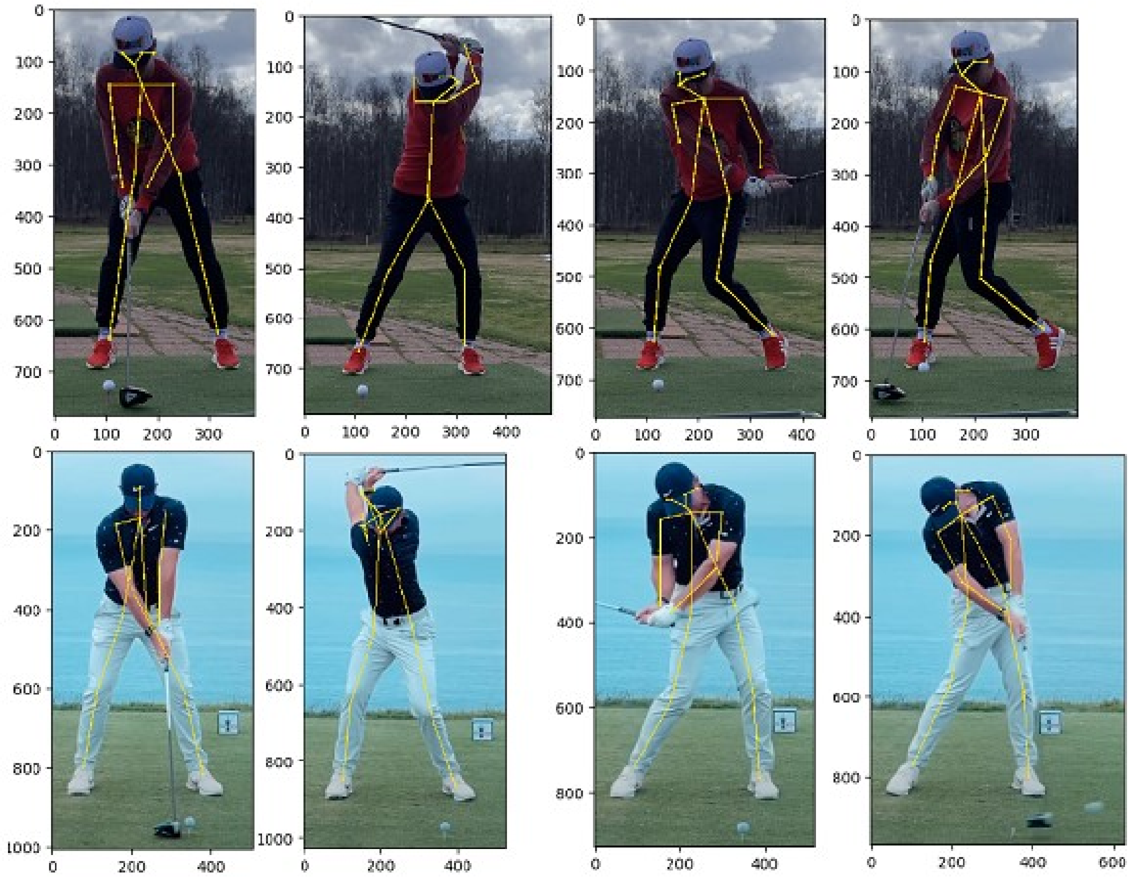


Figure 14. *Comparison pictures from the swings. Footage of the second player based on [21].*

4.3 Conclusions

From this short analysis we can already pinpoint many things where Tuomas can focus on his training. Although golf is such a precise sport and everyone's ideal swing is different, comparing a sports movement into a professional provides useful data on what our amateur does differently compared to an elite golfer. If this same picture would be given to a real golf coach, they could probably say more, and more accurate observations and on what to focus first on the training.

This is just a glimpse of what could be done with such a tool in the sport of golf. By adding more data such as various angles and a wider range of comparison swings, along with the analysis of a professional coach, this tool has the potential to make the efficiency of amateur golf coaching much better and more available. The coach would no longer need to be physically present during training session and still be able to provide valuable assistance to the golfer. There are still challenges on using this tool like accuracy of detecting all the joints. For example, the second picture from the top row, the algorithm points both hip coordinates on the same x coordinate, and therefore it skews the knee angles. If a golf coaching program would be made and it used joint angles as data, it could provide wrong information for its user. Overall, this example showed how a simple tool that recognizes joints could be used in the field of sports and in what direction it could be developed.

4.4 Implementation

This work was done using a Lenovo Yoga 730 computer. The device had an Intel® Core™ i5-8250U 1.60GHz CPU and no dedicated GPU. Because of that the OpenDR's Lightweight PoseEstimation tool was run in a CPU only mode. The OS in use was a Linux Ubuntu 20.04 LTS. The capturing device for Tuomas' footage was an Apple's iPhone XS and the capturing device from Rory's swing is unknown since the footage was acquired from a source [21]. After the footage was captured, the code was then run in an Jupyter Notebook environment similar as seen in Tutorial [20]. With slight changes to the code such as changing the device into CPU (Pose Estimation Tutorial section 2) or changing the image path (Pose Estimation Tutorial section 7), the wanted images could then be run for inference on our system.

5. SUMMARY

This thesis researched the fundamentals of computer vision and perception tools. We addressed that the five basic components of a vision system are, radiation source, camera, sensor, processing unit, and actor. Each of these components have an important part on a system. Then we examined on how a computer processes and understands images. We found out that it segments the image into many chunks and gives each a color value. After that we moved onto a 3D imaging system and presented three different ways on how to implement 3D vision. The three ways were triangulation that utilizes two cameras and a known distance between them, time-of-flight method is based on sending a light signal onto an object and measuring the time of how long was needed for it to come back to our sensor, and the third way used a known pattern of light that was projected onto an object and then pictured with another camera. After 3D imaging we researched object tracking and found out that the common factor in object tracking is separating the wanted object from the background and other moving objects. There wasn't a regularized way on doing this and different algorithms had different strengths and weaknesses. In this part we also did a little overview on OpenDR project.

On the second part we researched different vision systems that were used in sports. In total we presented a total of 4 different vision. Each of the systems demonstrated a different way on how computer vision was utilized. Our hockey system uses tags and locators to capture and provide a variety of data for stakeholders such as viewers or coaches. Secondly, we covered a system that was used to segment tennis player's silhouette. A system like that is an important step for future vision systems in the field of sports. Next, we covered a project in which a badminton robot was built. The challenges of this project were mainly on shuttlecock tracking and when a system like this is perfected, the benefits for sports training are great. Lastly a ball tracking system that is used in sports like cricket and tennis was introduced. Originally this system was designed to help officiating cricket, but since then the system has been adopted into many different sports.

The third part was a practical example of how pose estimating algorithm can be used in golf coaching. The algorithm extracts joints from a picture and then draws bone-like guidelines which can then be used to approximate different angles in the body. These lines and joint positions are then compared to a professional and the differences can be used as a guideline on what to focus on training.

Overall, this thesis provides a general picture of how computer vision and perception tools work, are used, and can be used in the field of sports. Chess is a good example of the potential how computers can revolutionize sports analysis, and accurate perception is a crucial step on the path of acquiring said potential.

SOURCES

[1] C. Premebida et al. Intelligent Robotic Perception Systems, 2018. URL: https://www.intechopen.com/chapters/62978/ (accessed May 23, 2023)

[2] R. Szeliski, Computer Vision Algorithms and Applications, Springer, 2011 https://link-springer-com.libproxy.tuni.fi/book/10.1007/978-1-84882-935-0

[3] K. Dawson-Howe, A practical Introduction to Computer Vision with OpenCV, John Wiley & Sons, Incorporated, 2014, https://ebookcentral.proquest.com/lib/tampere/detail.action?docID=7104138.

[4] B. Jahne, Computer Vision and Applications, San Diego: Academic Press, 2000. https://search-ebscohost-com.libproxy.tuni.fi/login.aspx?direct=true&AuthType=cookie,ip,uid&db=e000xww&AN=198901&site=ehost-live&scope=site

[5] 'Beginner's guide to depth', 2019. https://www.intelrealsense.com/beginners-guide-to-depth/ (accessed May 24, 2023)

[6] S. Challa et al., Fundamentals of Object Tracking, Cambridge University Press, 2011. https://ebookcentral.proquest.com/lib/tampere/detail.action?docID=667571.

[7] 'Video Processor Systems for Tracking in Ball Games', https://worldwide.espacenet.com/publicationDetails/originalDocument?CC=WO&NR=0141884A1&KC=A1&FT=D&ND=&date=20010614&DB=EPODOC&locale= (accessed May 25, 2023)

[8] 'OpenDR Factsheet' https://opendr.eu/wp-content/uploads/2020/07/OpenDR_Factsheet.pdf (accessed May 24, 2023).

[9] R. Mohan, A. Valada, 3D Multi-Object Tracking Using Graph Neural Networks with Cross-Edge Modality Attention, IEEE Robotics and Automation, 2022. https://opendr.eu/wp-content/uploads/2022/10/3D-Multi-Object-Tracking-Using-Graph-Neural-Networks-with-Cross-Edge-Modality-Attention.pdf

[10] N. Heidari, A. Iosifidis, Temporal Attention-Augmented Graph Convolutional Network for Efficient Skeleton-Based Human Action Recognition, 25th International Conference on Pattern Recognition (ICPR), 2020. https://opendr.eu/wp-content/uploads/2021/04/On-the-spatial-attention-in-Spatio-Temporal-Graph-Convolutional-Networks-for-skeleton-based-human-action-recognition.pdf

[11] 'Wisehockey' https://bitwise.fi/referenssit/wisehockey/ (accessed May 26, 2023)

[12] 'Real-Time Sport Analytics Platform' https://wisehockey.com/wp-content/uploads/2020/02/wh_overview_en.pdf (accessed May 26, 2023)

[13] V. Kanav, P. Walters, M. Fani, DA Clausi, JS Zelek, Player Tracking and Identification in Ice Hockey, Expert systems with applications Vol. 213, 2015. https://doi.org/10.1016/j.eswa.2022.119250

[14] R. Vito et al. Tennis Player Segmentation for Semantic Behavior Analysis, 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), IEEE, 2015. https://ieeexplore-ieee-org.libproxy.tuni.fi/search/searchresult.jsp?searchWithin=%22Publication%20Number%22:7406213&searchWithin=%22Document%20Title%22:Tennis%20Player%20Segmentation%20for%20Semantic%20Behavior%20Analysis

[15] Chen, Wei et al., Using FTOC to Track Shuttlecock for the Badminton Robot, Neurocomputing (Amsterdam), 2019. https://doi.org/10.1016/j.neucom.2019.01.023.

[16] D. Osokin, Real-time 2D Multi-person Pose Estimation on CPU: Lightweight OpenPose, arXiv preprint arXiv:1811.12004, 2018.

[17] 'What is deep learning'. https://www.ibm.com/topics/deep-learning (accessed May 27, 2023)

[18] 'What is deep learning'. https://www.mathworks.com/discovery/deep-learning.html (accessed May 27, 2023)

[19] T.-Y. Lin et al., Microsoft COCO: common objects in context, In ECCV, 2018.

[20] 'Pose estimation tutorial'. https://github.com/opendr-eu/opendr/blob/master/projects/python/perception/pose_estimation/lightweight_open_pose/demos/inference_tutorial.ipynb (accessed May 10, 2023)

[21] 'Footage of Rory McIlroy swing'. https://golfdiscountmall.com/watch-rory-mcilroy-close-up-swings-slow-motion-us-open-2021/ (accessed May 10, 2023)