FACULTY OF ENGINEERING
AND ARCHITECTURE

Crash during the bunch sprint in the fourth stage of Tour de Suisse 2010.
Photo credits - Graham Watson.

Jelle De Bock - doctoral dissertation

Data-Driven Performance and Safety Analysis in Cycling Races

Jelle De Bock

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Information Engineering Technology

Supervisors
Prof. Steven Verstockt, PhD - Maarten Slembrouck, PhD
Department of Electronics and Information Systems
Faculty of Engineering and Architecture, Ghent University

October 2023

2023

GHENT
UNIVERSITY

FACULTY OF ENGINEERING
AND ARCHITECTURE

Data-Driven Performance and Safety Analysis in Cycling Races

Jelle De Bock

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Information Engineering Technology

Supervisors
Prof. Steven Verstockt, PhD - Maarten Slembrouck, PhD
Department of Electronics and Information Systems
Faculty of Engineering and Architecture, Ghent University

October 2023

GHENT
UNIVERSITY

ISBN 978-94-6355-763-4
NUR 984
Wettelijk depot: D/2023/10.500/95

Members of the Examination Board

Chair

Prof. Hennie De Schepper, PhD, Ghent University

Other members entitled to vote

Kevin Baker, PhD, RouteYou
Bram Janssens, PhD, Ghent University
Prof. Steven Latré, PhD, Universiteit Antwerpen
Prof. Erik Mannens, PhD, Ghent University
Joachim Taelman, PhD, Ghent University

Supervisors

Prof. Steven Verstockt, PhD, Ghent University
Maarten Slembrouck, PhD, Ghent University

Acknowledgements

My passion for sports made the choice for a sports-related data science obvious. After the interesting master's thesis topic which was situated in the context of the Continuous Athlete Monitoring (CONAMO) project, professor Verstockt convinced me to join his research group. I was quickly convinced that this was the right choice, as he had the ambitious goal to start with sports-specific innovation projects. Together with professor Verstockt, we saw the research group transitioning from two passionate researchers working on one or two sports-related projects to a fully developed team working on more than five extensive projects with research partners all over the globe. I can say that I am proud to have my share in this research group! I will be forever thankful to Steven, both as the intelligent and motivating professor Verstockt and as the compassionate, understanding and kind person he also is. I think I can honestly say that I would never have delivered this result without him.

I also want to thank dr. Maarten Slembrouck for taking the role as supervisor, together with prof. Steven Verstockt. Especially, in the finishing stages of the dissertation you served as my big example in how to bundle, streamline and document the research trajectory of the past four years. Dr. Slembrouck's attention to detail helped me a lot to fine-tune my dissertation, both grammatically and content-wise. I also learned a lot from Maarten when working with him on the WCN and STRADA project. I especially enjoyed our time together soldering together wires when we were bonking (for those who do not understand, this is cycling jargon to say that we were hungry).

They often say that research is a team sport, and I could not agree more! During my PhD research, I had the luck to have shared the office with a lot of clever, but also very nice and kind people. I especially like to thank my ex-colleague dr. Florian Vandecasteele. When he was in the finishing stages of his PhD about fire safety, I joined the team. I now recognise a lot of the struggles and stress moments Florian had during the finishing stages. You were a great mentor and helped me find my way in the research world. Next, I can proudly say that I have worked together with the most awesome colleagues on the projects I worked on. The discussions about research and life in general with Robbe Decorte, prof. Dieter De Witte, Kenzo Milleville, dr. Joachim Taelman and Jelle Van Haeverbeke (we have two Jelles in our team). I can also not forget my former colleagues; thank you dr. Krishna Kumar Thirukokaranam Chandrasekar, dr. Dilawar Ali, Lore Simons and Alec Van den broeck. Alec also was one of my first master's students, but he transitioned to an intern and later into a nice and inspiring colleague.

Last, but by no means least I'd like to thank my family and loved ones! You were the ones that kept believing in me, even when I wasn't believing in myself anymore. You kept me motivated and helped me see the bigger picture. It wasn't always easy to combine a PhD with sports, but you helped me to put everything into perspective and brought me back up when I was down. Together with my friends, you helped me to forget the bad times and made the sun shine when it was dark! The next drink is definitely on me!

Moerbeke-Waas, September 2023
Jelle De Bock

Table of Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
BLE	Bluetooth Low Energy
CNN	Convolutional Neural Network
CSV	Comma Separated Values
DAIQUIRI	Data & Artificial intelligence for Quantified Reporting in sports
DEM	Digital Elevation Model
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GPX	GPS eXchange
HTML	Hypertext Markup Language
IoT	Internet of Things
JSON	JavaScript Object Notation
ML	Machine Learning
NLP	Natural Language Processing
NER	Named Entity Recognition
OCR	Optical Character Recognition
OSM	OpenStreetMap
PC	Personal Computer
POS	Part-of-speech

PTZ	Pan Tilt Zoom
R-CNN	Region based Convolutional Neural Network
RFID	Radio-Frequency Identification
RSSI	Received Signal Strength Indicator
SBC	Single Board Computer
TARS	Task-aware representation of sentences
UCI	Union Cycliste Internationale
USB	Universal Serial Bus
UWB	Ultra Wide Band
WCN	Wireless Cycling Network
XML	Extensible Markup Language

Samenvatting
– Summary in Dutch –

Sensoren, video en social media data zijn alomtegenwoordig in het hedendaagse wielrennen. Renners dragen tal van wearables; motorfietsen, helikopters en drones zorgen voor de videocaptatie en fans en teams raporteren wedstrijd- en wielergerelateerde randinformatie op social media. Uit deze grote hoeveelheid (big) data nuttige inzichten genereren is echter geen sinecure. Er is duidelijk nood aan tools, methodologiëen en algoritmes om beelden te analyseren, sensordata te verwerken en ongestructureerde naar gestructureerde informatie om te zetten.

Als we naar de geproduceerde data in het wielrennen kijken dan kan deze ruwweg in drie klassen onderverdeeld worden: video/audio, tekstuele en geospatiale en/of tijdsgeanoteerde data. Door deze types data te gaan analyseren en combineren kunnen we nieuwe, innovatieve services ontwikkelen die atleten, fans, coaches en organisatoren een betere beleving van de wielersport bezorgen. Deze thesis onderzoekt drie zulks een services: gepersonaliseerde storytelling voor fans, datagedreven veiligheidsanalyses voor organisatoren en multimodale prestatieanalyses voor de sporter en diens entourage.

Video is het eerste type data dat bestudeerd en geanalyseerd wordt. Een videobeeld is immers een waardevolle, vaak onontgonnen schat aan informatie. Daarbovenop is het professionele wielrennen één van de meest geteleviseerde sporten in Vlaanderen. Van alle belangrijke wedstrijden is er wel een videostroom voor handen. Om de waardevolle informatie in de opgenomen videobeelden te verkrijgen dienen er dikwijls meerdere analysestappen doorlopen te worden. In deze thesis worden met name pose-estimators, renners- en teamherkenning en geverkeersinfrastructuur-gerelateerde objectdetectie nader bekeken. Voor de pose-estimator werden verschillende state-of-the-art frameworks onderzocht en werd uiteindelijk Alphapose gekozen als framework voor de implementatie voor onze use cases. Voor de renners- en teamherkenning en het

verkeersinfrastructuur detectie model werd in beide gevallen dan weer voor een hertraind YOLOv5 model gekozen.

Professioneel wielrennen is voorts ook een onderwerp waarover, vooral in aanloop van de belangrijke wedstrijden zoals de Ronde van Frankrijk of de Ronde van Vlaanderen, veel inkt vloeit in de pers. In tegenstelling tot de periode waar enkel de papieren krant elke dag nieuwe informatie verspreidde, is de hedendaagse stroom van nieuwsfeiten haast continu. Deze informatiestromen vormen voor data-analyse een enorme, maar vaak ontongonnen schat aan kennis. De schrijvende pers wordt namelijk steeds meer aangevuld door sociale media-kanalen zoals Twitter, Facebook, Instagram of TikTok. Net zoals bij videoanalyse, is bij tekstuele data eveneens een conversie-proces benodigd om geschreven tekst om te zetten in gestructureerde data die bruikbaar is voor verdere analyse. Natuurlijke taalverwerking (Natural Language Processing, NLP) is de verzamelnaam van methodes en algoritmes om betekenis te halen uit geschreven en/of gesproken tekst. In deze thesis wordt bijvoorbeeld het sociale mediaplatform Twitter gebruikt om incidentgerelateerde tweets op te sporen en zo automatisch een incidentdatabank aan te leggen. Een tekstclassifier bepaalt of een Tweet al dan niet crash-gerelateerd is en haalt hier automatisch de betrokken renners en eventuele oorzaken uit. Op basis van het tijdstip van de Tweet of door eventuele kilometeraanduidingen in de de Tweet tekst of uit de beelden die aan Tweet kunnen gelinkt worden, kan soms ook al de locatie van het incident en de daaraan gerelateerde metadata geëxtraheerd worden (vb. afdaling, platteland/stadscentrum en moment in de race). Verder kan combinatie van de data ook parcoursbouwers en officials helpen bij het plannen van toekomstige wedstrijden en deze dus op die manier veiliger maken.

Een andere waardevolle informatiebron vormt de tijdruimtelijke data. Dit is data die voorzien is van een coördinaat en/of van een tijdsindicatie. Op zich hebben een geografische locatie of een tijdstip weinig waarde voor analyses, maar wanneer deze gelinkt worden aan bijvoorbeeld sensorwaarden dan krijgt de sensordata als het ware een extra dimensie. Een interessant storytelling voorbeeld in deze thesis is de segmentanalyse die uitgevoerd wordt op GPS-logdata. Door het verschil in de tijd die verscheidene renners nodig hebben om een segment te voltooien kan verteld worden welke segmenten de meest uitdagende zijn en dus het interessantst zijn om te verhalen en te bekijken. Dit is zoals vernoemd interessant voor storytelling, maar kan bijvoorbeeld ook voor performance doeleinden gebruikt worden. In de Ronde Van Vlaanderen kan bijvoorbeeld mede hiermee de helling gekozen worden waarop het best aangevallen kan worden of kan het een fan helpen beslissen op welke helling te gaan staan omdat hij daar het meest actie zal mogen verwachten. Het zou ook de broadcasters van een cyclocrosswedstrijd kunnen helpen om de camera plaatsing te optimaliseren.

De hier voorgaande analyses focusten telkens op één enkel type informatie. Als echter alle beschikbare types samen gebruikt worden, kunnen nog diepgaandere analyses voor een bepaalde use case uitgevoerd worden. We definiëren dan ook multimodale analyse als "een analyse die door meerdere types data samen te brengen nieuwere, rijke inzichten biedt die niet uit 1 individueel type data te verkrijgen is". De verschillende wielerdisciplines zoals wegwielrennen, veldrijden en baanwielrennen lenen zich uitermate tot dergelijke analyses. In dit werk focussen we voonramelijk op 3 services: safety screening voor een veiligere koers, performance analysis binnen het pistewielrennen en storytelling in cyclocrossverslaggeving.

De parcoursanalyse en incidentrapportering die voor de internationale wielerunie (Union Cycliste Internationale, UCI) ontwikkeld werd is een mooi voorbeeld van hoe multimodale data kan bijdragen tot een veiligere wieleromgeving. Videoanalyse op GPS-geannoteeerde videoframes wordt hierbij gebruikt om het wegdek te segmenteren en de gevaarlijke elementen te annoteren op het parcours van een wedstrijd. Verder wordt de geospatiale databank OpenstreetMap gebruikt om het GPS-bestand van het parcours te analyseren. Zoals reeds eerder vermeld wordt tekstuele data van Twitter gebruikt om eventuele incidenten in wedstrijden te rapporteren in een MySQL-databank.

Voor de storytelling demonstrator werd het veldrijden als discipline gekozen. Het cyclocrossen is immers in Vlaanderen erg populair waardoor meestal zowel de vrouwen- als mannenwedstrijd integraal te volgen is op televisie. Geautomatiseerde rijlijnanalyse is een mooi voorbeeld van zo een videogedreven analyse. Door middel van een vaste camera gericht op een bepaalde strook wordt met een rennerdetectie-module binnen een zone op het beeld naar de rijlijn gezocht die het meest efficiënt is op deze strook. Op basis van rennerdetectie/-tracking detecteren we van elke renner diens rijlijn, alsook de daarmee corresponderende timing. Door deze van alle renners te vergelijken kan men de meest efficiente rijlijn bepalen voor een bepaald segment. Een andere interessante cyclocross use case is de verhouding van lopen/fietsen in een segment. Hiervoor werd een YOLOv5 model gebouwd dat in staat is lopende en fietsende renners te herkennen. Wanneer dit gecombineerd wordt met de segmenttijden, kunnen interessante inzichten geboden worden. Zo kan het bijvoorbeeld blijken dat iets vroeger van de fiets gaan voordeliger is dan heel ver te proberen fietsen.

In onze laatste use case focussen we op het baanwielrennen. In samenwerking met Cycling Vlaanderen werd de piste in het Centrum Eddy Merckx centrum uitgerust met een set-up om real-time sensor en video data van de renners op te nemen. Samen met de context van het type training en/of oefening dat uitgevoerd werd en de rondetijden van een timingssysteem kunnen diepgaande (video)analyses van de trainingen aan de coaches gepresenteerd worden. Dit werd gebruikt in de Madison discipline, meer bepaald in de wissel tussen twee renners. Aan de hand van de centraaal verzamelde data is het door een op cadans- en power-gebaseerd algoritme mogelijk de wissels te detecteren (i.e., tijdstip en waar ergens op de piste). Met een camerasetup bovenop het datacaptatie framework konnen er vervolgens ook beelden van de gedetecteerde wissels uit de opnames geknipt worden voor verdergaande analyses zoals de kwaliteit van de armbeweging en de lichaamshouding tijdens het wisselen (door computervisie modellen en/of coaches).

Als conclusie kan gesteld worden dat data alomtegenwoordig en van onschatbare waarde is binnen het professionele wielrennen. Vaak is de data voorhanden maar is de mate waarin ze geanalyseerd wordt nog eerder beperkt. Deze thesis onderzoekt de meerwaarde die de huidige, state-of-the-art machine learning technologieën kan bieden binnen het wielrennen. Hiervoor gebruiken we tekstuele, video en geospatiale databronnen om storytelling-, veiligheids- en prestatieanalyses uit te voeren. De gepresenteerde bouwstenen in deze thesis zijn de fundamenten voor volledig geautomatiseerde en gepersonaliseerde samenvattingen of volautomatische veiligheidsanalyses van wielerwedstrijden.

Summary

Sensors, video and social media data are omnipresent in modern bicycle racing. The riders wear numerous wearables; motorcycles, helicopters and drones take care of video recording and fans and teams report race and race-related side information on social media. The generation of useful insights from the vast amounts of (big) data is not a straightforward job. It is obvious that there is a need of tools, methodologies and algorithms to analyse the video images, crunch the sensor data and convert unstructured into structured information.

If we look further into the produced data in cycling, we can roughly divide it in three classes: video/audio, textual and geospatially and/or temporally annotated data. When these types of data are analysed and combined, we can develop new, innovative services and provide athletes, fans, coaches and organisers a better experience of the sport. This thesis proposes three of those services: personalised storytelling for fans, data-driven safety analysis for organisers and multimodal performance analyses for athletes and their entourage.

Video is the first type of data that is studied and analysed. A video frame is often a rather unexploited source of information. Furthermore, professional cycling is one of the most televised sports in Flanders and the video streams of all important races are usually available. The big challenge for this type of data convert a video frame into actionable insights. This conversion often consists of multiple analysis steps and usually also a fair amount of computing power. In this thesis, we will further look into pose-estimators, rider- and team-recognition models and traffic-infrastructure-related object detectors. For the pose-estimator several state-of-the-art frameworks were investigated and benchmarked. With this knowledge, we selected the Alphapose framework for the implementation for our use cases. For both the rider- and team-recognition and the traffic infrastructure detection model a retrained YOLOv5 model was chosen.

Furthermore, in the press much ink flows about professional cycling. This is especially the case in the run-up to important races such as the Tour de France or

the Tour of Flanders. The rise of social media platforms such as Facebook, Twitter and Instagram allowed the masses to broadcast additional news or just their opinions at any time of day. Unlike the time when only written newspapers were daily spreading new information, the current stream of news facts is almost continuous. For data analysis, these streams of information represent an enormous, but often unexplored, wealth of information. Nowadays, the written press is complemented by social media channels. Similar to video analysis, textual data also requires a conversion process to turn written text into structured data that is suited for further analysis. Natural Language Processing (NLP) is the collective term for all methods and algorithms to extract meaning from written text. For example, in this thesis the social media platform Twitter is used to detect incident-related tweets and automatically create an incident database. A text classifier determines if a Tweet is crash-related and automatically extracts the riders involved and the possible causes of the incident. Also, based on the publication time of the Tweet and possible kilometre indications in the Tweet text or in images linked tot the Tweet, the location and its metadata (e.g., descent, rural/urban and moment in the race) could be extracted. These insights can ultimately help race organisers to plan out their courses and make them safer.

Another valuable source of information is geospatially and/or temporally annotated data. This is data that is linked to a geographical location and/or that has a time indication. In itself, a geographical location or a time might have little value for analysis, but when these are linked to sensor values, the sensor data gains an extra dimension. An interesting story-telling example in this thesis is the segment analysis performed on GPS log data. Through the difference in the time it takes several riders to complete a segment, interesting analyses can be performed to highlight which segments are the most challenging and thus the most interesting to watch. As mentioned, this is interesting for story-telling but can also be used for performance purpose. In the Ronde Van Vlaanderen, for instance, this can be used to choose the best hill to attack on or for a fan to choose the segment where it is expected to see the most action. It could also help in cyclocross to assist the broadcasters in optimising the camera positioning.

The previous analyses each focused on a single source of information. However, if all the types are used together even broader analyses can be performed for a particular use case. We define multimodal analysis as "an analysis that combines multiple types of data to provide new, more enriched insights that cannot be provided by a single source of data". The various cycling disciplines such as road cycling, cyclocross and track cycling are very suitable for these types of analyses.

In this work, we mainly focus on three services: safety screening for a safer racing, performance analysis in track cycling and storytelling in cyclocross broadcasting.

The course analysis and incident reporting developed for the International Cycling Union (Union Cycliste Internationale, UCI) is a great example of how multimodal data can contribute towards a safer cycling environment. In this use case, video analysis on GPS annotated video frames is used to segment the road surface and annotate the dangerous elements on the course of a race. Furthermore, the geospatial database OpenstreetMaps is used to analyse the GPS file of the course. As previously mentioned, textual data from Twitter is also used to report incidents in races and store them in a formatted manner in a MySQL database.

For the storytelling demontrator we focused on cyclocross. The discipline is very popular in Flanders and the big races are usually broadcasted from start to finish for both the men and the women. Automated ride line analysis is a nice example of such a video-driven analysis. With a fixed camera focused on a particular sector of the course, a rider detection module searches within a predefined zone on the image for the riding line that is most efficient for the selected sector. Based on rider detection and tracking we can detect a rider and his/her riding lines and link it with the corresponding timings. When all riders are compared, the most efficient ride line can be obtained for a certain segment. Another interesting cyclocross-specific use case is the riding and running distribution in a segment. For this purpose, a YOLOv5 model was built that is capable to distinguish between riders on the bike and riders that are running with their bike. When this information is combined with segment timings valuable insights are provided. It might for instance be beneficial to dismount from the bike a bit sooner rather than trying to cycle as far as possible in a sandpit.

In our last use case, we focus on track cycling. In collaboration with Cycling Vlaanderen we have equipped the Centrum Eddy Merckx cycling track with a setup to collect real-time sensor data and video of all riders on the track. Together with the context of the type of training and/or exercise that was performed and the lap times of a timing system, in-depth analyses of the workouts can be presented in real-time to the coaches. Our setup has already been successfully used in the Madison discipline, more specifically to detect the handsling changes between two riders. The centrally captured data was used to develop a cadence- and power-based algorithm to detect the handsling changes (i.e., a timestamp and location on the track). With an additional camera setup on top of the sensordata capturing platform it was possible to extract clips from the detected handsling changes from the recordings. These clips can be used by coaches or further anal-

ysed with computer vision to provide comprehensive analyses of the arm movement and posture whilst preforming the handsling.

In conclusion, data is ubiquitous and invaluable within professional cycling. Often the data is available, but it is not being used to its full potential. This thesis explored the possibilities of current, state-of-the-art machine learning technologies applied to professional cycling. We use textual, video and geospatial data as the main data sources to perform storytelling, safety and performance analyses. The presented building blocks in this thesis lay the foundations for fully automated and personalised summaries or fully automated safety analyses of cycling races.

1
Introduction

"Luctor et emergo."

–Wapen van Zeeland (1948)

This chapter situates the conducted research work, summarises the main contributions and outlines the structure of this dissertation. It also provides an overview of the publications that were authored related to this dissertation.

1.1 Context

Over the past years, an exponential growth of cycling sensor-generated data has been observed. In contrast, the impact of these new technologies on storytelling, safety analysis, coaching/training and injury prevention is still limited. For storytelling, for example, most of the existing applications just add an additional layer of raw sensor data on top of the live broadcast (this is, for example, what the likes of Dimension Data, NTT and Velon have been doing over the past years [1, 2]).

Another noticeable trend is the increase in the number of live broadcasted races the last couple of years. Gone are the days that only big one-day classics and stage races were broadcasted on television. Additionally, quite some races are even entirely filmed from the start to the finish. This is not only very entertaining for the spectators, but also provides a lot of potential for computer vision techniques to learn from the video footage and provide automatic analyses on the recorded footage. But, similarly to the generated sensor data, the available video data is mostly left unexploited.

The main challenge and the current shortcoming of most of the innovations is that only a single source of data is used at a time which makes it often difficult to interpret the real context of the situation. This is for instance the case with the infographics that Velon posts on their social media. It is interesting to see the average power of a rider the last 10 kilometers of a race, but an extra source of data (e.g., video or the course profile) could provide extra meaning to these statistics.

The real goal of sports-related storytelling and analysis should be thorough understanding of what the raw sensor values and images mean. This can be achieved with a unique mix of feature engineering and machine learning and by bringing together multiple sources of data (e.g., automatically combining a video stream combined with the course elevation profile).

To exemplify this set goal we can take the example of the time-trial performances of the Belgian and Italian powerhouses Wout Van Aert and Filippo Ganna. It is easy to understand that the rider with the fastest average speed along the time-trial course will arrive first. But it is not necessarily the case that the "fastest" rider did put in the biggest effort. Filippo Ganna does probably have the aerodynamic advantage over Wout Van Aert, but Wout's cyclocross background helps him having the edge on overall cornering speed and the accompanied short, vigorous accelerations after those corners. These are the differences that the classical analyses don't reveal. In other words, we know who won, but we would probably also like to know why this person won, and this is exactly where multimodal data analysis can provide insights that spawn beyond a single-source data analysis.

As a final remark, it should be made clear that this dissertation is the end result of very broad research on data science applied to cycling. This will become evident by the magnitude of techniques and use cases presented in this thesis. This has the implication that the performed research is very result- and implementation-driven, and thus, is focusing on the applicability of state-of-the-art data-science and computer vision techniques on cycling. All the aspects presented in this PhD thesis could be PhD-topics on its own, but the deliberate choice on the broad/wide applicability of data science techniques on cycling was made from the ground up.

The overall goal of this dissertation can be formulated as follows: **"To what extent can multimodal data mining improve safety, storytelling and performance in professional cycling?"**.

1.2 Challenges

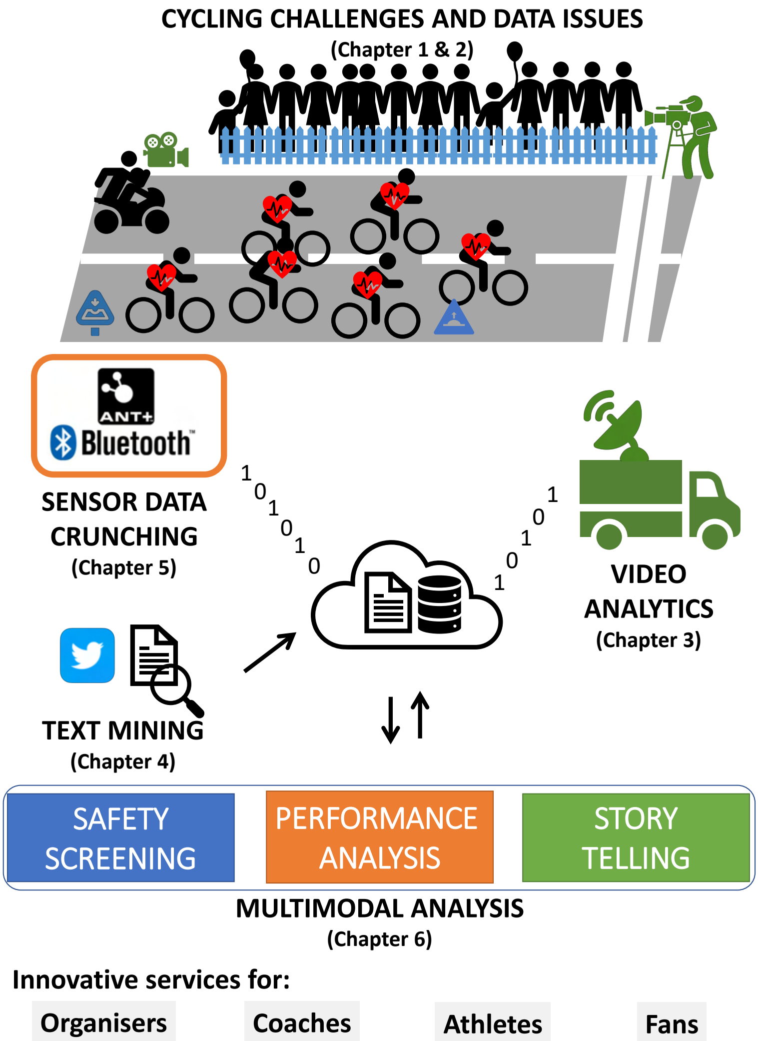


Figure 1.1: Schematic position of the different chapters in this dissertation

Sensors, video and social media are nowadays an integral part of modern cycling (see Figure 1.1). Cyclist are equipped with numerous wearable devices, motorcycles and helicopters capture the video footage of a race and fans and teams report about the races on social media.

These vast amounts of big data can provide valuable insights and services to stakeholders, but this is not a straightforward task. There is a need for tools, algorithms and methodologies to analyse video, capture and process sensor data and convert unstructured text to structured data.

With all the data available and as mentioned in the research goal of this publication, the aim is to use multimodal data analysis to make a technological contribution towards professional cycling. This dissertation will mainly focus on three domains within professional bicycle racing: ‘safety’, ‘storytelling’ and ‘performance’. In the following subsections, we will introduce the domains and explain why they are relevant and how information technology can make a valuable contribution within each of the domains.

1.2.1 Safety

Road cycling is an endurance sport in which athletes ride their bicycles on courses that mainly contain public roads. The number of vehicles using these public roads is slowly increasing through the years [3] and, as a side effect, the number of traffic incidents is also increasing. Safety measures such as traffic calming infrastructure (e.g., roundabouts or speed bumps) [4–6] on public roads are often introduced to cope with this issue and have a significant impact on the number of road fatalities [7]. Unfortunately, although these imposed measures might be highly beneficial for regular road usage, it is usually the opposite when they are used for road cycling races. In the past, severe accidents happened due to sudden narrowing of roads, roundabouts, traffic islands or speed bumps. A great, but unfortunate example of this statement is the crash of the Deceuninck Quickstep rider Yves Lampaert during Milano-Torino in 2020 (Figure 1.2). Within the last ten kilometers, the Belgian classics rider crashed at full speed into a traffic island, causing a collarbone fracture and a couple of weeks out of competition.

The biggest challenge of safety in cycling is the coordination and assistance of what and how to signal to riders, what sectors to avoid in a course and how to learn from historical incidents. With a better understanding of the contributing factors of big and severe crashes and by relating them to road circumstances and their moment in the race (e.g., start, climb or finish) it is possible for our stakeholders (i.e., organisers and jury) to take appropriate measures to reduce the chance of severe accidents.

Figure 1.2: Crash of Yves Lampaert on a traffic island that was cushioned, not signalled. (source: incycle Youtube channel)

1.2.2 Storytelling

"If we want to keep bicycle race broadcasts competitive amongst the likes of Netflix and co., we need to come up with formats that make bike races more popular to watch.". This rather direct quote of Tomas Van Den Spiegel, CEO of Flanders Classics, in the lead-up to the 2022 edition of Flanders' biggest one-day classic "De Ronde van Vlaanderen" perfectly captures and represents the storytelling challenges we face in professional cycling.

A similar, more scientific publication by Van Reeth et al. [8] shows similar conclusions. In Figure 1.3 the decreasing interest in the Tour de France can be observed for most countries. Additionally, they found out that the average age of TV spectators is rather high [9]. In France, more than half of the TV audience watching the races is sixty-plus, while only 14% is younger than 35. This is probably not because cycling is not the most attractive sport to watch, but rather because of the way that it is presented to the viewers. In the modern era, where everything has to be fast and shiny, watching a 4 hour long broadcast of a race might not sound the most appealing for a younger audience. Well-thought TV productions and new technologies might be needed to whet the appetite of the youngsters. Recent advancements such as in-car video, audio recordings and real-time sensor information are already a huge step in the right direction [10]. With all this new data available and in order to avoid data-overload to the audience, the real challenge for data scientists and computer vision experts lies in the combination of sensor data, metadata and video data to combine short appealing stories that are personalised and on a per user-level. For instance, even if a spectator might not be interested to watch a very long, general broadcast, they might for instance be interested in a five minute summary which focuses on his/her favourite rider, team or nationality and automatically shows the most engaging race situations given these preferences.

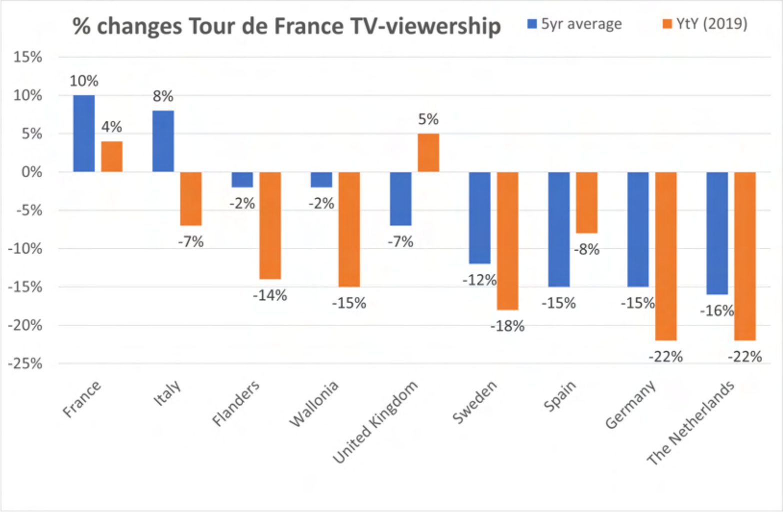


Figure 1.3: Stats from professor Daam Van Reeth's Twitter channel (KU Leuven) illustrating the decreasing interest in the Tour de France in countries where cycling is popular and mediatised.

Last but not least there is the fact that a lot of the fans are also amateur/recreational cyclists themselves. The last couple of years the cycling industry is booming. The pandemic got even more people on bicycles and this is resulting in a huge boost of sales in both bicycles and bicycle equipment [11]. Based on the company's annual reports, this increase in sales is also illustrated in Figure 1.5. The fact that viewers are also interested in the technical aspects of cycling (e.g., which gears are they using, what bike brand, which wheels, etc.) could also be used towards the improvement of storytelling and viewer engagement. As already mentioned, all riders collect and share heaps of performance-related data. The role of information technology could also lie in the fact that it can assist in the full immersion of the sporty viewer. The current state of the art equipment such as smart indoor cycling trainers, powerful home computers and all the data collected might perfectly allow viewers to not only watch, but also ride along with the exciting parts of races.

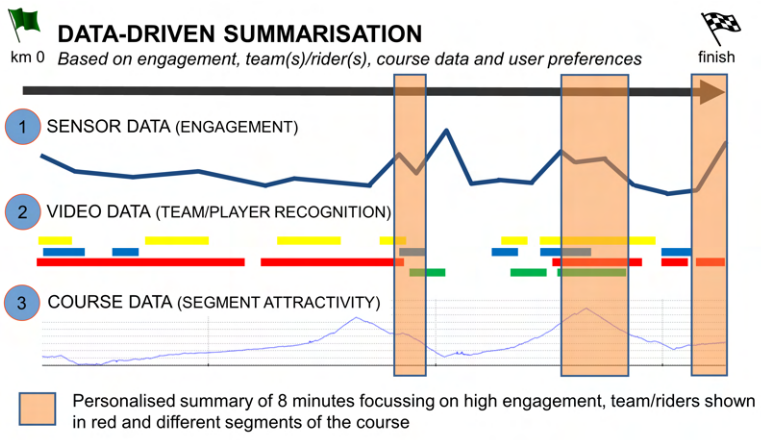
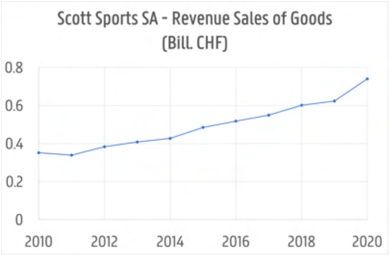
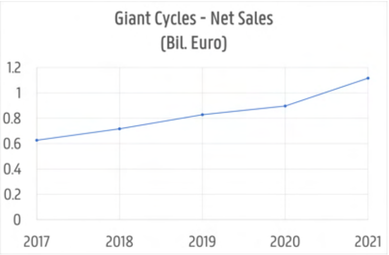


Figure 1.4: Data-driven summarisation as a solution to keep a broader audience entertained and offer a more personalised viewing experience to the fan.



(a)



(b)

Figure 1.5: The clear trend in two of the main cycling companies (Scott Sports (a), Giant Cycles (b)) is the increase of sales. (source: online available annual financial reports)

1.2.3 Performance

Data doesn't win matches or races, but it can definitely help during the process of building and training a successful team to win important races. Although we might have just invented this quote, its relevance might have already been proven multiple times in the past. Sports- and data-minded people undoubtedly heard about 'Moneyball' [12], a quite popular film directed by Bennett Miller, which is inspired on the same-titled book by Michael Lewis. In this film, although it might be slightly over-romanticised, the makers tried to shake up the conventional way baseball scouts selected their team roster by introducing data and algorithms as a helping factor to shape the optimal team for a given (limited) budget. To a certain extent, data plays and has played a huge role in understanding performance and gain statistical insight in what might give the edge over the opponent.

Additionally, a study by Laird and Waters [13] states that experienced coaches can only recall 59 percent of the critical events that occurred in football matches. This percentage even drops when coaches are less experienced. It is to this problem that data-driven and/or video-based match/race analytics might come into play. Algorithms and models do not replace coaches or trainers, but the ideal solution will help them to filter out and summarise the key takeaways from a race/match.

If we focus on data-driven performance in cycling, we can see two major potential data sources for performance analysis and enhancement: sensor and video data. The sensor data can even be further divided in two major categories. A first category are sensors that are worn and/or move together with the cyclist (e.g., power meter, glucose meter or heart rate strap). A second category are external sensors such as the MyLaps timing loops. This produces a wealth of information, and as Tim Cusick, product lead of WKO quite correctly phrases it: "If we look at how big data has led the way, we see a three-step evolution in data analytics: descriptive to predictive to prescriptive"[1]. Although many app developers, researchers, product designers and teams already figured out a way to develop valuable descriptive statistics (i.e., through graphs and dashboards), evolving towards predictive and prescriptive is still mostly uncharted territory and there is a lot of room left for improvement within these two phases.

An extra challenge within this sub-challenge is that usually multiple disciplines have to interact in order to gain thorough understanding from the data [14]. For instance, if we would want to train a model that gives training advice to cyclists, a data scientist would first need to gain an understanding of how and why certain energy systems need to be trained and how an athlete reacts on training sessions in order to find its way in the provided data sets. The best models and insights can only be achieved with the shared knowledge and collaboration of domain experts in both sports and data science. Luckily and in response to this challenge, several conferences emerged over the last years that bring together researchers and industry from different background (i.e., coaches, physiologists and data scientists).

The three presented challenges are the main topics that were researched within this PhD dissertation. The three challenges will be approached from a data-driven angle. In the remainder of this dissertation, the different types of data will be identified and studied with the presented challenges in mind. For this dissertation we will mainly focus on cyclocross, track cycling and road cycling and will research how the disciplines can be made safer, how a rider can gain

[1] https://www.netapp.com/blog/cycling-data-endurance-sports-analytics/

competitive advantage and how the races can be optimally brought to fans and stakeholders.

1.3 Outline

In Section 1.2, the problems and challenges within professional cycling and how they could be improved with the use of data and algorithms were discussed. Figure 1.6 provides a schematic overview of the research contributions within each of the focus areas in this dissertation. In the columns the focus areas within this PhD are provided. The rows represent the analysis techniques, depending on the analysed data type. The logos at the intersections between rows and columns are the data sources and companies that provided a dataset for this research. In the last row, for the multimodal data analysis, the projects in which all the different analyses come together are presented. The DAIQUIRI project was a storytelling project in collaboration with the media broadcasters, focused on making data-driven stories during cyclocross races. “Museum in de Living” was a cycling-related cultural heritage project in collaboration with the KOERS museum. WCN and STRADA are projects in collaboration with the Flemish cycling federation. COURSE is the road safety and crash reporting project we did in collaboration with the Union Cycliste Internationale (UCI). BAS-X was a project in collaboration with RouteYou, where we developed route screening building blocks for recreational purposes. Some of these were repurposed for the COURSE project.

Figure 1.6: Schematic overview of the contributions within each of the focus areas in this dissertation.

Within this section, we give an overview of the remainder of this dissertation and explain how the different chapters are linked together. In Chapter 2, we start with a global overview of the different available data sources in cycling and will also focus on what the current limitations and hurdles are when using the data for analysis and machine learning. In the following three chapters (i.e., Chapters 3, 4, 5), we will further elaborate on the possible analyses that can be performed on each of the data sub-types introduced in the data chapter. Next, in Chapter 6, we will bring multiple types of analysis together in use cases that have been tackled throughout the PhD trajectory. In Chapter 7 we will draw the major conclusions and point out future work that can be performed within data-driven professional cycling analysis.

This dissertation came to existence by a number of publications that were realised in the context of this PhD. The selected publications provide an integral and consistent overview of the work performed. Furthermore, this dissertation serves as a composed overview of all individual publications, brought together within this dissertation. The different research contributions are detailed in Section 1.4 and the complete list of publications that resulted from this work is presented in Section 1.5.

1.4 Research contributions

During the research phase of this dissertation a lot of contributions with regard to the described challenges in professional bicycle races have been made. The main contributions can be summarised as follows:

- Implementation of a Twitter methodology using Natural Language Processing (NLP) to convert unstructured Tweets into structured database input to document incidents within professional men and women road cycling.

- Implementation of a YOLOv5 road-traffic infrastructure detector, a U-NET with a ResNet34 encoder road quality estimation model and an edge-computed SqueezeNet course crow density estimation methodology for visual course safety analysis.

- Presentation of the cycling course safety and the reported incidents in a web-based React application.

- Data-driven sector engagement analysis in cyclocross and cycling races based on segment times deducted from riders' GPS data.

- Real-time ANT+ sensor data collection in professional track cycling with a WASP capturing network and the subsequent presentation and analysis of the captured data with a C# back-end and a React front-end.

- Proximity-based camera selection in cyclocross races with Quarq Qollector real-time location data of the riders.

1.5 Publications

The research results obtained during this PhD research have been published in scientific journals and presented at a series of international conferences. The following list provides an overview of the publications during my PhD research.

1.5.1 Publications in international journals (listed in the Web of Science [2])

1. **De Bock, J.**, Verstockt, S. (2021). Video-based analysis and reporting of riding behavior in cyclocross segments. SENSORS, 21(22). https://doi.org/10.3390/s21227619

2. **De Bock, J.**, Verstockt, S. (2022). Road cycling safety scoring based on geospatial analysis, computer vision and machine learning. In Multimedia Tools and Applications. Springer Science and Business Media LLC. https://doi.org/10.1007/s11042-022-13552-1

1.5.2 Publications in other international journals

1. **De Bock, J.**, Verstockt, S. (2021). SmarterROUTES-a data-driven context-aware solution for personalized dynamic routing and navigation. ACM TRANSACTIONS ON SPATIAL ALGORITHMS AND SYSTEMS, 7(1). https://doi.org/10.1145/3402125

2. **De Bock, J.**, Verstockt, S. (2019). SmarterRoutes: data-driven road complexity estimation for level-of-detail adaptation of navigation ser-

[2]The publications listed are recognised as ‘A1 publications’, according to the following definition used by Ghent University: Articles included in one of the Web of Science databases ’Science Citation Index’, ’Social Science Citation Index’ or ’Arts and Humanities Citation Index’. Limited to the publications document type article, review, letter, note, proceedings paper.

vices. ADVANCES IN CARTOGRAPHY AND GISCIENCE OF THE ICA, 2. https://doi.org/10.5194/ica-adv-2-2-2019

1.5.3 Publications in international journals currently under review

1. **De Bock, J.**, Verstockt, S. (2022). COURSE - an artificial intelligence based toolbox to create safer road cycling courses. Manuscript submitted for publication in the ACM Multimedia Tools and Applications journal.

1.5.4 Publications in international conferences (listed in the Science Citation Index [3])

1. Braeckevelt, J., **De Bock, J.**, Schuermans, J., Verstockt, S., Witvrouw, E., Dierckx, J. (2019). The need for data-driven bike fitting: data study of subjective expert fitting. In J. Vilas-Boas, P. Pezarat-Correia, J. Cabri (Eds.), IC-SPORTS: PROCEEDINGS OF THE 7TH INTERNATIONAL CONFERENCE ON SPORT SCIENCES RESEARCH AND TECHNOLOGY SUPPORT (pp. 181–189). https://doi.org/10.5220/0008344701810189

1.5.5 Publications in other international conferences

1. **De Bock, J.**, Vandewiele, G., Verstockt, S., Ongenae, F., De Turck, F. (2018). Towards an automated workout compliance model. In World Congress of Performance Analysis of Sport XII ISPAS (pp. 53–62). Opatija, Kroatia: International Society of Performance Analysis of Sport (ISPAS).

2. Verstockt, S., Mannens, E., **De Bock, J.** (2019). Data-driven summarization and synchronized second-screen enrichment of cycling races : using live and historical sports data to reinvent traditional reporting. In AI4TV '19 : Proceedings of the 1st International Workshop on AI for Smart TV Content Production, Access and Delivery (pp. 10–16). Nice, France: ACM. https://doi.org/10.1145/3347449.3357481

[3] The publications listed are recognised as ‘P1 publications’, according to the following definition used by Ghent University: P1: Proceedings included in one of these Web of Science indexes: 'Conference Proceedings Citation Index - Science' of 'Conference Proceedings Citation Index - Social Science and Humanities'. Limited to publications document type: article, review, letter, note, proceedings paper, with exception of publications classified A1.

3. Verstockt, S., Van den broeck, A., Van Vooren, B., De Smul, S., **De Bock, J.** (2020). Data-driven summarization of broadcasted cycling races by automatic team and rider recognition. In icSPORTS 2020, 8th International Conference on Sport Sciences Research and Technology Support, Proceedings (pp. 13–21). online: SCITEPRESS. https://doi.org/10.5220/0010016900130021

4. **De Bock, J.**, Van den broeck, A., Verstockt, S. (2020). GPS driven camera selection in cyclocross races for automatic rider story generation. In icSPORTS 2020, 8th International Conference on Sport Sciences Research and Technology Support, Proceedings (pp. 67–74). Online: SCITEPRESS. https://doi.org/10.5220/0010016200670074

5. Steyaert, M., **De Bock, J.**, Verstockt, S. (2022). Sensor-Based Performance Monitoring in Track Cycling. In International Workshop on Machine Learning and Data Mining for Sports Analytics (pp. 167-177). Springer, Cham.

6. Dumez, P., Prevost, G., Slembrouck, M., **De Bock, J.**, Marbaix, J., Verstockt, S. (2022). Video- and location-based analysis of cycling routes for safety measures and fan engagement. In R. Lienhart, T. Moeslund, & H. Saito (Eds.), MMSports '22 : proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports (pp. 29–37). https://doi.org/10.1145/3552437.3555706

1.6 Published (Online) news articles

1. "UGent onderzoekt BK veldrijden: grootste tijdverschillen op schuine kant, Sporza", 2020, https://sporza.be/nl/2020/01/11/ugent-onderzoek-bk-veldrijden-grootste-tijdverschillen-op-schui/

2. "Universiteitsonderzoek geeft inzicht in beslissende stukken BK-parcours: Schuine kant speelde grote rol", HLN, 2020, https://www.hln.be/veldrijden/universiteitsonderzoek-geeft-inzicht-in-beslissende-stukken-bk-parcours-schuine-kant-speelde-grote-rol˜afbc2f63/

3. "Veiligheidsmodel analyseert Tour: Etappe 5 en 10 het gevaarlijkst", NOS, 2020, https://nos.nl/collectie/13847/artikel/2346116-veiligheidsmodel-analyseert-tour-etappe-5-en-10-het-gevaarlijkst

4. "(On)veilige sprints in de Tour? Etappe 5 en 10 zijn risico-aankomsten", 2020, Sporza, https://sporza.be/nl/2020/09/02/veiligheid-sprints-ronde-van-frankrijk-id-lab/

5. "Could AI be the future of bike race coverage?", 2021, Cyclingtips, https://cyclingtips.com/2020/12/could-ai-be-the-future-of-bike-race-coverage/

6. "UCI gooit data in de strijd voor meer veiligheid", De Standaard, 2021, https://www.standaard.be/cnt/dmf20210224_98014045

7. "UGent wil wielrennen veiliger maken met uniek onderzoek: Wat zijn de oorzaken van zware valpartijen?", Nieuwsblad, 2021, https://www.nieuwsblad.be/cnt/dmf20210202_94146841

8. "Het grote crashrapport: de harde data achter valpartijen in het wielrennen", Knack, 2022, https://sportmagazine.knack.be/wielrennen/het-grote-crashrapport-de-harde-data-achter-valpartijen-in-het-wielrennen/

9. "Vuelta aan Nederland goed besteed: Behoort organisatorisch tot beste leerlingen", NOS, 2022, https://nos.nl/artikel/2441495-vuelta-aan-nederland-goed-besteed-behoort-organisatorisch-tot-beste-leerlingen

References

[1] Y. Kataoka and P. Gray. *Real-Time Power Performance Prediction in Tour de France*. In U. Brefeld, J. Davis, J. Van Haaren, and A. Zimmermann, editors, Machine Learning and Data Mining for Sports Analytics, pages 121–130, Cham, 2019. Springer International Publishing.

[2] T. Wade. *Tour de France 2021 – The technology used to create the world's largest connected stadium*. Technical report, June 2022. Available from: https://hello.global.ntt/tourdefrance/-/media/ntt/tdf/2021/tour-de-france-whitepaper.pdf.

[3] Deac, Cristian and Tarnu, Lucian. *Considerations on the role of modernizing the road infrastructure in the prevention of road accidents*. MATEC Web Conf., 290:06004, 2019. Available from: https://doi.org/10.1051/matecconf/2019 29006004, doi:10.1051/matecconf/201929006004.

[4] R. Elvik. *Road safety effects of roundabouts: A meta-analysis*. Accident Analysis & Prevention, 99:364–371, 2017. Available from: https://www.sciencedirect.com/science/article/pii/S0001457516304560, doi:https://doi.org/10.1016/j.aap.2016.12.018.

[5] Gitelman, A. S. Hakkert, E. Doveh, and A. Cohen. *A STUDY OF SAFETY EFFECTS OF ROAD INFRASTRUCTURE IMPROVEMENTS UNDER ISRAELI CONDITIONS*. 2001.

[6] L. Jateikienė, T. Andriejauskas, I. Lingytė, and V. Jasiūnienė. *Impact Assessment of Speed Calming Measures on Road Safety*. Transportation Research Procedia, 14:4228–4236, 2016. Transport Research Arena TRA2016. Available from: https://www.sciencedirect.com/science/article/pii/S2352146516304 008, doi:https://doi.org/10.1016/j.trpro.2016.05.394.

[7] United Nations Economic Commission for Europe. *Statistics of road traffic accidents in Europe and north America, VOLUME LVI, 2021*. United Nations, New York, NY, February 2022.

[8] D. V. Reeth and D. J. Larson, editors. *The Economics of Professional Road Cycling*. Springer International Publishing, 2016. Available from: https://doi.org/10.1007/978-3-319-22312-4, doi:10.1007/978-3-319-22312-4.

[9] D. B. John Lombardo. *Going gray: Sports TV viewers skew older*, Jun 2017. Available from: https://www.sportsbusinessjournal.com/Journal/Issues/201 7/06/05/Research-and-Ratings/Viewership-trends.aspx.

[10] Van Reeth, Daam and Lagae, Wim. *A blueprint for the future of professional cycling*. SPORT BUSINESS AND MANAGEMENT-AN INTERNATIONAL JOURNAL, 8(2):195–210, 2018. Available from: http://dx.doi.org/10.1108/sbm-02-2017-0010.

[11] E. Heinena and T. Gotschid. *Cycling: Past, current and future*. Cycling, page 1, 2022.

[12] M. Lewis. *Moneyball: The art of winning an unfair game*. WW Norton & Company, 2004.

[13] P. Laird and L. Waters. *Eyewitness Recollection of Sport Coaches*. International Journal of Performance Analysis in Sport, 8:76–84, 02 2008. doi:10.1080/24748668.2008.11868424.

[14] U. Brefeld, J. Davis, M. Lames, and J. J. Little. *Machine Learning in Sports (Dagstuhl Seminar 21411)*. Dagstuhl Reports, 11(9):45–63, 2022. Available from: https://drops.dagstuhl.de/opus/volltexte/2022/15917, doi:10.4230/DagRep.11.9.45.

2
Data in cycling

"Data doesn't decide, data informs."

–Arne Jaspers (sport scientist, RBFA)

This chapter discusses the central element of the research: the data powering all the analyses and visualisations. We will approach how data can be collected and further focus on the availability, sensitivity and limitations of data within the cycling context. This chapter will also briefly introduce the different types of data that were used in our research. Furthermore, we will discuss the used data in a general way including its quality, security and legality. The research question on which we try to find the answer in this chapter is **"What types of data can be identified within the professional cycling use case and what are the main opportunities and challenges of them?"**

2.1 Data in sports

Data has a very important role in our everyday lives and professional sports is no exception to this rule. Modern decision making is almost synonymous with using data as an important informing factor. In this chapter we will already briefly introduce the different types of data and provide a detailed explanation on how and why they each contribute towards this more general research goal (i.e., improve cycling safety, performance and story telling through multimodal data analysis).

2.1.1 Opportunities

Data is everywhere. After the 2022 edition of the "Tour de France" and complying to their yearly habit, tech giant NTT published their "Tour de France whitepaper" [1]. In this paper, it is very accurately described how data is ingested in the NTT servers. The data is very diverse, as it comes from riders, TV broadcasters, in-race cars etc. Bringing together and orchestrating this data flow is a huge challenge, but ultimately can result in nice visualisations and analyses that inform teams, fans and broadcasters. But in order to provide valuable data insights, the data often has to be pre-processed in quite a rigorous way. To lay a solid foundation, we will provide an in-depth focus on big data and data science with all its strengths, challenges and limitations. After this discussion, we will further drill down on the types of data that are especially useful in professional cycling.

2.1.2 Big data

In their overview work: "Big Data: A review", Sagiroglu and Sinanc [2] define it as a *term for massive data sets having large, more varied and complex structure with the difficulties of storing, analysing and visualising for further processes or results*.

Furthermore, big data can also be further characterised by its three main characteristics, which are often called the *3 V's*: standing for Variety, Volume and Velocity. Variety is pretty self explanatory but it basically means that data in a big dataset is not standardised, but rather composed of different types of data (e.g., images, text, GPS logs, etc.). Volume is the fact that datasets in the orders of terabytes and even petabytes are not an exception anymore. Finally, velocity is to distinguish and show the variation between data that should be processed in real-time versus in batch and the speed of which new data becomes available (e.g., every second, every day, every week, etc.). There is also a variant on the 3 V's which are the 5 V's [3]. In this variant veracity (which include the trustworthiness and authenticity of the data) and value (its relative importance to stakeholders and its impact on decisions) were added.

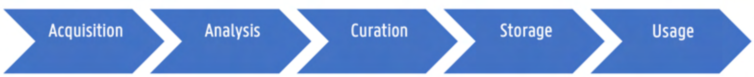


Figure 2.1: The big data value chain

Owning and mining data is one aspect, extracting value out of it and presenting it to stakeholders is a different one. For this purpose, a workflow called the

“Big Data Value Chain” was introduced by Curry et al. (2014) [4]. In Figure 2.1 an overview of the stages in the value chain is provided. In this Chapter we discuss the acquisition. In the remainder of this thesis we will largely focus on the second step in the workflow (i.e., the analysis) and the fifth step (i.e., the usage of the data and information for our use cases by the cycling community). Furthermore, in each of the chapters dedicated to the analysis of a certain type of data, we will also not forget to discuss how the dataset was annotated and validated (Step 3) and how data can be stored and/or knowledge can be shared (Step 4).

If we apply the introduced principles about big data on sports science we can conclude that the sports industry is no exception when it comes to data. A lot of data is generated and collected nowadays. For many sports including professional road cycling the collected data can be subjected to the who, what, where, when questions. Although it is in reality probably a bit more complex, recording and collecting data that contains answers on each of the four questions already ensures that data is ready for further analysis. For successful data science projects, the real challenge should be to provide an answer on the why and how questions [5]. But this is only possible if the provided data is of sufficient quality. Last but not least, we should also mention that sports often have a deliberately built-in factor of unpredictability [6, 7], so this should never be forgotten when collecting and analysing data (i.e., the risk to search for elusive patterns).

2.1.3 Data science

Data science is a buzz word that is widely and often wrongly used. Provost and Fawcett defined data science as a set of fundamental principles that support and guide the principled extraction of information and knowledge from data [8]. Likewise, data mining is the actual extraction process of knowledge with algorithms and technology, incorporating the data science principles. With this information the “what is it?” question might be answered, but the actual reason why we would like to apply data mining in any given context still remains unanswered. Even if the exact reasons why one might want to “mine data”, the general motivation should be to *help and improve the decision making process*. Additionally, we use algorithms and the help from computer programs for assistance in decision making, mainly because the volume of the data have outstripped the capacity of manual analysis and the fact that we have increasing and more readily and cheaply available computing power at our disposal. In the modern era we can use this computing power to extract valuable and relevant information out of the mass of data to support the decision making. Morgulev et al. further specify that a sports specific data science methodology actually consists of a three-stage process (see Figure 2.2) [9].

In the first stage, data is captured from different resources. It can be either

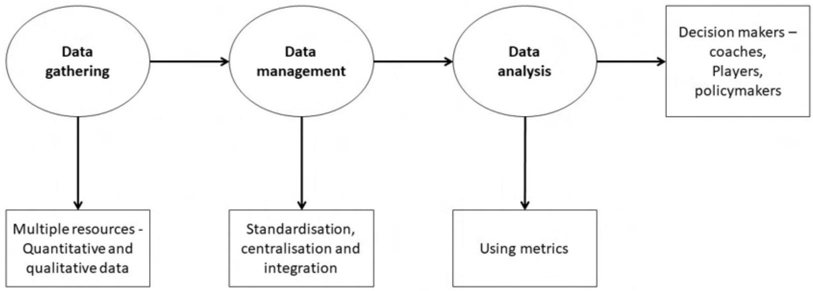


Figure 2.2: From Morgulev et al. [9]: a sports analytics framework

quantitative (i.e., measurable) or qualitative (i.e., properties, labels and identifiers). In the next stage, all these heterogeneous data sources should be brought together and unified. Usually, within this stage, data clean(s)ing is also performed. Depending on the quality of the data sources, this second phase can be quite labour-intensive. It is only in the final stage that we are actually mining the data - it is now that we find the answers on our questions and confirm/reject our hypotheses. As as final step, this knowledge is shared with decision makers, which within sport, can be anyone ranging from athletes, coaches and fans to broadcasters.

2.1.4 Sensor data in cycling

In cycling a lot of the data originates from sensors. According to the Oxford dictionary, a sensor is defined as "a device which detects or measures a physical property and records, indicates, or otherwise responds to it". As mentioned before, sensors can be split in two different categories: sensors worn on the athletes' bodies or bikes (e.g., power meter or heart rate monitor) and sensors that are not travelling with the rider (e.g., wind vane measuring the course circumstances). There is also a hybrid type of sensors where part of the system is not travelling with the rider and part of the system is on the rider's body or bike (e.g., a timing system with a transponder and measurement loops).

As described, a sensor measures a physical property with the aim to be monitored or analysed in greater detail. For our first category of sensors (i.e., those travelling along) the sensor is often measuring a performance related characteristic. The trajectory of the measured data onto a computer is very interesting to investigate. In a first step, the property needs to be captured by a device. For the power that the rider is producing with its legs, for example, a solution that is able to capture this force needs to be developed. Strain gauges are the perfect electronic components for this purpose, as they measure deflection through changes in resistance. Other popular sensors use electrocardiography (heart rate straps), magnetism (speed/cadence) or photoplethysmography (optical heart rate monitors) as capturing mechanisms.

In a next step, the captured parameter needs to be transferred to a monitoring or storage device. In early iterations of these sensors, this was often not wireless, but most recent versions have wireless communication built-in. Within the cycling sensor industry, ANT+ and Bluetooth Smart are the mostly used protocols to send data from sensor to monitoring/recording device. One of the big strengths of the ANT+ protocol is that there is no pairing procedure required to be able to capture sensor data. Every device that is listening for ANT+ data should be able to capture and decode it. This is useful but also implicates that one's sensor data is freely distributed in the ether. This is somewhat mitigated that a person cannot be directly identified by sensor values (i.e., the raw sensor data does not include the owner's identification). This also implicates that the pairing of a person's identity with the sensor identification is something that often needs to be performed manually.

Finally, timing systems also can be considered as as a (special) type of sensor data that is very useful and beneficial for analysis. Timing systems can use different technologies, but active/passive ultra-high frequency Radio-Frequency Identification (RFID) gateways are the most widely adopted ones. Aside from the exact technology, these systems detect when an athlete crosses a certain point on the course. This not only gives us timing information (i.e., how much time was needed to move from checkpoint x to y), but also tell exactly when an athlete was at a certain location (if the mapping between checkpoint and location on the course was made).

As soon as data is recorded and linked to an athlete the data analysis can be performed. In this thesis we will further focus on the possible analyses with this data.

2.1.5 Challenges

2.1.5.1 Quality and/or noisiness

Quality is a term that is used outside the scope of data as well. The first association of data with quality was by the MIT Total Data Quality Management research group [10]. They concluded that the much older and general definition of quality, i.e., "fitness for use" is mostly adopted in data-specific literature. If we interpret this statement, we would falsely conclude that data that is appropriate for the use case (e.g., a set of crash-related Tweets to automatically detect causes) is the only requirement for data to be qualitative. In this subsection, we will further define a more modern definition of qualitative data.

Furthermore, we could also define why we need qualitative data. Although this is rather obvious, qualitative data are a precondition for guaranteeing the value of performing data science techniques [11]. In 2020, Ramasamy and Chowdhury [12] defined the concept of "Data Quality Dimensions (DQD)" as "a set of data quality attributes that represent a single aspect or construct of data quality". The most common dimensions are Accuracy, Completeness, Consistency, Freshness, Uniqueness and Validity. Accuracy is a measure for how well data represents the real-world values. Completeness measures if data is sufficient to deliver meaningful decisions. Consistency tells us if the same data that might be also stored elsewhere matches. The freshness of data indicates if the data represents the reality at the required point in time. With uniqueness, we check if there are not any duplicates in your data that might affect processing. Finally, integrity is an indication if data is maintained correctly and not changed unintentionally when it was stored elsewhere. Throughout the years, these DQD's have been slightly modified and new ones were added to the list. However, research within this area is still evolving. More research should be performed on selecting the DQD's that are truly relevant for big data and the studying of big data quality related issues projected on these data quality dimensions.

In conclusion, we can state that ensuring data quality in a data science project is critical, although not straightforward. This is why data exploration before tackling a data science project is crucial to get an insight in the quality/noisiness of the provided dataset. In data mining, we define noisy data as data with a large amount of additional, meaningless information called noise, making it very hard or sometimes impossible to mine insights from the data. This is why, in our research, all the datasets were often manually pre-checked if they were not containing outliers or irregularities. Outliers or irregularities can be defined as observations that are abnormally far away from other values of a random sample from a population. For the stage hardness metric for the UCI-safety project for instance, we had to carefully check the courses' elevation data, and ultimately had to correct it with digital elevation model data as it was very inconsistent across the dataset.

2.1.5.2 Availability

The availability of health- and sports-specific data is not that broadly discussed in literature. If we would define "data availability" we could define it as: a measure or indication of how readily available the data is for data science. The best available data is data that can be directly imported in programming frameworks such as Python or R. Within cycling there are not really great examples of such easily accessible datasets. Usually, some preliminary steps need to be performed to obtain a clean and programming/data science friendly dataset. An example of mediocrely available data are pro cycling race results, which are usually scraped from the HTML source code of websites such as https://www.procyclingstats.com/ or https://www.uci.org/. For video data we can classify the availability in a similar manner. Some video is directly shared from the owner, but other videos need to be downloaded from online streaming providers (e.g., YouTube, Vimeo, Eurosport Player, TIZ-cycling). Some data, that is often more personal (e.g., sleep data, resting heart rate, Heart Rate Variability) is not publicly available, and is usually only shared when close collaboration between researchers and data owners exist. During our research, we wrote multiple scrapers. Two examples are the results scraping of Procycling stats for the bunch sprint predictor and the Mapillary scraper to get street views for road traffic infrastructure model training. This brings us seamlessly to the next challenge, data sensitivity, which we will discuss in the following subsection.

2.1.5.3 Personal data sensitivity

In May 2018, the European Union imposed a set of new rules that are intended to protect personal digital data. This set of rules is better known as the General Data Protection Regulation (GDPR). As with all governmental regulations, the exact formulation and extent is quite long, but in this section we will sum up the key takeaways and considerations that have to be made when working with sports and athlete-specific data in a research-specific context.

As a starting point, the following data types are often discussed within a GDPR-specific context: personal data, sensitive personal data, pseudonymised data and anonymised data. Only truly anonymised data fall without scope of the GDPR rules and principles. For all other types of data the GDPR principles still need to be taken into account.

As a very brief summary, we can conclude that GDPR is based on six basic principles: lawfulness, purpose limitation, data minimisation, accuracy, storage limitation and confidentiality/integrity. Lawfulness is probably the most abstract principle of them all. This rule determines when data processing can be considered as lawful. We have a total of six options and lawfulness is ensured if at least one has been respected. For research, we can either specifically ask for specific consent or we can also enforce lawfulness by proving that a task carried out in the public interest or in the exercise of official authority vested in the controller [13, 14]. The purpose limitation imposes to only use and process personal data for the purpose of your research, and this processing must be reasonable and proportionate to the purpose of your research. The minimisation requirement means that you may only use the personal data necessary to achieve the objectives of your research. The personal data has to be processed accurately and it may not be kept longer than necessary for your current research or for possible further analyses of the data.

This brings us to a final GDPR-specific consideration: data controller or data processor. If an organisation or researcher is the data controller, they decide 'how' and 'why' data should be processed. A data processor processes data on behalf of the controller [15]. If we apply this on the research performed in this thesis we can conclude that a lot of the studies that have been performed were as a data processor. External industry partners and/or organisations delivered the data and asked a specific question on what to do with the data they delivered. In the incident database of the UCI, for instance, these principles needed to be carefully implemented. As an example, the database contains all information required to provide a statistic of riders that cause or were involved in the most crashes, but this is not the type of information that can be made publicly available to other riders, teams or media.

2.1.5.4 Limitations

In the previous subsections we covered the more obvious data-specific challenges, but there are still some final remarks and considerations that have to be made. These considerations can be easily called the last hurdles in convincing or pitching a data science project.

The first one is the fact that data science projects are often treated quite similarly to other kind of projects (e.g., building a new sports hall, recruiting a new coach and staff, ...). For instance, if a data scientist designs a workflow for a team or federation to help scouts and talent developers to find the new "Remco Evenepoel" with the help of data, they ideally want to see return on investment. Often they want that return rather quickly, so it might be challenging to pitch your idea knowing that the benefits of proper talent detection and management will only become clear in a couple of years. This is also very much the case for our data capturing project with Cycling Vlaanderen. The learning curve was steep and set-up cost was relatively high, so the idea had to be pitched with a future-based mindset to the stakeholders. After one year of implementation, testing and data capturing, the real value of the set-up became apparent when it was used to optimise the team pursuit pacing plan.

The next thing, which might probably also sound very familiar to a more general data scientist is the lack of a data set that is large enough and of decent quality. It is not uncommon that a stakeholder is convinced to have a qualitative and quantitative dataset, but after thorough data exploration it becomes clear that the data has not been recorded consistently or that it contains only a few test subjects. Within data and computer science the general rule of thumb "garbage in, garbage out" is often very relevant in these kind of scenarios. An unfortunate example of this "garbage in, garbage out" principle was during a project with some Belgian professional soccer clubs. The club staff had to take various fitness and anthropometric tests to help predict injury before they actually happen. The problem in this project was that testing procedures across teams was often so different that the test results could not be used to train such an algorithm.

The last important consideration that has to be made is the gap that often exists between data scientists and stakeholders. It is not always easy for data scientists to fully understand the context they're working in. For instance, it might be extremely hard for a data scientist to understand the training logs of athletes if they do not have sufficient knowledge about sports physiology or if they do not know how to interpret heart rate monitor or power meter data. The same is true the other way round, when discussing the insights mined from the data, data scientists should be extremely careful to present and communicate their conclusions and findings in such a way that it is understandable by a person that does not have the same technological knowledge as they have. An example of this last consideration was noticed during the "Becoming an Olympian in Sports Analytics" summer school that was held by the research group. In this summer school, different people, each with different professional background (e.g., computer scientists, physiologists or coaches) were brought together to teach them the core principles of sports analytics. During the hackathon, where they tackled a sports-specific research problem in groups, we could see that people with different backgrounds did not always understand each other.

2.2 Types of data

2.2.1 Video data

Video data is a recorded sequence of still images which can be accompanied by an audio signal. The images in the video have a certain resolution, which is the number of pixels contained in each frame and is a measure for the amount of detail (see Figure 2.3) in a single video frame. Common video resolutions are 1920x1080 (High-Definition) and 3840x2160 (4K). The frame rate of a video determines how many frames per second are recorded and determines how much movement in between frames can be captured. This also influences what information and what accuracy can be expected from the footage. When the pedalling behaviour of a cyclist is analysed for instance, we might not get sufficient details when we record at 25 frames per second (i.e., one frame every 40 milliseconds). On average, the pedalling cadence of a cyclist is around 80 rotations per minute, so one rotation takes only 750 milliseconds, which results in around 19 frames per pedalling cycle (i.e., a frame every 19 degrees). With this frame rate, we might for instance miss the top or bottom dead centre.

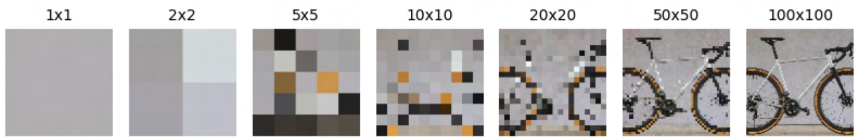


Figure 2.3: Impact of image resolution on the amount of detail that can be captured in a frame, the numbers 'WxH' represent the number of horizontal and vertical pixels in each image

One aspect that has to be considered when handling video is that it needs to be encoded. With video encoding, the recorded sequences of images are prepared for output, storage and proper playback and processing by video viewers. If we would record video in a raw sequence (i.e., a sequence of images) the video sizes would be huge. For instance, if we would record 10 minutes of video at 30 frames per second in Full HD, we need to store 18 000 (10x60x30) single images. A single 1920x1080 image at a common bit depth (i.e., level of colour detail) of 3 bytes/pixel (1 byte for every colour channel) needs 6.2 Megabytes (MB) (1920x1080x3). So in summary this 10 minutes of video would take 112 Gigabytes (GB) of space!

This is why encoding and compression is needed for video data. As an example, for the UCI safety project, we had to work with GoPro recordings of the courses of multiple stage races and one day classics. For 2022, we analysed over 600GB worth of video data! To keep the time to upload and download the course recordings to a minimum, compression and encoding is obviously needed. A term that is often used within the encoding and decoding process of video are codecs. These are a collective term for the standards that are available for video encoding and decoding. Two of the more popular codecs are H.264 and the newer H.265. An extensive explanation of how codecs do the, sometimes rather impressive, compression would lead us too far out of scope in this thesis. A brief explanation, however, is required to fully grasp why it is beneficial to compress videos. Compression (and its counterpart decompression) always has costs attached to it. These costs are both time (i.e., the time needed for encoding) and loss of information. The latter, is the result of spatial and temporal encoding. Spatial encoding reduces the amount of information on a single frame and temporal encoding reduces the amount of information that is needed to produce a "moving image". As an example, for a course analysis of the Amstel Gold Race of 2023, nearly 3 hours of GoPro video data was compressed from 31GB to 9GB. GoPro action cameras already perform minor compression, but this is limited to reduce the stress on battery and computing capacities.

For the processing of video in the Python programming language, we can count on the OpenCV computer vision library. With this library a video can be automatically read in memory frame-by-frame (so with decoding behind the scenes). Furthermore, the OpenCV library contains lots of pre-implemented computer vision algorithms (see Figure 2.4 for an example) to ease the development for researchers.

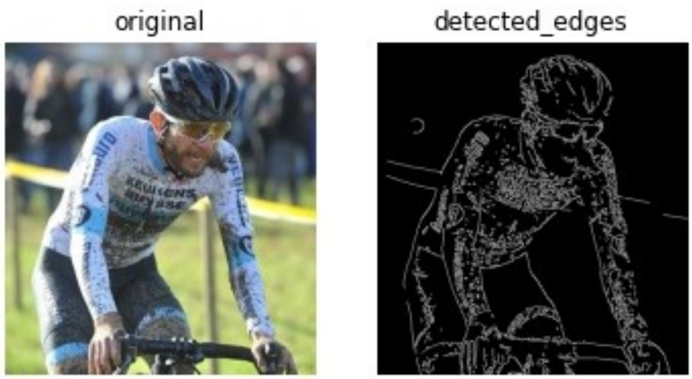


Figure 2.4: Demonstration of the Canny edge detection in the OpenCV computer vision package.

2.2.2 Textual data

A search for the word "text" in the Macmillan dictionary learns us that text can be defined as *the part of a book, magazine, or computer document that consists of writing and does not include pictures or notes.* If we apply this definition on our topic we will define textual data as anything that is either structured or unstructured text. Structured data are datasets that comply to a preset data model or structuring convention. The data is often stored and exchanged in structured file formats (e.g., JavaScript Object Notation and Comma Separated Values) or databases (e.g., MySQL or MongoDB). This makes that this sub-type of data is very suitable for data science as it greatly reduces the time spent in cleaning and pre-processing the datasets.

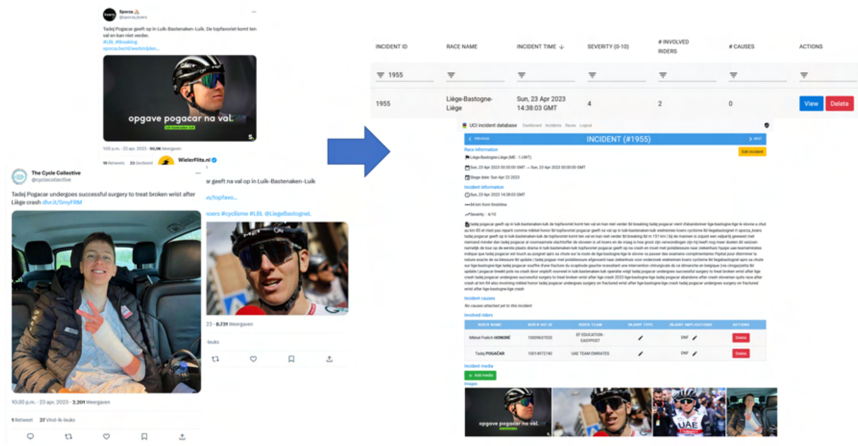


Figure 2.5: Illustration of the conversion process of unstructured Tweets about Tadej Pogačar's crash in Liège-Bastogne-Liège 2023 into structured records in the incident database.

Unstructured data on the other hand can be defined as data that does not have a predefined data model and/or it is not defined in a predefined way. Although its structured counterpart is much more data-science friendly, research suggest that around 80% of the available data was unstructured in the average company in 1998 [16]. It is true that most companies are making a digitisation catch-up effort and the exact percentages might have changed a bit, but it still nicely illustrates that a potential wealth of information is hidden within unstructured datasets. A great example of unstructured treasuries are the comments on social media platforms. Zhi et al. [17] for instance used all basketball sub-Reddit comments to analyse the spectators' sentiment before/during/after basketball games. Based on these comments, they performed a behavioural study to find patterns in the comments of the fans. They found that most of the comments consisted of player and/or team appraisal. Furthermore, they also developed a prototype application that could quickly extract key game information based on comments in order to give a data driven game-experience for both sport fans and story writers. Within our own research, the value of the conversion process is illustrated in Figure 2.5 where the unstructured text from a crash related set of Tweets was transformed into structured information in the database which can be used to gain insights in the causes and severity of road cycling incidents.

With this knowledge in mind, we will further investigate how we can use both structured and unstructured data in our data-analysis for the professional cycling use case. An in-depth study of how textual data is used within this PhD is presented in Chapter 4.

2.2.3 Geo-spatial data

According to IBM's website, geo-spatial data can be defined as: *"information that describes objects, events or other features with a location on or near the surface of the earth"* [1]. In sports, geographically tagged data is rather ubiquitous and especially useful for further analysis. With Global Positioning System (GPS) chipsets readily and cheaply available and built-in in portable devices such as smartphones, watches and dedicated GPS units a lot of location-tagged data is generated. GPS coordinates are unique identifiers of a precise location on the earth. The exact location is often provided as a latitude, longitude pair in a coordinate system. But location can also be provided in other ways. It is, for instance, possible to infer location of players on a pitch with one or multiple calibrated cameras. The location that comes out of this localisation is then relative to a reference point on the field (e.g., centre spot of a soccer field) [18].

Within field team sports some companies focus on localising the players on the playfield and calculate metrics based on the collected GPS samples of the players (e.g., accelerations, decelerations, jumps, ...) [19]. For this purpose, players are equipped with dedicated data loggers that collect their playing behaviour throughout training sessions and matches. Additionally, the platforms often also provide the ability to equip players with heart rate monitor and collect a measure for exertion throughout the session. This can have favourable effect on injury monitoring and prevention [20].

In (professional) cycling Global Positioning System (GPS) data is also very much inherent to the sport. The use of GPS-enabled watches and head-units (e.g., Garmin, Polar or Wahoo) is very much embedded in the behaviour of cyclists out on the road. Furthermore, cyclists are also very keen on uploading their workouts to online platforms such as Strava (Figure 2.6), Komoot, RouteYou or Trainingpeaks to analyse their performances and compare themselves with others. But apart from its competitive purposes, cycling GPS data could also be used to improve road infrastructure [21] or to understand and improve routing choices [22].

Furthermore, within cycling, the last couple of years we can observe the tendency to provide power, heart rate and location data during races in real-time [23] [24]. This opens a lot of possibilities for real-time analysis and predictions in a race situation. This extra information provided to fans will further help to improve the engagement and fan experience.

In this research, GPS data will be a centrepiece in most of the implemented use cases within the studies. Additionally, combination of geo-spatial data with

[1] https://www.ibm.com/topics/geospatial-data

other sources/types of data is a crucial part in analysis and data mining within our research. A thorough overview of the geo-spatial research performed in this PhD will be presented in Chapter 5.

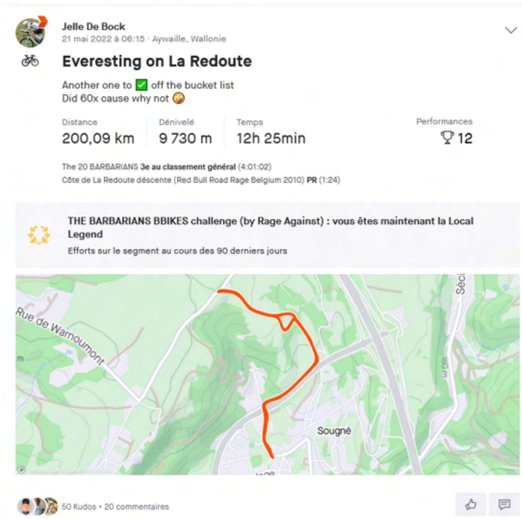


Figure 2.6: Geo-spatial data on Strava, one of the most popular GPS workout sharing platforms

2.2.4 Time series data

The emerge of wearable devices has greatly facilitated digital health and sports monitoring. Some works even refer to the current era as the age of the “Quantified-Self” [25]. With wearables, users gather more data about themselves, so this necessitates appropriate analysis methods and tools. For instance, Whoop, the wearable health monitoring device, generates more than 150 MegaBytes of data every single day [26].

This specific type of data has one important characteristic, it is timestamped. Every data point has a specific point in time when it was captured. The type of information that is collected over time can range from heart rate to the number of people entering/exiting a cyclocross venue. Just as with video data, time series data has a resolution at which the data was captured. The resolution at which data is recorded is both technology and use case dependent. As an example, in order to capture the fast direction changes of football players on the field, the resolution of the GPS of the devices usually ranges from 5Hz for regular up to 15Hz for elite use cases [27].

In cycling, geospatial (GPS) data is almost synonymous for time series data. The data captured by wearables and GPS head units record time stamped logs of the coordinates and the sensors attached to the athlete. Typical sampling rates are 1 recording per second (i.e., 1 Hz), so a one hour bike ride with 4 additional sensors (e.g., heartrate, cadence, power and body temperature) provide 4 x 3600 = 14 400 data points annotated by time, latitude and longitude. It needs no further explanation that this is a wealth of information for the data scientist. In the publication of Jobson et al. [28], for instance, the authors try to provide a theoretical, data driven approach to understanding training intensity quantification and its correlation with physical performance. In a more recent effort, Hilmkil et al. [29], managed to predict heart rate of an athlete based on elapsed time, speed, accumulating distance, instantaneous power, instantaneous cadence, instantaneous power to weight ratio, altitude and his/her last 30 second average heart rate.

2.3 State of the art data processing techniques

In this thesis we will use state-of-the-art methodologies to contribute towards safety, storytelling and performance in cycling. In this section, a global overview of the most important frameworks and methodologies for each data type is given.

2.3.1 Video

For video processing, we mainly rely on two types of processing techniques. A first family of techniques are the traditional computer vision methodologies such as edge detection, Scale-Invariant-Feature-Transform (SIFT) or mean-shift. These methodologies are relatively easy to understand and are easy-to-implement in code thanks to the popular OpenCV library. These methodologies were mainly used to complement the next category of video processing methods, being the machine learning (ML) based methods. In this dissertation several video-based ML methodologies are used that use convolutional neural networks (CNNs). Without going into too much technical detail, a CNN is a deep learning model specialised for visual data, like images. It uses layers for feature extraction (convolutional), dimension reduction (pooling), and prediction (fully connected). By learning patterns and features from data during training, CNNs excel in tasks like image recognition and object detection, making them a vital tool in modern machine learning. For object detection, we used the YoloV5 object detection model for several tasks (e.g., rider modus detection and traffic infrastructure detector). For mapping the joints on the body of a person (i.e., pose estimation), we compared a number of frameworks, but finally selected the Alphapose model to be included in our computer vision methodologies. Finally, for image segmentation, which is the process of dividing an image into distinct, meaningful regions or segments to analyse and understand its content, we used several different frameworks such as Dectron2, Humanparser's body segmentation model and DeeplabV3+.

2.3.2 Text

In text processing, we mainly need to perform two important tasks. The first task is getting the data ready for analysis. Text data can be either structured or unstructured and be readily available or needs to be scraped. To reveal information and extract meaning from unstructured text (i.e., written and or spoken text) we need methodologies and tools to perform the task at hand. Natural Language Processing (NLP) is the collective term for the methodologies that have this exact purpose. In this thesis, we perform part-of-speech tagging and named entity recognition (NER) with the Spacy library. For text classification (i.e., classifying if a piece of text based on certain keywords and how they are correlated we use a Bidirectional Encoder Representations from Transformers (BERT) that is used to detect if Tweets on the Twitter platform were bicycle crash related or not.

2.3.3 Geospatial and time series analysis

For processing of GPS annotated time series we use several methodologies and tools. To provide a complete overview, the geospatial database "OpenStreetMap" (OSM) has to be mentioned before anything else. This geospatial database was queried with the Overpass Query Language. With the OSM database we can query geospatial metadata for every coordinate pair. As this is a community based database, the level of detail returned for a location might vary. In this research the OSM database was used to provide the course safety checking mechanism with traffic infrastructure elements such as roundabouts or speed bumps. Course information or geospatially annotated performance data (e.g,. power/heart rate files) are usually stored in a specific file format. The most common file types are GPS eXchange (GPX) and Flexible and Interoperable Data Transfer (FIT). These files can be easily parsed in the Python programming language with libraries such as gpxpy and fitparse. Last but not least, coordinate pairs can also be enriched with elevation data. Elevation features help to train the sprint speed and bunch sprint prediction models for the safety methodologies. Elevation data can be directly captured by GPS recording devices (e.g., Garmin or Wahoo GPS head units) or they can be queried from Digital Elevation Models (DEMs).

2.4 Conclusion

In this chapter, we tried to provide the reader with holistic and hopefully clear view on data within modern professional sports and cycling. Collecting, storing, owning and using data offers a lot of possibilities but also provides plenty of challenges. In the remaining chapters we will focus on the use cases that can be tackled with each of the four discussed data types (i.e., video, text and geographically annotated data that is either timestamped or not).

References

[1] T. Wade. *Tour de France 2022 – The technology behind the world's largest connected stadium*, 2022. Accessed: 2023-04-12. Available from: https://services.global.ntt/en-us/about-us/tour-de-france-report.

[2] S. Sagiroglu and D. Sinanc. *Big data: A review*. In 2013 International Conference on Collaboration Technologies and Systems (CTS), pages 42–47, 2013. doi:10.1109/CTS.2013.6567202.

[3] P. Chandarana and M. Vijayalakshmi. *Big Data analytics frameworks*. In 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), pages 430–434, 2014. doi:10.1109/CSCITA.2014.6839299.

[4] E. Curry. *The big data value chain: definitions, concepts, and theoretical approaches*. In New horizons for a data-driven economy, pages 29–37. Springer, Cham, 2016.

[5] D. Patel, D. Shah, and M. Shah. *The intertwine of brain and body: a quantitative analysis on how big data influences the system of sports*. Annals of Data Science, 7(1):1–16, 2020.

[6] J. Stockrahm. *The Unpredictability of Sports (1997)*. 1997.

[7] R. Hristovski. *Unpredictability in competitive environments*. BOOK OF, page 32, 2017.

[8] F. Provost and T. Fawcett. *Data science and its relationship to big data and data-driven decision making*. Big data, 1(1):51–59, 2013.

[9] E. Morgulev, O. H. Azar, and R. Lidor. *Sports analytics and the big-data era*. International Journal of Data Science and Analytics, 5(4):213–222, 2018.

[10] R. Y. Wang and D. M. Strong. *Beyond Accuracy: What Data Quality Means to Data Consumers*. Journal of Management Information Systems, 12(4):5–33, 1996. Available from: https://doi.org/10.1080/07421222.1996.11518099, arXiv:https://doi.org/10.1080/07421222.1996.11518099, doi:10.1080/07421222.1996.11518099.

[11] L. Cai and Y. Zhu. *The Challenges of Data Quality and Data Quality Assessment in the Big Data Era*. Data Science Journal, 14, 05 2015. doi:10.5334/dsj-2015-002.

[12] A. Ramasamy and S. Chowdhury. *Big Data Quality Dimensions: A Systematic Literature Review*. Journal of Information Systems and Technology Management, 17(0), 2020. Available from: http://www.jistem.tecsi.org/index.php/jistem/article/view/3126.

[13] *GDPR and the Research Process: What You Need To Know*. Available from: https://www.openaire.eu/blogs/gdpr-and-the-research-process-what-you-need-to-know.

[14] *Research Data Management (RDM)*, May 2022. Available from: https://www.ugent.be/en/research/datamanagement.

[15] *What is a data controller or a data processor?*, Dec 2019. Available from: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/controller-processor/what-data-controller-or-data-processor_en.

[16] S. C. C. *Enterprise Information Portals*. Merrill Lynch In-depth Report, 1998. Available from: https://cir.nii.ac.jp/crid/1572261550727302528.

[17] Q. Zhi, S. Lin, P. Talkad Sukumar, and R. Metoyer. *GameViews: Understanding and Supporting Data-Driven Sports Storytelling*. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery. Available from: https://doi.org/10.1145/3290605.3300499, doi:10.1145/3290605.3300499.

[18] R. Zhang, L. Wu, Y. Yang, W. Wu, Y. Chen, and M. Xu. *Multi-camera multi-player tracking with deep player identification in sports video*. Pattern Recognition, 102:107260, 2020. Available from: https://www.sciencedirect.com/science/article/pii/S0031320320300650, doi:https://doi.org/10.1016/j.patcog.2020.107260.

[19] L. Hennessy and I. Jeffreys. *The current use of GPS, its potential, and limitations in soccer*. Strength & Conditioning Journal, 40(3):83–94, 2018.

[20] F. E. Ehrmann, C. S. Duncan, D. Sindhusake, W. N. Franzsen, and D. A. Greene. *GPS and injury prevention in professional soccer*. The Journal of Strength & Conditioning Research, 30(2):360–367, 2016.

[21] K. C. Heesch and M. Langdon. *The usefulness of GPS bicycle tracking data for evaluating the impact of infrastructure change on cycling behaviour*. Health promotion journal of Australia, 27(3):222–229, 2016.

[22] J. M. Casello and V. Usyukov. *Modeling cyclists' route choice based on GPS data*. Transportation Research Record, 2430(1):155–161, 2014.

[23] Velon. *About Velon*. Available from: https://www.velon.cc/about-velon.

[24] R. Maker. *A Look At Velons’ Live Race Tracking System*. Available from: https://www.dcrainmaker.com/2018/01/velon-live-tracking-system.html.

[25] D. Lupton. *The quantified self*. John Wiley & Sons, 2016.

[26] Whoop. *Whoop closes $12m, announces availability of its performance optimization system exclusively designed for elite athletes and teams*, Jun 2022. Available from: https://www.whoop.com/thelocker/whoop-closes-12m-announces-availability-of-its-performance-optimization-system-exclusively-designed-for-elite-athletes-and-teams/.

[27] E. Schulze, R. Julian, and S. Skorski. *The accuracy of a low-cost GPS system during football-specific movements*. Journal of sports science & medicine, 20(1):126, 2021.

[28] S. A. Jobson, L. Passfield, G. Atkinson, G. Barton, and P. Scarf. *The analysis and utilization of cycling training data*. Sports medicine, 39(10):833–844, 2009.

[29] A. Hilmkil, O. Ivarsson, M. Johansson, D. Kuylenstierna, and T. van Erp. *Towards machine learning on data from professional cyclists*. arXiv preprint arXiv:1808.00198, 2018.

3
Video analysis in cycling

"In ictu oculi."

–The Bible, 1 Cor. 15:52

This chapter starts with a brief discussion about how video is used within professional cycling. It focuses on the state-of-the-art computer vision techniques and algorithms that were used within a cycling context. This chapter discusses the entire flow from video capturing to algorithms' and computer vision models' output. The research question of this chapter can be formulated as: **"Can we use a well-thought combination of computer vision techniques and artificial intelligence algorithms to obtain cycling specific scene understanding and identification of entities (e.g., riders, bikes or spectators)?"**

3.1 Video data in cycling

As we discussed in Chapter 1 there is a decreasing interest in the traditional way of race broadcasting. The Belgian media information institute (CIM) provided the viewing numbers for all cycling races between 2017 and 2022. If we directly compare 2022 with 5 years ago (i.e., 2017), almost all races face a remarkable decline over the last years (see Figure 3.1). David Hil, Chairman and CEO of Fox Sports Television, states it quite boldly: *"if sport is not working on TV, is not attracting an* *audience, is languishing, the problem lies with the people controlling that sport. They have either failed to keep that sport vital and alive, through laziness or mismanagement, or they have allowed people presenting their sport to the public through television to get away with sloppy, lazy or inattentive production."*

		2017	2022	
Omloop het Nieuwsblad	OMLOOP Nieuwsblad	812 060	675 632	-17 %
Kuurne-Brussel-Kuurne	KUURNE BRUSSEL KUURNE	808 096	587 224	-27 %
E3 Harelbeke	E3 HARELBEKE SAXO BANK	421 371	413 746	-2 %
Ronde Van Vlaanderen	DE RONDE	891 265	880 049	-1 %
Paris-Roubaix	PARIS ROUBAIX	797 053	790 293	-1 %
Amstel-Gold-Race	AMSTEL GOLD RACE	819 524	681 302	-17 %
Liège-Bastogne-Liège	LIEGE BASTOGNE LIEGE	581 820	675 990	+16 %
Tour de France (avg. stage)	le Tour de France	472 857	569 518	+20 %

Figure 3.1: Remarkable decline of number of spectators for 6 of the most popular races in 2017 and 2022

The reason for this decline is probably a bit more nuanced than David Hill's opinion, but it is definitely not singular, i.e., there are multiple contributing factors to the decreasing interest of the television audience. The first reason is that races are often too long to be attractive all time and for several of them the end result is rather predictable. Based on the race/track features (such as the amount of mountains/hills, total ascent and the hardness of final), one can easily predict if it will end in a bunch sprint or not. For this purpose a random forest machine learning algorithm trained on the workout files of multiple sprinters in multiple races was built. This is discussed in more detail in Section 5.4.2 of Chapter 5. Secondly, the way in which the race is captured also did not change much over the last decades. All these factors contribute to a weakening interest during the last years, especially of young generations. Related to this, it is important to mention that not only cycling, but also other sports see a quick rise in average age of TV viewers as younger fans shift to digital platforms or drop out [1]. There is an increased interest in short-term digital interactions such as team/rider stories and highlights that can be shared and viewed on different platforms. Most sports reporting, however, is still old fashioned and does not fully exploit the existing technology and digital platforms/tools. We need more personalised, interactive experiences to keep the end user happy and to get back the youngsters. However since road cycling broadcasts are also used to promote tourism (i.e., it offers cities and other points of interest an invaluable exposure to a broad audience), we must ensure that this economic aspect can continue to play a role in the proposed new formats [2].

The growing availability of data sources from video or Internet of Things (IoT) devices (such as sensors attached to bikes or athlete wearables) are offering a huge potential to make reporting more interactive, informative and attractive. These data sources, however, are still hardly used in sports events broadcasts. The missing gap is the adequate translation of the sensor data into useful narrative elements ("something happened") and the selection of the correct video fragments that tell this story. A striking example of a story that was not seen during the live broadcast of the Tour of Flanders 2019 - but reported in post-race summaries afterwards - was the team tactic of team Sky/Ineos to slow down the entire peloton at the beginning of several hills to save power for the rest of the race.

3.2 Video processing

3.2.1 Introduction

Video processing can be defined as a particular case of signal processing. It is basically the processing of time-stamped images. Each still image, which is often called a "video frame", contains a number of pixels that record light/colour intensity on a certain spot on the sensor, providing a visual representation of the captured "world". The higher the frame rate (i.e., the numbers of frames recorded every second), the more information the video will contain. The human eye and brain will perceive a video excerpt as smooth starting from 24 frames per second. If the frame rate is lower it will appear stuttery. This implicates that higher frame rates will allow video makers to make slow motion videos that will still look natural to the human eye. The processing that will be elaborated on in this thesis will focus on the frame content, i.e., we will focus on everything but the compression and encoding process, which we think deserves a (or multiple) thesis(ses) on its own.

In the following sections we will discuss the key video processing techniques and methodologies that exist and can be applied to our professional cycling use case. Please note that the provided techniques list is not exhaustive, but we limited ourselves to the most promising ones for cycling specific analytics.

3.2.2 Team and rider identification

An import aspect within professional road cycling video processing is the detection of who is in the field of view of the camera at which moment in the race. This information provides us a fundamental building block towards the goal of automated race summarisation. To perform this task an advanced computer vision/machine learning methodology was developed and combined with information retrieved from sensor data and race metadata. *Please note that this is a partial rework of the publications "Data-driven summarisation of broadcasted cycling races by automatic team and rider recognition" [3] presented at the ICSports conference in 2020 and the MDPI article "Video-Based Analysis and Reporting of Riding Behavior in Cyclocross Segments" [4].*

3.2.2.1 Related work on team/rider recognition

Related work on rider and team recognition in cycling is rather scarce or not existing, i.e., there was not a single hit in literature. However, in other sports like basketball, soccer and football, several approaches have been proposed over the last decade on how to identify players.

The jersey number recognition proposed by Liu and Bhanu, 2019 [5], for example, makes use of a Region-based Convolutional Neural Network (R-CNN) trained on persons/players and digits. Region proposals are geometrically related and fed to an additional classifier, ultimately generating jersey number proposals. Some results of this approach are shown in Figure 3.2. When a player is turned backwards to the camera the recognition goes well, but in any other orientation problems can occur, as shown in the last two examples. Furthermore, for any frontal view this approach will depend on the tracking of previous detections. Tracking in sports videos, especially when people are wearing similar sportswear and have a lot of occlusions, is also error-prone.

Mahmood et al., 2015 [6] introduced Adaboost based face detection and recognition of baseball players in action. A shortcoming of this algorithm is that it requires that the detected players' faces are frontal or near frontal. Again, when the orientation of a player changes, tracking issues can obfuscate the identification. In cycling this requirement is almost never fully met as riders wear helmets and often cover their faces with sunglasses. Broadcasted cycling videos continuously switch between cameras and viewpoints, which makes tracking even more difficult, i.e., none of the previously mentioned solutions would give us satisfying results. This is the main reason why we decided to develop a pose-based methodology that works on frontal, lateral and dorsal views, and tracks riders within the same shot when no occlusions occur. In case of occlusions, the tracking stops and the detection algorithm tries to resolve them. Based on the pose and the type of

Figure 3.2: Results of the jersey number recognition proposed by Liu and Bhanu, 2019

shot, decisions between using face recognition, jersey recognition and/or number recognition are made, and available sensor data is used to further filter or verify the set of possible candidates.

3.2.2.2 Sensor data enhancement

As highlighted earlier in this thesis, several third parties (such as Velon and Gracenote) provide structured and very detailed live data of sports events at a high frequency. Velon, for example, uses a tracker that can capture heart rate, power, and speed of each cyclist during several stages of Tirreno Adriatico 2019. This information is sent in real-time over a 4G connection to their servers, allowing live data display and analysis. Gracenote provided exact location of each group of riders during the Tour of Flanders 2019. If such sensor data of the race is available, it is definitely the most accurate and computationally most interesting solution for geo-localisation and event detection. When there are multiple groups of riders, however, an additional method (such as team or cyclist recognition) is needed to know which particular group is shown in the live video stream.

If detailed sensor data were available, several events can be detected in it, such as breakaways, crashes, or difficult sectors (e.g. barriers and sandpits in cyclocross or gravel segments in road cycling). For the latter type of events, the approach of Langer et al. [7] for difficulty classification of mountainbike downhill trails can, for example, be tailored to cyclocross and road cycling segment classification. The work of Verstockt et al. [8] in 2014 also shows that this is feasible. For breakaway/crash detection, experiments revealed that simple spatiotemporal analysis across all riders will already provide satisfying results.

3.2.2.3 Skeleton and pose detection for team/rider detection

The proposed team and rider detection methodology both start from the output of a skeleton recognition algorithm (such as OpenPose [9], tf-pose [10] and AlphaPose [11]). To start, a clear definition of what is understood by the term "skeleton" should be formulated. In computer vision, a skeleton is defined as the joints (e.g., left ankle, right knee or left shoulder) that can be distinguished on a person's body by a computer vision algorithm. The family of algorithms that draws these joints on the human body are also called "pose estimation algorithms". Figure 3.3 shows an example of the skeleton detection (front and side view) of these algorithms – tested in our lab set-up. In order to measure the accuracy of each of the available pose estimation libraries, tests were performed in which ground truth annotations of the rider joints are compared to the algorithms' output. In total, 15 frontal and 14 lateral images were manually labelled to perform the accuracy study. A visual representation of a manually annotated frame (purple dot is the manually annotated joint location) together with the pose estimators' results is presented in Figure 3.4. We compare the prediction of the keypoints by the model to the ground truth data. To say that a prediction is valid we introduce an allowed deviation in pixel distance. To normalise the distance we divide it by the length of the diagonal of the bounding box of all the keypoints. Five percent is a decent threshold value to see the differences between the predictions. As can be seen in the results shown in Figure 3.5, none of these skeleton trackers is outperforming the others in all situations, but AlphaPose, Detectron2 and OpenPose are definitely outperforming tf-pose.

Figure 3.3: Rider skeleton detection (lab set-up)

The skeleton detection is providing the keypoints, i.e., the main joints of the rider's body. From the keypoint locations (i.e., pixel coordinates) we can detect the pose and orientation of the rider. If the left shoulder is left of the right shoulder, then it is most likely to be a frontal shot. If the left shoulder is on the right of

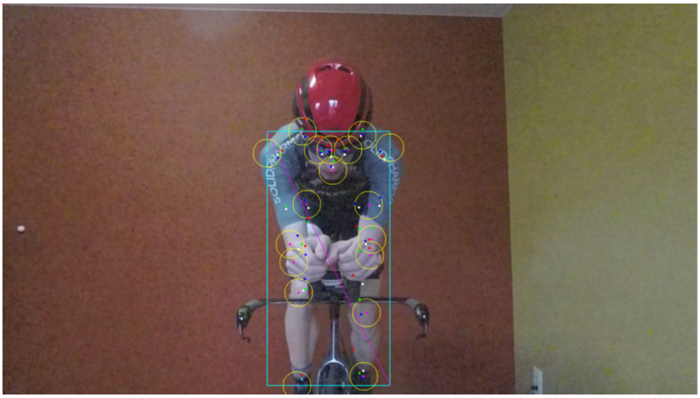


Figure 3.4: Visual representation of an image used for evaluation. Around every manually labelled keypoint (in purple) a deviation of 5% of the length of the diagonal of the bounding box is allowed to be a valid prediction (yellow circle). Other dots represent predictions of the tested pose estimators (white=detectron2, blue=openpose, green=alphapose, red=tf-pose-estimation)

the right shoulder, then the frame was most likely shot from a rear perspective. The dimensions of the skeleton with respect to the size of the video frame can also help to determine the type of shot (e.g., close-up or long shot). Based on this information, different techniques can be selected for further identification. For instance, if we see a rider from the back, face detection will not work, but a combination of number and team recognition will make it possible to detect the rider from that side. If we have a frontal view of the rider, the number will of course not be visible, but now the face recognition algorithm can take over. If available, sensor data can help to limit the number of candidate riders that can be expected in a particular shot or frame (this principle is further explained in Section 5.3).

In addition to detection of the orientation of the rider, skeleton detection can also be used for shot type classification. Based on the number of detected skeletons and their size/location in the video footage, a close-up shot can easily be distinguished from a long shot or landscape view, as is shown in Figure 3.6. Furthermore, scene changes can also be detected by analysing the skeleton size/location changes over time. As a result, we know exactly when it is safe to start and stop tracking (if needed) and we can also easily further crop the video into logical story units.

Finally, we also use the skeleton output to crop out the faces and upper body regions of the riders – results of this step are shown in Figure 3.7. In this way the accuracy of the next steps (face, number and team recognition) is improved

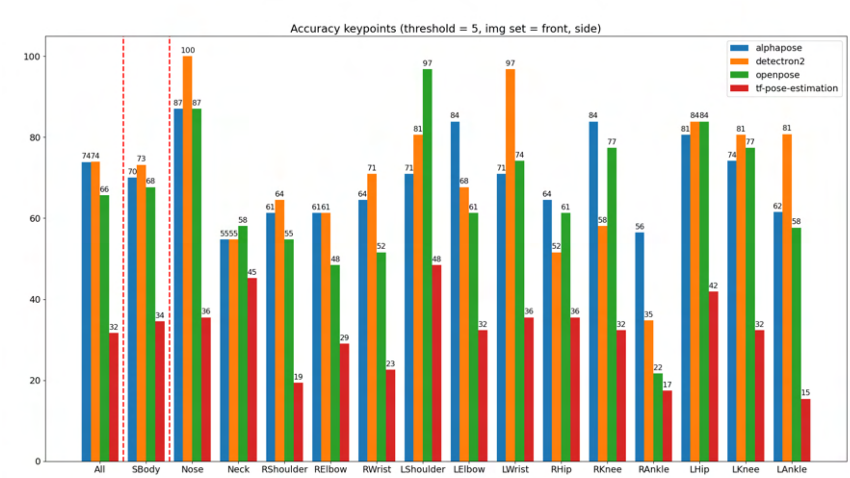


Figure 3.5: Accuracy of skeleton trackers for each of the rider's joints based on a manually labelled training dataset

Figure 3.6: Results of pose detection and shot type estimation on a Tour de France 2019 broadcast

a lot as background noise and non-rider information (e.g., numbers/text of other objects) are limited to the bare minimum.

Since the skeleton detection output is used in several of our building blocks, its higher computational cost is spread across all these steps (in our tests we get a maximum processing speed of 1 frame per second on a CPU), making it a very interesting building block that can probably be even further reused in other types of automatic sports analysis.

3.2.2.4 Team recognition

Team recognition should be capable of detecting which team a rider belongs to based on the team jerseys he/she is wearing. Team jerseys usually have distinct

Figure 3.7: Results of skeleton-based upper-body extraction

patterns with some sponsors on them. If team jerseys do not change over the years, it might be perfectly feasible to train a state-of-the-art object recognition model that has a couple of hundreds of images for each team. However, in practice, team jersey designs, sponsors, and even colours change usually every year (or sometimes even faster), which makes this approach rather unfeasible. To overcome this limitation, a methodology that only uses relatively few examples for each jersey class should be implemented. For this purpose, a transfer learning approach based on Sung et Al. was used [12]. In neural network transfer learning, the trained knowledge of an existing neural network is reused to do the classification for detection task for another unseen but related problem [13]. This is usually done by removing the last output layer of the network and adding another one instead. All weights of the original network are frozen (i.e., are not trained any further), but the last layer's weights are trained based on the (limited amount of) provided problem-specific training data. For our team classification module, we trained a RESNET18 model, with its last fully connected layer replaced by a linear layer that was retrained. In the first attempt of preparing the training data, the team jerseys were rectangularly cropped from a larger image. This introduced a lot of background noise in the image, which had a negative impact on the trained predictor's accuracy. This shortcoming was mostly solved by an extra model that crops the relevant body parts from an image with background information [14]. For our model's training data, we retrieved the torso from the humanparser's generated body part segmentation output. The model is a Resnet-101 backbone adopting the Context Embedding with Edge Perceiving (CE2P) to segment the body in the different body parts.

For our methodology, a network trained on the Pascal dataset to detect the following body parts: 'Background', 'Head', 'Torso', 'Upper Arms', 'Lower Arms', 'Upper Legs', and 'Lower Legs'. For our predictor, we are only interested in the torso as these contain the logos and the distinct team patterns. The humanparser's generated segmentation mask is used to crop out the jersey from the original image

Figure 3.8: Illustration of the cropping and masking of the torso of a cyclist.

(see Figure 3.8). This significantly improved the model's capability to accurately classify teams based on their jerseys. The F1 score on unseen validation data (different resolutions and view angles) increased from 49% to 81% with background subtraction and only the torso that was analysed.

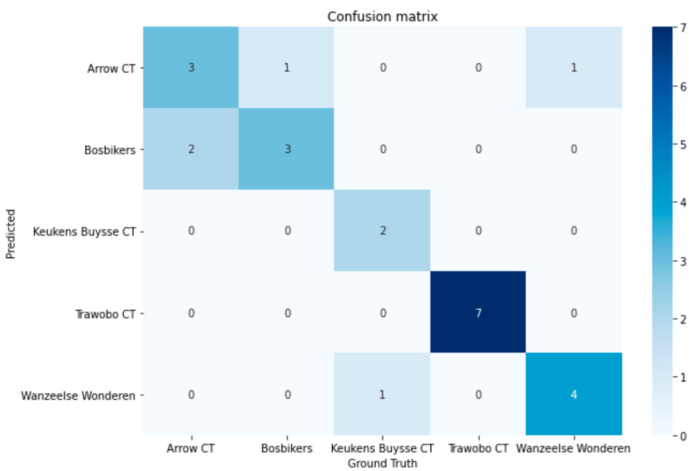


Figure 3.9: Confusion matrix of team predictions on unseen validation images trained on five team jersey classes.

To further prove the effectiveness and its relative easiness to train with few sample images, a model was trained on five teams (see Figure 3.10 for an overview of the corresponding jerseys). For each team, a total of 8 images were used to train the final output layer of the RESNET18 model. The training images were pre-processed using a set of (random) image transformations randomly performing light condition changes, horizontal flips, slight rotations and/or cropping. The other 3 images were used for testing purposes. The model achieved a 96% validation accuracy after 10 epochs of training time. The model was further validated on a number of unseen images for each team. The confusion matrix of this extra validation data is shown in Figure 3.9. As could be seen, some teams are still mistaken one for another, but the provided shots are from multiple camera angles and zoom levels, so additional more clever cropping might still improve the model's prediction. This experiment with the five sample teams shows the effectiveness of this approach for the specific cycling use case. An additional test was performed with another four additional teams and a similar confusion matrix was obtained, which shows that the methodology is scaleable to more teams. As team jerseys change at least once a season, the solution is required to be easily trainable and retrainable. The experiments show that good results can be achieved with relatively few training samples. In the future work section, we will discuss other possible improvements to make the model better performing and adoptable.

Figure 3.10: Overview of the five team jerseys that were used to train the team jersey recognition model.

3.2.3 Cyclocross line choice and ride mode detection

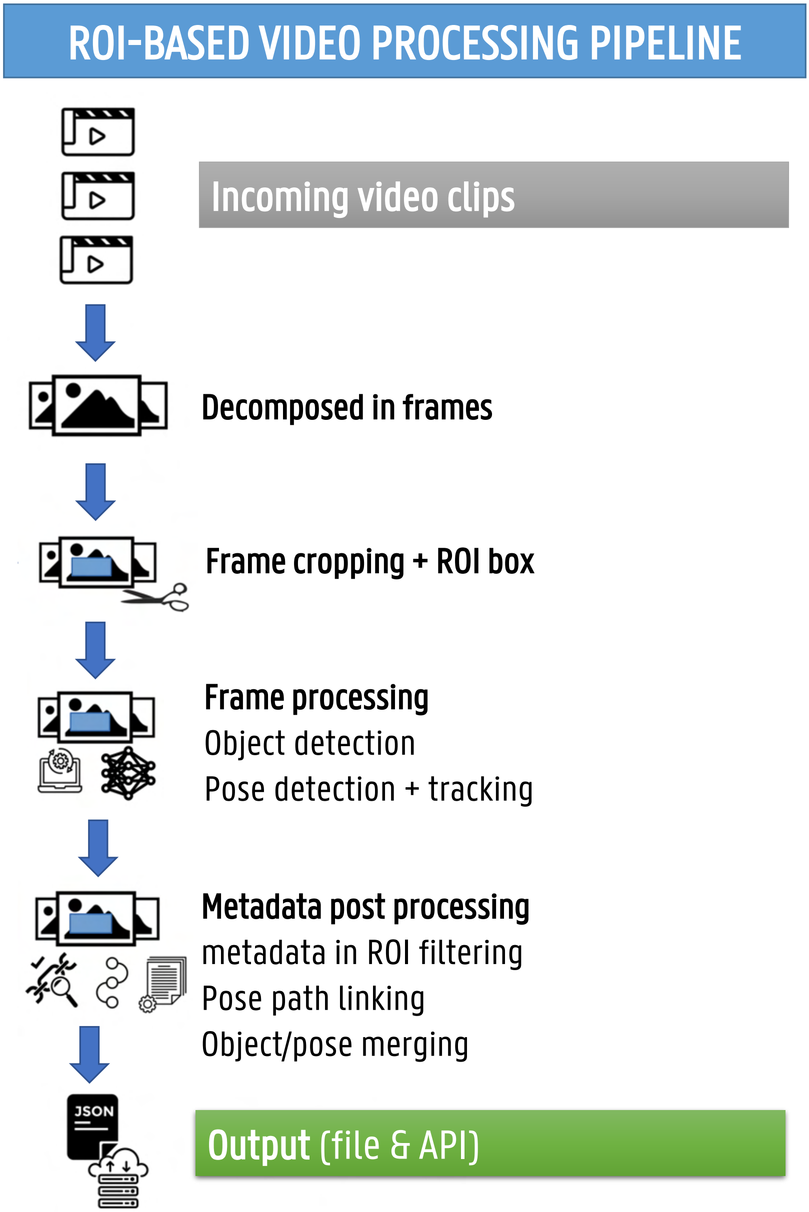


Figure 3.11: Different steps in the video processing pipeline to produce line choice data.

3.2.3.1 About the experiment

In cyclocross, riders ride or run with their bikes, based on the technicality and surface conditions. In the next step, a clear distinction between each of the possible rider modi is made. Technical sectors might, for instance, be perfectly rideable for a rider with great technical prowess but might be completely unrideable for another less technical rider. The barriers are a great example of such a technical sector. If the barriers are relatively high and are placed at a challenging part of the course (e.g., uphill or after a corner), some less technical riders will be forced to dismount their bike and run over this course feature. Monitoring these differences among riders within the video region-of-interest can be very valuable to help understand why, how, and where riders are taking a certain line and to explain why one rider is slower than another. To answer the riding mode question, a YOLOv5 neural network to detect running or riding cyclists was trained (see section 3.2.3.5). This knowledge is attached as metadata to the previously detected poses and its tracked pose identifiers.

3.2.3.2 Video capturing process

The first step of the analysis consists of the decomposition and initial preprocessing of the incoming video source. The pipeline accepts both recorded video clips and live video data delivered by popular streaming formats, such as HTTP Live Streaming (HLS) and the Newtek Network Device Interface (NDI) protocols. The video data is decomposed into frames and based on the frame rate of the recorded video, some frames are periodically skipped for an optimal balance between accuracy and processing time. Experiments with high-definition footage (1920 × 1080 pixels) at a frame rate of 30 frames per second show that down-sampling and processing every third frame gives the best balance between pipeline detection accuracy and processing speed (tested on an Intel i7-10700F processor, 32 GB RAM, Nvidia RTX 2060 Super GPU). By processing every third frame, the video is processed in (near) real time.

3.2.3.3 The region-of-interest principle

For pre-processing, the frame can be cropped to leave out irrelevant background information for further analysis (and further speed up processing times). For now, this is a manual procedure where the desired analysis region is selected before the analyses, but this could also be automated by techniques such as background subtraction or by optical flow algorithms. Next, a rectangular region-of-interest (ROI) is defined within the cropped region. The ROI is defined as the region of the video in which movements and the behaviour of riders are analysed and is illustrated by the measurement zone rectangle in Figure 3.12. The difference between the cropped area and the ROI/measurement zone is that subtle. The presented detectors and trackers are used on the entire cropped area, but only measurements (e.g., time in zone, ride modi in zone or line choice in zone) are considered within the defined ROI.

Figure 3.12: Illustration of the ROI principle on a video frame of a cyclocross training session. The red rectangle is the region of interest. Results of pose detection, tracking and post processing are illustrated by the yellow path within the ROI.

3.2.3.4 Pose estimation

In 3.2.2.3, pose detection methodologies for a cycling specific context were introduced. For our cyclocross video processing pipeline, an Alphapose pose estimator [11] is run on the frame to detect the riders and their body key points. To track riders through the frames, Alphapose offers various pose tracking implementations (e.g., PoseFlow, Human ReID or detector based). However, after thorough experimentation with the different trackers and its parameter configurations we were not able to achieve satisfiable results. The trackers work great on pedestrians, but on skeletons that are pedalling a bicycle, the tracking often fails. Our finding that traditional tracking algorithms make considerable mistakes in a sports context is also fortified in the literature [15] [16]. The tracker can fail in two different ways: first, in a tracking identifier swap, especially after partial occlusion of one skeleton behind the other. Another tracking failure occurs when a new tracking identifier is assigned to an already seen skeleton. However, for further path analysis we decided to position the camera in such a way that it films the region-of-interest from not a frontal but an overhead camera angle. This has the advantage that the selected ROI more accurately represents the real-life coordinates and that skeleton swaps are less likely to occur as there is a better unobstructed view of the riders (e.g., riders will not be hidden behind each other).

With this extra prerequisite in mind, we implemented a more straightforward yet purpose tailored spatio-temporally aware tracking mechanism that mostly circumvents the mentioned shortcomings of the trackers included within Alphapose. Full details of the tracking methodology can be found in Algorithm 3.1, but we will briefly discuss the main working principles of the technique. The technique keeps track of the skeletons seen in the last five frames with its last known coordinates within the ROI.
When a new frame is processed, the distance matrix between the old poses' centre locations and the new pose centres is calculated. The new pose matches with an older pose if it has the minimum distance to that old pose and the distance is smaller than 25% of its diagonal size of the bounding box around the new pose (see Figure 3.13 for an illustration of this approach). This percentage is a good starting point and can be tweaked depending on the course and segment type. For an extreme uphill sector, for instance, we can reduce this value as the riders are moving slower through the image.
Each time a new frame is processed, the poses older than five frames ago are also removed from the pose match dictionary.

This approach works very well in cycling as cyclists travel from a starting to an end point within the ROI (i.e., they never go back in the ROI), so the corresponding bounding boxes are also moving similarly through the ROI over time.

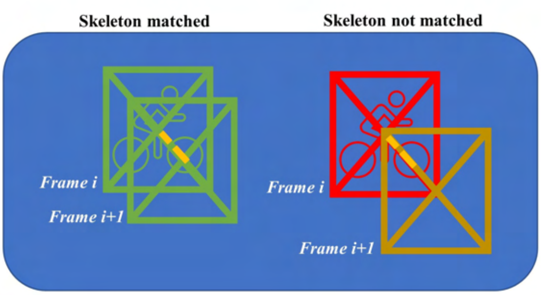


Figure 3.13: Illustration of the pose matching methodology. The yellow dotted line is the 25% threshold of the diagonal size. The poses on the left are matched between frame i and i + 1 as its distances between bounding boxes are less than 25% of the diagonal size, and the right poses are too far away to be matched.

Algorithm 3.1 Custom skeleton tracking methodology

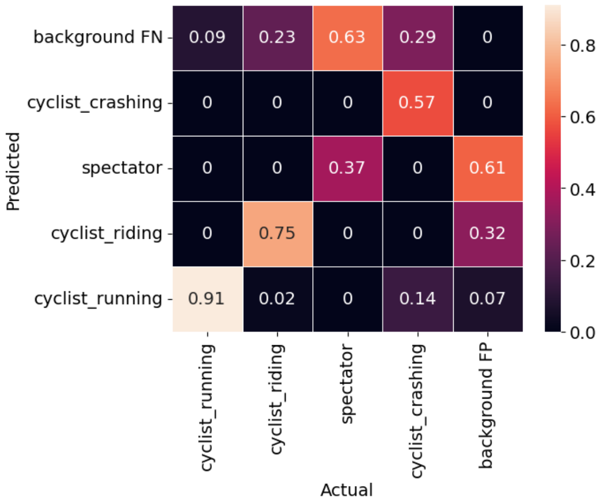
```
lastId ← 0                      ▷ When a new skeleton is found it gets last id + 1 as its id
trackedSkeletons ← []           ▷ Skeletons that were recently seen in video
nFramesInHistory ← 5

for i ← 0 to length(videoFrames) do
    frameNr ← videoFrames[i]
    poses ← frameResults[frameNr]
    poseCenterXY ← [[x,y] for pose['coordinates'] in poses]
    poseIds, trackedSkeletons ← mapPoses(poseCenterXY, trackedSkeletons)
    trackedSkeletons ← cleanupOldPoses(trackedSkeletons)
    ▷ Distance between tracked and new pose skeletons (omitted for readability)
    distMatrix ← buildDistanceMatrix(trackedSkeletons, poseCenterXY)
    usedPoses ← []
    mappedIds ← [null] × length(poses)
    for trackedKey in distMatrix do
        minIndex ← index of minimum of distMatrix[trackedKey]
        minDistance ← value of minimum of distMatrix[trackedKey]
        distThrehsold ← 25% of diagonal length of pose at minIndex
        if minIndex not in usedPoses and minDistance < distThreshold then
            trackedSkeletons[trackedKey]['lastSeen'] ← frameNr
            trackedSkeletons[trackedKey]['coordinates'] ← poseCenterXY[minIndex]
            append minIndex to usedPoses
            mappedIds[minIndex] ← trackedSkeletons[trackedKey]['idx']
    notUsedPoses ← indices of poses not in usedPoses
    for notUsedPoseId in notUsedPoses do
        newIndex ← lastId
        trackedSkeletons[newIndex] ← {
          'lastSeen': frameNr,
          'coordinates': poses[notUsedPoseId]['coordinates'],
          'idx': newIndex
          }
        mappedIds[notUsedPoseId] = newIndex
        lastId ← lastId + 1
    ▷ Add mappedIds to poses data structure here (omitted for readability)
```

3.2.3.5 Ride modus detector

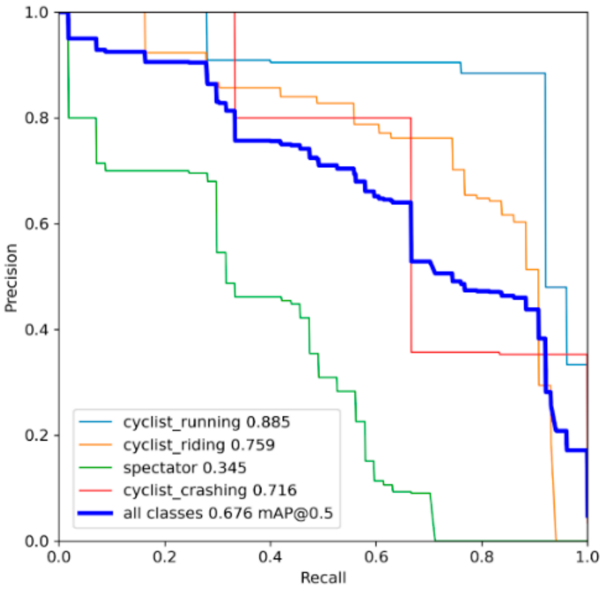
To train the ride modus detection model, a training dataset of 869 images was constructed, with 747 cyclists that are riding, 457 running, 116 crashing, and 1038 spectators. The dataset was split uniformly across the categories in 75% for the training data, 20% for test, and 5% for validation. Frames of videos were captured from different cyclocross races, under varying circumstances in the 2020 season. The data was used to train a YoloV5 model (yolov5s variant) for 100 epochs and achieved a mean average precision (mAP) of 68%. As can be seen in Figure 3.14a, it is the spectator class that is degrading the model's overall performance quite a bit as it classifies most spectators as background. This is not really a problem for this proposed methodology as we are only interested in riders and the confusion matrix shows us that there is no confusion between riders (either running or riding) and spectators. Furthermore, not all spectators were consistently in every frame that was used for training so this can also confuse the model and its evaluation. The real power of this detector is when it is used on a video sequence. In combination with the tracking results of the rider skeletons, the rider modi can be tracked over time. Figure 3.15 shows an illustration of the ride mode model that was run on a sequence of a men's pro cyclocross race. The ride mode detector ran on each frame of this sequence and the color-coded arrows show the history for that rider across the video sequence. The combination of the video frame rate and the ROI analysis results can give an indication of the total time required to finish the sand sequence and how they did it (i.e., riding and running ratio).

3.2.3.6 The final product: from pipeline elements to a methodology

The previously introduced elements in the cyclocross ride line methodology can now be combined into a workflow to detect riders in certain zones on the course. For this purpose, the analyses across subsequent frames have to be synthesised into a "run". If we want to perform real-time path analysis the post processing is initiated whenever the ROI remains empty for 10 consecutively analysed frames. In post-processing the actual paths travelled by the tracked skeletons are determined. If desired, the coordinates of the paths can also be transformed into real life coordinates using a homography perspective transformation [17]. The main challenge in this post-processing step is the handling of the re-identification of the pose tracker. A pose is re-identified if the pose tracker assigns a new tracking identifier to a pose that was actually already seen in a previous frame. The possible causes for re-identification can be usually reduced to two different categories. The first takes place when the pose estimator does not succeed to map the skeletons for one or more consecutive frames, which causes a jump in the subsequent positions which is too high for our simple geospatial pose tracker to link it



(a) Confusion matrix



(b) Precision-Recall curve

Figure 3.14: Training output of the cyclocross riding mode detector

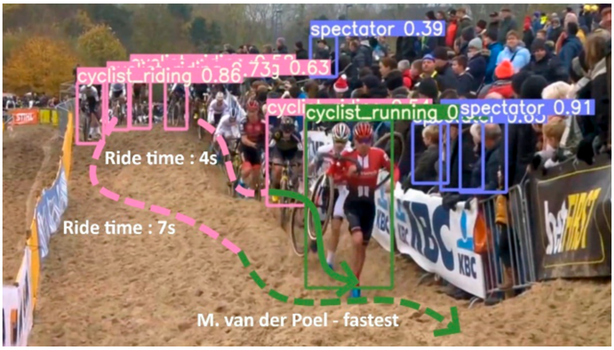


Figure 3.15: Ride mode detector run on a video extract of the World Cup race of Koksijde. The arrows show the history of the detected ride modi for that specific rider.

to a previous pose instance. Another culprit is when two poses are basically overlapping each other and both identifiers are mistaken for each other. With these limitations in mind, a post processing strategy can now be implemented to search and solve the aforementioned re-identifications. The strategy exploits the fact that in our ROI-based approach the skeletons will always travel in a consistent direction within the ROI (e.g. left to right, right to left, top to bottom or bottom to top). With this added constraint, a straightforward, yet powerful geospatially aware pose path merger can be implemented. The merging process is illustrated in Figure 3.16 and will be briefly discussed in the next paragraph.

In summary, the merging process consists of three steps. In the first step, the paths are split based on jumps in frame numbers of the different tracks. A track is defined as a pose that was tracked over time in the ROI and was assigned a tracking identifier. The criterion to split a track is that tracks that have a non subsequent frame sequence are split into two separate tracks. This prepares the tracks for step two of the merging process where the tracks are attempted to be merged again based on a spatiotemporal weighting function. As illustrated in Figure 3.16, a track has a number of match candidates that can be matched. The selected candidate is the one with minimum spatiotemporal distance and below a certain threshold that is set based on the ROI's dimensions. In the last step, the spatiotemporally linked tracks are iteratively matched based on the index lists of frame numbers within a track. This process stops if all paths have index lists that are non-overlapping.

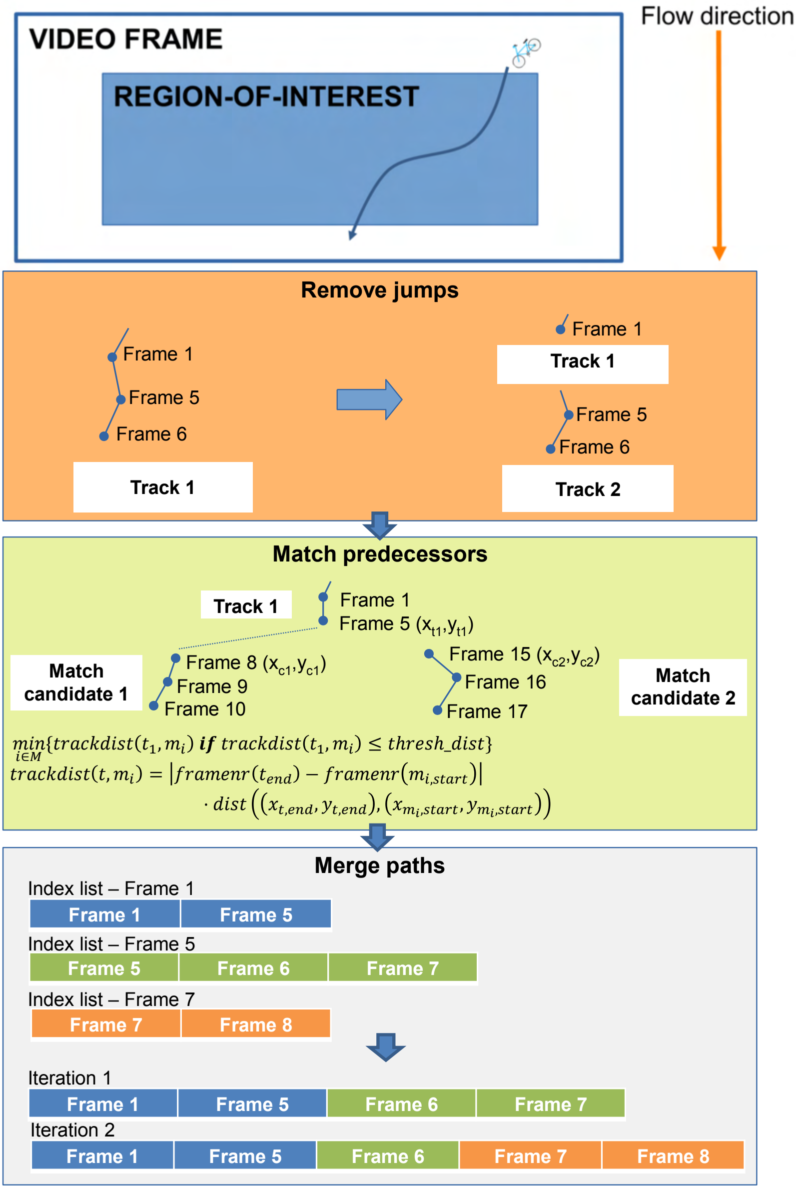


Figure 3.16: A schematic overview of the three-step pose tracking merging strategy.

The proposed pose tracking methodology combined with the merging strategy produces a set of pose tracks that can be directly used in the following steps in the video pipeline. In this next step the pose tracks (or parts of the tracks) that are within the boundaries of the desired ROI's coordinates are extracted. To check if a skeleton of a path is within the ROI, the centre of the bounding box around its joints was used. Using this approach, the tracked skeletons' coordinates stay much more consistent and less spiky as when a joint such as the knee or foot were used. The skeleton is considered within the ROI if its centre is within the ROI's bounding box (see Figure 3.17). This check is performed for every frame in which the skeleton in the track was detected. A valid path in the ROI is defined as a rider that is entering and exiting the defined bounding box (and has multiple detections within the ROI).

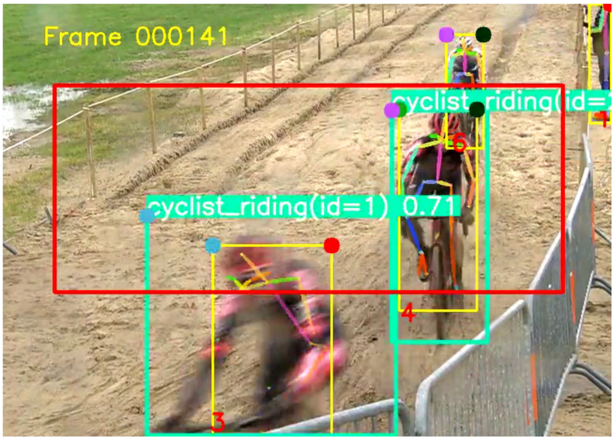


Figure 3.17: Rider entering (id = 6) and rider exiting (id = 3) the sand pit region-of-interest (red rectangle)

Once the valid paths have been detected, all information is available to create insightful statistics of the path a rider did follow within the ROI. The extra metadata such as the rider modi and the team jersey recognition results are used to annotate the path with a major riding mode and the team probability scores of the rider. The duration of the video clip from entry to exit of the region of interest equals the time spent in that zone. A first application is the direct analysis of the riding lines within the ROI. Figure 3.18 shows how this data can be visualised. As mentioned in the previous section, the coordinates of the poses were first mapped on real-life coordinates to optimally represent the true shape of the ROI and paths that were followed. In the ride line graph presented in Figure 3.18 we can see that the blue and orange rider tend to better follow the flow of the course than the green rider (i.e., the course curves to the right). This finding is further fortified by the screenshots of the actual video footage in Figure 3.18. The green rider made a technical mistake in the sand, causing a deviation from the other riders' lines, which were more logical when the flow of the course is considered.

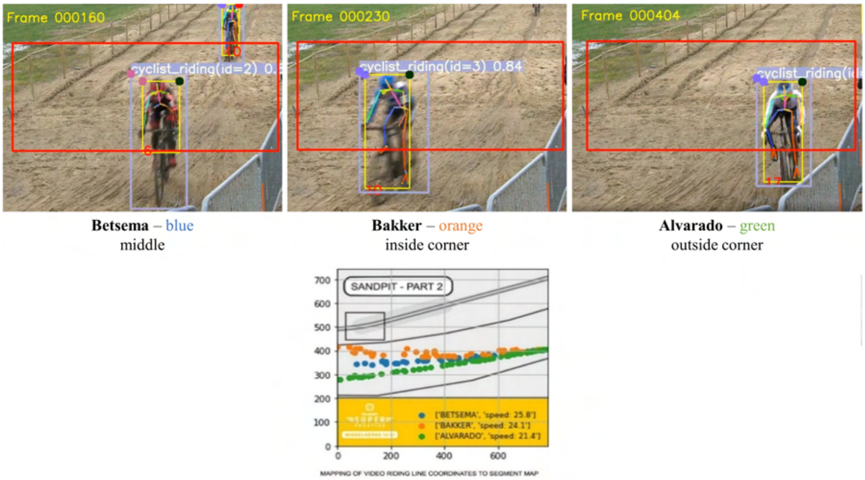


Figure 3.18: Three female riders riding different lines through the sandpit in a professional cyclocross race. The red numbers are the tracking identifiers of the pose tracker. Schematic overview of the produced ROI paths of the various riders that went through the ROI (riders go through the ROI from the right to the left).

As a final step in the processing pipeline, the raw ride line data is also published to an API, so this information could also be used by video broadcasters to directly incorporate these near-real-time stats in the live video feed (or in race summaries or recaps afterwards).

3.3 Conclusion

In conclusion, we can summarise that video plays an important role within professional cycling. Cycling is a sport that is greatly televised with a massive archive of recorded cycling broadcasts. Video images can contribute in all of the focus areas (performance, storytelling and safety) of this thesis. We presented an overview of how races are broadcasted and elaborated on the key techniques that can be applied in cycling specific video analytics. In Chapter 6, when we tackle multimodal analysis, it will become obvious that video analysis plays a key role in the overall analysis of the selected use cases, especially when it is enriched with other types of data.

References

[1] J. Lombardo and D. Broughton. *Going gray: Sports TV viewers skew older*. Sports Business Journal, pages 1–8, 2017.

[2] D. Van Reeth and R. Simmons. *TV demand for the Tour de France: the importance of stage characteristics versus outcome uncertainty, patriotism, and doping*, 2013-02-01.

[3] S. Verstockt, B. Van Vooren, S. De Smul, J. De Bock, et al. *Data-driven summarization of broadcasted cycling races by automatic team and rider recognition*. In icSPORTS 2020, the 8th International Conference on Sport Sciences Research and Technology Support, pages 13–21. SCITEPRESS, 2020.

[4] J. De Bock and S. Verstockt. *Video-based analysis and reporting of riding behavior in cyclocross segments*. Sensors, 21(22):7619, 2021.

[5] H. Liu and B. Bhanu. *Pose-guided R-CNN for jersey number recognition in sports*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 0–0, 2019.

[6] Z. Mahmood, T. Ali, S. Khattak, L. Hasan, and S. U. Khan. *Automatic player detection and identification for sports entertainment applications*. Pattern Analysis and Applications, 18(4):971–982, 2015.

[7] S. Langer, R. Müller, K. Schmid, and C. Linnhoff-Popien. *Difficulty Classification of Mountainbike Downhill Trails utilizing Deep Neural Networks*. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 270–280. Springer, 2019.

[8] S. Verstockt, V. Slavkovikj, P. De Potter, and R. Van de Walle. *Collaborative Bike Sensing for Automatic Geographic Enrichment: Geoannotation of road\/terrain type by multimodal bike sensing*. IEEE Signal Processing Magazine, 31(5):101–111, 2014.

[9] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. *OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.

[10] W. Mao, Y. Ge, C. Shen, Z. Tian, X. Wang, and Z. Wang. *TFPose: Direct Human Pose Estimation with Transformers*, 2021. Available from: https://arxiv.org/abs/2103.15320, doi:10.48550/ARXIV.2103.15320.

[11] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu. *RMPE: Regional Multi-person Pose Estimation.* In ICCV, 2017.

[12] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales. *Learning to compare: Relation network for few-shot learning.* In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1199–1208, 2018.

[13] S. Tammina. *Transfer learning using vgg-16 with deep convolutional neural network for classifying images.* International Journal of Scientific and Research Publications (IJSRP), 9(10):143–150, 2019.

[14] P. Li, Y. Xu, Y. Wei, and Y. Yang. *Self-correction for human parsing.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

[15] L. Bridgeman, M. Volino, J.-Y. Guillemaut, and A. Hilton. *Multi-person 3d pose estimation and tracking in sports.* In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pages 0–0, 2019.

[16] N. Seethapathi, S. Wang, R. Saluja, G. Blohm, and K. P. Kording. *Movement science needs different pose tracking algorithms.* arXiv preprint arXiv:1907.10226, 2019.

[17] E. Malis and M. Vargas. *Deeper understanding of the homography decomposition for vision-based control.* PhD thesis, INRIA, 2007.

4 Text analysis in cycling

"Sapere aude."

– Horatius (Epist. I,2,40)

This chapter discusses the performed text analysis within cycling. We provide an overview of state-of-the-art natural language processing (NLP) and how these methodologies were used to extract meaningful information from "unstructured" text data in sports and cycling. The research question can be formulated as follows: **"Is the current state-of-the-art NLP able to extract structural race specific information (e.g., crash or breakaway), rider specific information (e.g., rider name, team or sustained injury) from Tweets?"**.

4.1 What is text analysis?

Text is defined as a collection of words that provide meaning and understanding about a certain topic and can have a certain sentiment or nuance in it. In this chapter we will provide an introduction on how text can be understood and processed by computers. We will present the major text processing techniques prevalent and apply them on our professional cycling data use cases.

As mentioned in Chapter 2, text can be either structured or unstructured. Structured data is the easiest data to use for further (computer) processing. Unstructured data needs more pre-processing. The best example of unstructured text is our natural language, which is the language that humans use to communicate, is nothing but a collection of words that have a certain meaning (provided if its structure, spelling and grammar is correct). The individual words have a definition that can be searched in a dictionary, but it is the combination of words that gives sense to a sentence. This can be perfectly illustrated by the word "race". Depending on the context of the sentence, it can be used either as a sports events (i.e., a bicycle race) or as a kind of biological taxonomy (e.g., a race of a dog). Understanding these nuances in natural language is relatively straightforward for human beings, but is way harder for a computer program. Extracting individual words and punctuation marks is relatively straightforward, but getting an idea of the exact content and meaning of a phrase is much less trivial. In an interesting article about the power of unstructured corporate data, Harbert Tam cites that 80 percent of the data in a company might be unstructured [1]. Luckily, throughout the years, computer aided text processing has made huge leaps forward, and several great techniques exist to get meaning from natural text.

4.2 Text retrieval

The first step to start analysing textual data is by getting it ready and available for our computer algorithms. This process is often referred to as "text retrieval". Getting the actual information might be either extremely convenient or very challenging. In this section we will discuss the major methodologies that exist to get textual data ready for analysis.

4.2.1 Databases and/or Application Programming Interfaces (APIs)

As mentioned in the introduction of this chapter, structured data is (textual) data that has a repetitive nature. This data is very easy to process by computer programs if it is stored in a convenient manner. Storage can be either file-based or in a data management system (i.e., database). Often these entries comply to a data model and each entry has a unique identifier.

The modernisation of storing and managing data online has found its way nicely into professional cycling. The international cycling federation (UCI), for instance, has all of its races and its results digitised and made publicly available through their website and via an application programming interface (API). In Figure 4.1 the architecture of a Representational State Transfer API (REST API), one of the more popular API mechanisms, is presented.

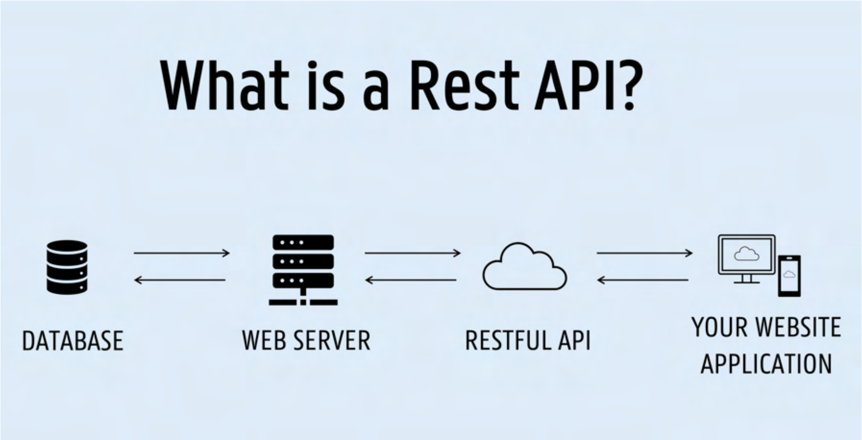


Figure 4.1: Schematic overview of how the REST API principle works.

Furthermore, races became also more digitised through the years. Gone are the days that race officials could only rely on pen and paper and a very basic video camera to record finish line crossings. Fast forward to 2023, when timing chips are used in most of the professional races (in Flanders, for instance, the timing chip even found its way into amateur and youth racing as well). Although the system itself is relatively simple, it provides a very solid, automatically recorded and accurate answer to the classic finish line crossing problem. The MyLaps system, for instance, is used by the Flemish cycling federation on the track and in most of their cyclocross, road, mountainbike and BMX races. The system relies on riders wearing an actively powered RFID timing chip that gets picked up by detection antennas positioned along the course (these can be in mats on the course or by loops stuck to the tarmac). This principle is schematically illustrated in Figure 4.2. By gathering this information, and by linking antenna identifiers with riders, an accurate “gate” crossing is provided to the race officials. According to their website they can provide accuracy up to 0.003 seconds [1]. The data can be made available online on their data collection and distribution platform Sporthive. This data can be either easily be downloaded from MyLaps’ web applications, but is also directly accessible on the network on which the decoders are installed. Multiple decoders can be installed on a single network, allowing for intermediate timings between measurement loops. As this is a timing system, the decoders should then agree

[1] https://www.mylaps.com/timing-solutions-active/prochip/prochip-timer/

on the time. Current MyLaps decoders do not use Network Time Protocol (NTP) servers to sync, but rather rely on an additional GPS timing module to do the synchronisation of the time.

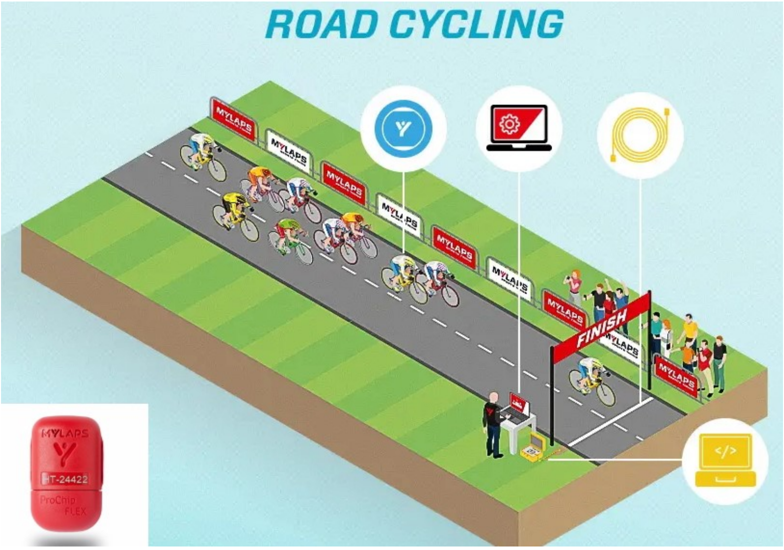


Figure 4.2: Schematic overview of the MyLaps timing setup in road cycling context and an example of a MyLaps timing chip (left bottom) (retrieved from: https://www.mylaps.com/active-sports/road-cycling/)

4.2.2 Scraping

As mentioned in the introduction of this chapter, eighty percent of the data in a company might be unstructured. A major part of this data might be “hidden” in the HyperText Markup Language (HTML) of the web pages on the world wide Web. Although, strictly speaking, HTML data is structured (i.e., it uses XML as storage mechanism) a lot of the information is unstructured as the pages often contain long sentences and text blocks. The Indeed Editorial team describes web scraping nicely as “a software tool that can access data on any public website and extract and copy data from it” [2]. A scraper is in fact a very broad term for a variety of different applications and use cases. A major consideration that has to be made when creating scrapers is its legality. As a general rule of thumb: web scraping is legal as long as the data is publicly available on the internet and/or if a website allows it. The latter is usually checked by looking if and what is in a website’s `robots.txt` file. Although mystified by legal confusion, scraping is often the only feasible manner to get hold of data for further analysis.

4.2.3 Image-to-text

Sometimes the information we want to retrieve might be concealed in video. As an example, it is not uncommon that the road is painted full of rider names when riders ride a famous climb in France. Another example are the barriers that usually contain publicity banners of the sponsor. In Figure 4.3, for instance, we can see Bob Jungels' name painted on the famous slopes of the "La Redoute" climb in Liège, Belgium.

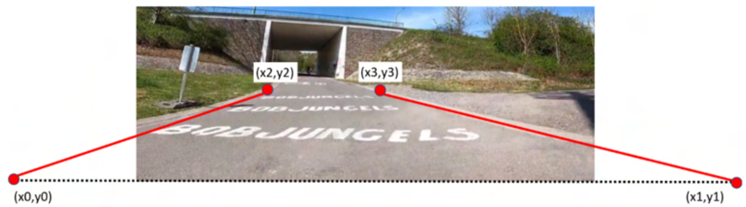


Figure 4.3: Example of Bob Jungels' name painted on the "La Redoute" climb. The coordinates for the homography are also annotated on the image. The coordinates are roughly representing the road boundaries.

In this subsection we introduce a solution to automatically extract the aforementioned texts on the road from dashcam-like recordings.

To achieve this, a pre-selection process was implemented in order to limit the images provided to the model. The goal of this pre-processing step is to only analyse the images that are more likely to contain text on the road. This pre-selection is broken down into two steps. For both of them, the assumption that the text is written in lighter colour than the road was made. This is almost always the case (i.e., white paint on grey asphalt). The first step is an intensity-based method to determine if there is any light colour present in the image. To do this, an adaptive threshold based on the brightness of the image was set (to cope with shadows). The thresholding operation creates a binary mask and the number of 1's are counted. If this count exceeds a certain threshold (1% of the total number of pixels in the image), it is assumed that potentially text is present in the frame. If an image is accepted by the first check, it has to pass through a second, more selective process, in order to determine whether to send it to the model. In order to focus only on the road for the second step, semantic segmentation was performed on the image using the DeepLabV3+ model [3] trained on the cityscapes dataset [4]. The segmented road is morphologically filtered (closing after opening) and only the largest road blob is kept to remove outliers.

To perform this check, a histogram of the segmented image is made. Next, the peaks of this histogram are analysed, considering that the highest peak corresponds to the road (it is assumed to be the most present intensity because we use the segmented image of the road). Thus, if there is another peak (i.e., another local maximum) that is brighter than the road, then it is hypothesised that there is potentially white, i.e., text.

These two selection steps allow to filter the input images and keep only a part of them. However the presence of shadows or white paint lines for traffic disrupts these checks, which accept images that do not contain text and slow down the overall system performance.

In order to make the text painted on the road readable by an OCR model, the main idea is to use a homography on the image to modify the perspective of the road. If the untransformed image of the road would be used, the OCR models will almost certainly fail as they are trained on printed text that is usually perfectly straight in one line (see Figure 4.5 for an illustration of the OCR methodology failing when images were not transformed). To solve this "issue" we first have to rectify the painting on the road surface. This is performed by a transformation and is parameterised by minimum 4 points (x_i , y_i), where $i \in [0, 3]$ whose purpose is to describe the position of the road with respect to the camera that captured the footage (see Figure 4.3 for those points annotated on the footage). Please note that this homography transformation does not result in a perfectly straight line of text. This is because some assumptions of the real-world have been made. Firstly, the road is not perfectly flat and straight and secondly, the paint on the road surface is often also skewed and/or not perfectly straight. The camera was mounted on the windshield of a car or the handlebars of a bicycle that rode the course, so its position (with respect to the road) remains constant (suspension will create relative movement between the plane defined by the wheels of the car/bicycle and the dashboard/handlebars and will be the worst while breaking, accelerating and taking corners). In order to optimise the transformation in terms of memory and computation time, the transformation was limited to the area enclosed by these points. The position of the camera is not varying that much with respect to the road, so a fixed homography can be defined. However, in order to optimise and fine-tune the transformation, an accurate estimation is made based on the position of the four points from the segmentation of the road used for the pre-selection of the images. The points (x_0, y_0) and (x_1, y_1) are placed at the intersection of the lateral lines delimited by the road segmentation and the line defining the bottom of the image, in order not to lose information on the text written at the bottom of the image. Thanks to the homography, the names appear readable in the transformed image, facilitating the task of text recognition for an

OCR model. This is illustrated in Figure 4.4

Figure 4.4: Homography transformation being applied to the "La Redoute" painted text. The painted text is now easily readable.

Once an image was selected and processed, it is provided as input of the text detection and recognition model. Several OCR models were compared, such as EasyOCR [5], Tesseract [6] and Keras-OCR [7]. Ultimately, the EasyOCR model was chosen because the technique offers a suitable trade-off between the recognition of the text painted on the road's surface and the computation time. It takes 0.4 seconds per frame on the Google Colab cloud notebook, and about 25 seconds on a Raspberry Pi model 4B. The model was not further fine-tuned because of limited data and to avoid overfitting. On Figure 4.5 (left side) it is illustrated that without the transformation, the model cannot read the text due to the perspective, but after applying the pre-processing pipeline, the OCR model can correctly read the text painted on the road surface, as can be seen in Figure 4.5 (right side). Before the preprocessing, the model detected only one text: "YEd" and "Siin" whereas with the geometric transformation, it detects "VLIEGEN" and "YOMAMTIME", which is more similar to the expected results. Finally, the predictions are associated with the cyclists' names available in either the UCI's members or race's participants list. To compare similarity, the Levenshtein distance [8] is used, which calculates a distance between two strings. The five most similar names are kept in memory and compared with the results of the last ten predictions to prevent outliers. As a result, predictions of Figure 4.5 "VLIEGEN" and "YOMAMTIME" are respectively associated to "Loïc Vliegen" and "Eri Yonamine", which is correct.

In order to make our solution work in real time on the setup, the computation time needs to be reduced significantly. The current implementation of the algo-

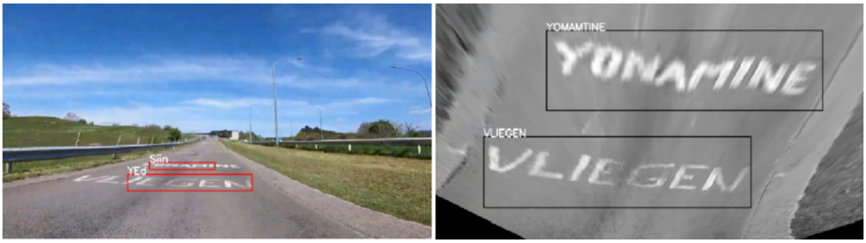


Figure 4.5: Positive improvement of the post processing (homography) on the EasyOCR text recognition (left the non transformed and right the transformed image).

rithm takes about 25 seconds per image (tested on a Raspberry Pi with camera module attached), which makes it impossible to use it in real time, so the number of images processed by the algorithm needs to be reduced significantly.

As a test, the methodology was employed on the Planche des Belles Filles climb, in France's Vosges, and often featured in the Tour de France. The video was recorded from a bicycle's handlebars with a Garmin Virb Ultra 30 action camera that has GPS metadata embedded in the recorded videos. This implicates that frames can be linked to a geographical location. The video was analysed at 3 fps, for 5,295 frames, corresponding to 29 minutes and 25 seconds. The image pre-selection deleted 1,536 images, which represent 29% of the frames. By analysing images that passed the checks, we could observe that his number could have been much lower, but in the most part of the video road markers are painted which confused our image pre-selection. For the results of the text recognition, Figure 4.6 illustrates that the model detected areas where cyclist's names were painted on the road. It detected "Thibaut Pinot" 42 times and "Alex Dowsett" 14 times. Some other names were detected but only the two most common cyclists were plotted.

4.3 Natural Language Processing

Natural Language Processing (NLP) is a research area within computer science that tries to understand spoken and written language in a similar way as humans do. As mentioned previously, on the first hand, this might seem trivial for human beings, but for computers it is actually a hard task [9].

Luckily, the research with respect to natural language processing is continuously evolving and some interesting and helpful models and methodologies exist to understand the context of the sentences. In the following subsections we will further focus on some of the major NLP building blocks and show its usefulness for a cycling specific context. In the Chapter about multi-modal analysis (Ch. 6),

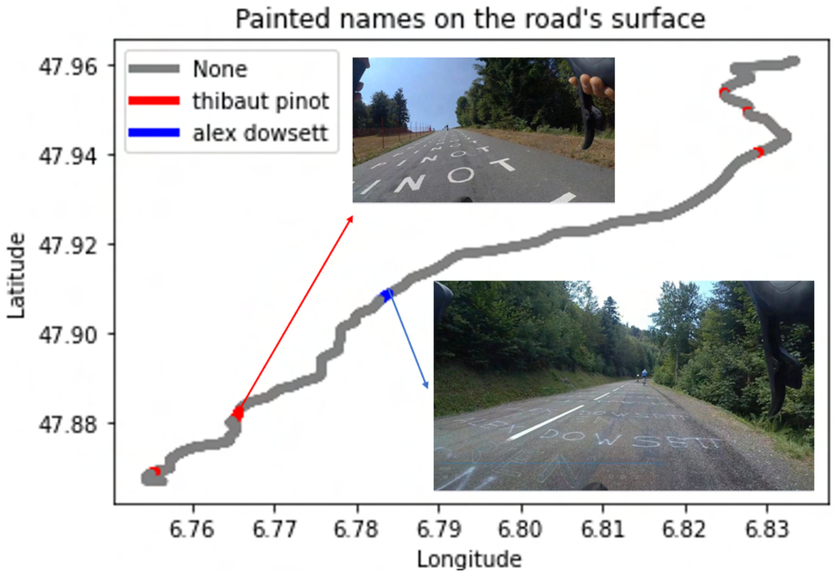


Figure 4.6: Painted names on the road's surface during a race on Planche des Belles Filles

we will further elaborate on how we used NLP methodologies within the cycling safety use case. The key principles within NLP will be exemplified with the rather "simple" sentence about probably our best known Belgian cyclist, Eddy Merckx. The sentence we will use is: *"Eddy Merckx, a famous Belgian cyclist, won the Tour de France five times."*. For the various NLP analyses we will use the well-known Spacy Python package. As it is mentioned on their web page, this package is promoted to have "industrial-strength natural language processing".

4.3.1 Tokenisation and part-of-speech tagging

In Part of Speech (POS) tagging, every word (including punctuation) is categorised as noun, proposition, verb, etc. This is illustrated in Table 4.1, where we use our "Eddy Merckx-sentence" in conjunction with the Spacy tokeniser. Apart from separating the sentence in different parts, the package provides a lot more insights. The 'position' attribute (Table 4.1), for instance, is the Universal Position (UPOS) of the word. These are considered to be a universal naming convention for part-of-speech categories. In Table 4.2 an overview of the universal names of the different tags is provided. The tag column in Table 4.1 is an extra refinement of the UPOS of the token. The dependency column is probably the most abstract, but it shows the syntactic dependency, i.e., the relation between tokens. Figure 4.7 might help to understand these dependencies that were being made between the parts of the sentence.

Table 4.1: The POS tagging performed by the Spacy library on our example sentence.

i	text	lemma	upos	tag	dependency	shape	is_alpha	is_stop
0	Eddy	Eddy	PROPN	NNP	compound	Xxxx	True	False
1	Merckx	Merckx	PROPN	NNP	nsubj	Xxxxx	True	False
2	,	,	PUNCT	,	punct	,	False	False
3	a	a	DET	DT	det	x	True	True
4	famous	famous	ADJ	JJ	amod	xxxx	True	False
5	Belgian	belgian	ADJ	JJ	amod	Xxxxx	True	False
6	cyclist	cyclist	NOUN	NN	appos	xxxx	True	False
7	,	,	PUNCT	,	punct	,	False	False
8	won	win	VERB	VBD	ROOT	xxx	True	False
9	the	the	DET	DT	det	xxx	True	True
10	Tour	Tour	PROPN	NNP	compound	Xxxx	True	False
11	de	de	ADP	IN	compound	xx	True	False
12	France	France	PROPN	NNP	dobj	Xxxxx	True	False
13	five	five	NUM	CD	nummod	xxxx	True	True
14	times	time	NOUN	NNS	npadvmod	xxxx	True	False
15	.	.	PUNCT	.	punct	.	False	False

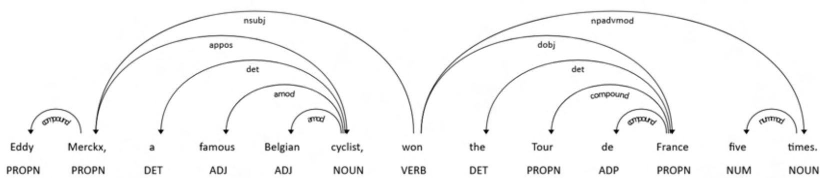


Figure 4.7: The NLP part-of-speech visual representation performed by the Python Spacy package.

4.3.2 Named Entity Recognition

Named Entity Recognition (NER) is the process of information extraction by recognising named entities in unstructured texts. A named entity, sometimes also called "a real world object", is often subdivided in a number of predefined categories such as "person names, organisations, locations, etc.". Those name entities have the main purpose to identify and characterise key information (i.e., by

Table 4.2: Overview of the UPOS defined tags

UPOS	Description
ADJ	Adjective
ADP	Adposition
ADV	adverb
AUX	auxiliary
CCONJ	coordinating conjunction
DET	determiner
INTJ	interjection
NOUN	noun
NUM	numeral
PART	particle
PRON	pronoun
PROPN	proper noun
PUNCT	punctuation
SCONJ	subordinating conjunction
SYM	symbol
VERB	verb
X	other

entities) in the text. These definitions will probably be more clear with a simple example. Let's for instance analyze our sentence about Eddy Merckx. If we consult Figure 4.8, we can see the named entities recognized by the Spacy Python package on top. The package found 4 entities that were in the trained model (i.e., person, Nationalities or religious or political groups (NORP), event and a cardinal). If we bring together those 4 entities we can see that this sentence is about a person called 'Eddy Merckx', he is Belgian and that he has something to do with the event called 'Tour de France'. The term 'five' probably needs more background information to make sense of that entity, but by following the dependencies in the POS tagging (see Fig. 4.7) it can be derived that he won 5 times.

```
Eddy Merckx 0 11 PERSON
Belgian 22 29 NORP
the Tour de France 43 61 EVENT
five 62 66 CARDINAL
```

Figure 4.8: Named Entity Recognition results of the Spacy package.

4.3.3 Text classification

Spoken and written text can be easily understood by human beings. Nuances and subjectivity are usually well grasped by the reader or listener. It is only in the last decade or so that computer technology is able to understand these nuances, emotions and opinions within text. Text classification is a subbranch of Natural Language Processing (NLP) that basically trains an algorithm to classify a piece of text based on certain keywords and how they're correlated. The classical example is the use of sentiment analysis on your product reviews to classify if customers are happy about a certain product.

For text classification in Python, great libraries and frameworks do exist. Scikit learn is an example of such a package that is able to perform basic pre-processing and converting the human text to a computer-understandable format. Two very widely adopted techniques to convert words to numbers are the bag-of-words and the more advanced TFIDF vectorisation. Once this step is performed similar models as would be used on numerical data can be used (e.g., random forests, boosting or neural networks).

4.4 NLP use case: Automatic bicycle racing incident reporting from Tweets

Road cycling is one of the most popular sports talked about on Twitter and a lot of valuable race information is widely shared in near real-time. Popular race events (such as incidents, breakaways, and sprints) are mostly well reported by different users (e.g., fans, journalists, broadcasters and cycling experts). Twitter has a powerful REST API (representational state transfer-application programming interface) which allows us to easily get access to those messages. By grouping all the information within these tweets, a detailed description of an event can be given. Such event data can be very valuable for several stakeholders such as teams, race organisers, broadcasters, jury, and fans. However, finding the relevant messages and clustering them in a correct way is not straightforward. It is, for example, important to know which accounts to follow, i.e., we cannot analyse all of them. Accounts of teams, organisers, journalists, and cycling fans are the most valuable since they actively report about events during the race. By limiting the number of accounts, we speed up the querying (and all subsequent steps) and avoid noise from non-official, lower quality accounts. In our current setup, a set of 24 official team accounts, 41 official race accounts (e.g. 'Paris Roubaix' and 'ParisNice'), 5 media/broadcaster accounts (e.g. 'sporza_koers' and 'NOSwielrennen') and the accounts of RCS Sport, ASO and UCI are used – 73 accounts in total. This account list is a simple JSON file and can easily be updated if needed.

By combining all the information related to a particular event from these different sources, a detailed story of that event can be produced. As an example, we look at the tweets posted by our targeted account types during the 2020 edition of the Tour of Flanders in which an incident with Julian Alaphilippe took place. As illustrated in Figure 4.9, Deceuninck-QuickStep's (Alaphilippe's team) twitter account already provides detailed information about the event. From those tweets, we can extract the event type (a crash), the location (35km to go), the involved rider (alafpolak1), the race (RVV2020), and the possible cause (a motorbike). Additional tweets, posted on other accounts, will help to further enrich and validate the retrieved incident data.

Figure 4.9: Deceuninck-QuickStep tweets of Julian Alaphilippe's Tour of Flanders crash (Twitter, Oct 2020).

Extracting event info from unstructured tweets in an automatic way is challenging. It requires i) natural language processing (NLP) tools to find the different types of entities/information in the unstructured, noisy text messages and ii) classification tools to classify a tweet as a particular race event. We also experienced that considering official race information (such as the race contestants lists published on the UCI website) helps to further improve the rider/race name detections and to validate if the tweet is related to a race happening at the time of tweeting. However, since race and rider names are not always correctly spelled on Twitter or are sometimes abbreviated, fuzzy string-matching techniques are needed to find the correct matches. As can be seen in the example in Figure 4.9, tweet texts sometimes contain hyperlinked strings such as hashtags, usernames, and image/video uniform resource locators (urls) related to a particular event and are often providing additional race info. Those can be detected with regular expressions –the Regular Expression (regex) Python library was used in our tool. Finally, and to further enrich the event data with location information, the timing info and kilometre indications of tweets can be used to query a race route (mostly available prior to a race in GPS exchange format) and retrieve the coordinates corresponding to that time or kilometre indication. In addition, timing info can be used to crop/find video shots that matter in race coverage (if available) and automatically feed them to an incident video database, which – on its turn - can be further investigated with computer vision tools.

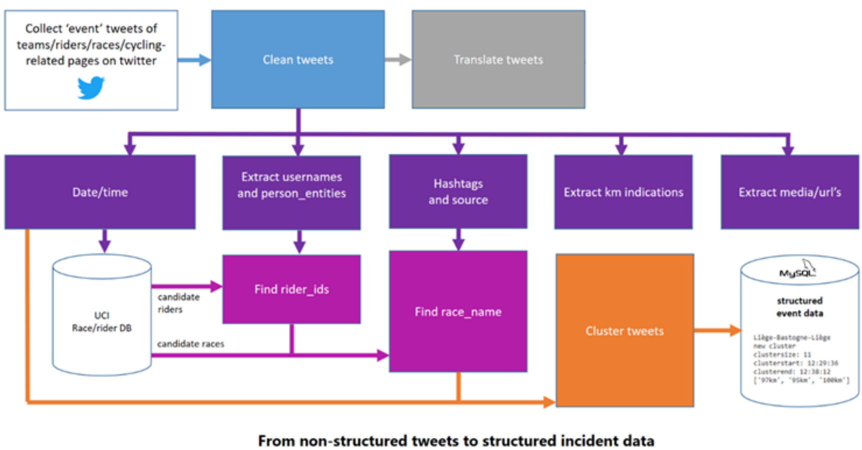


Figure 4.10: Proposed methodology to transform non-structured cycling tweets into structured race event data.

The general architecture that is used to transform non-structured cycling tweets into structured race event data is shown in Figure 4.10. In what follows we will discuss the functioning of the most important building blocks of this architecture and demonstrate them on some tweets of recent race events.

4.4.1 Tweet classification

Since we are only interested in certain topics (i.e., race cycling events), it is necessary to classify the tweets into the topic categories we want to focus on. Tweets, however, do not provide enough shared context required for most content similarity measures [10]. Some authors use ensembles of classifiers to cope with this issue [11], others solve the problem by enriching the tweet context by using external data sources, or distributed language representations trained on a huge amount of data [10]. For race cycling tweets (with some very specific categories and limited training data), we found that transfer learning-based approaches, like Task-aware representation of sentences (TARS) [12] are the most appropriate. The TARS model we used is implemented in Flair by the TARSClassifier class. This model uses a Bidirectional Encoder Representations from Transformers (BERT) encoder, which is renowned for its state-of-the-art performance for NLP tasks [12] [13]. Our dataset was split in 80% training and 20% test data. With a learning_rate of 0.02, a mini_batch_size of 50, and 10 max_epochs, the model didn't require a lot of training to do a decent job, which is illustrated by a low loss for both training and validation and a validation and test F1 score above 99% (Table 4.3). The dataset itself can be downloaded from our Github [2].

Table 4.3: Comparison of the different text classification techniques. The BERT encoders significantly outperform the traditional support vector classifiers

	SPACY SVC (TFIDF Tokenizer)	**SPACY SVC (Count Vectorizer)**	**FLAIR TARSBASE (Bert Encoder) [12]** *10 epochs*	**TF model (BERT Encoder) [13]** *10 epochs*
Accuracy	89%	85%	99%	99%
F1 score	89%	85%	99%	99%
Precision	90%	83%	99%	99%
Recall	88%	88%	99%	99%

4.4.2 Named entity recognition of persons, organisations and locations

Not all riders, organizations and locations are referred to by hashtags or handles. Some of them also occur as plain text. In order to detect those entities as well, we use FlairNLP's SequenceTagger [3] with a pre-trained English CoNLL-03 model that

[2]https://gist.githubusercontent.com/jelledebock/fa941a9038339bcba4dc57380270385f/raw/crash_-no_crash.json

[3]https://github.com/flairNLP

can detect persons, organisations and locations. Tweets are inputted as a character sequence into a pre-trained bidirectional character language model. From this model, a contextual embedding is generated for each word by extracting the first and last character cell states. These embeddings are then passed into a vanilla BiLSTM-CRF sequence labeler, achieving robust state-of-the-art NER results.

4.4.3 Rider and race matching with detected entities

For each race date the UCI Dataride platform contains the name of all races on that date and all riders that participated in them. We scripted a scraper (based on basic REST requests) that converts all this information into JSON format and built an API on top of it that can be queried by date, rider/team name and race name. Since the spelling of race and rider names in tweets and the UCI database can differ it is not always possible to directly link tweets to a particular race and a set of involved riders. In order to cope with this issue, we scripted a rule-based mechanism to automatically generate some alternative 'candidate' rider/team/race names and combined it with a fuzzy string-matching algorithm. For Liège-Bastogne-Liège, for example, we scripted string operations to automatically add the following candidate race names: lbl, Liege, lbl2020. In a similar way, different formats of rider names are generated (by mixing their first and last name and making different combinations of them). Based on these candidate lists, the fuzzy string matching tries to find a good hit for each detected person/race entity. The string matching itself is based on fuzzywuzzy, a string similarity matching library that uses Levenshtein distance to calculate the differences between two strings. If the matching score is above 70%, the match (i.e., the official UCI race or rider name) is included in our dataset.

4.4.4 Temporal clustering

In this last step of the tweet analysis, we first group the tweets based on the detected race names and race dates. For each of these groups we analyse the timestamps when the tweets have been posted within a running time window. If we have 3 or more tweets with a maximum time difference of 240 seconds, a new cluster (event) is found. The minimal number of tweets in a cluster is chosen to be three to limit the amount of false positive detections (and to focus on the most important events). However, for less popular races we would advise to lower it to two. Race popularity (i.e., the total amount of tweets of previous editions of a particular race) could maybe be used to automatically set this threshold and will be part of further research.

4.4.5 Use case example: Remco Evenepoel crashes in the 2023 Giro d'Italia

The overall Twitter pipeline is illustrated in Figure 4.11 where the original Tweets are showed on the left. The right part of the image shows how the incident was documented in the incident database after the Tweets were processed by the presented techniques.

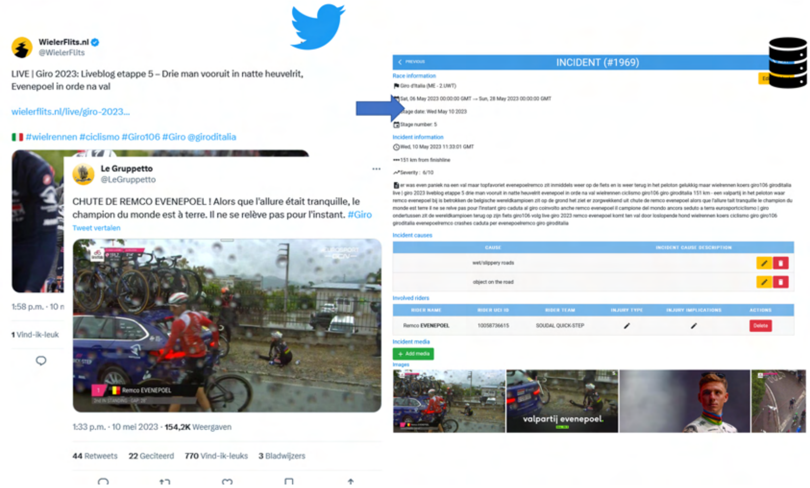


Figure 4.11: Example of a processed incident from Twitter. From Tweets (left) to a record in the database (right)

4.5 Conclusion

A lot of written and spoken text is being collected, and cycling is no exception to this rule. A lot of the data is structured (e.g., in databases, Excel-files or via web services), but a lot of the data remains unstructured (e.g., Tweets, incident reports or race radio).

Structured data has the huge advantage it does not need advanced preprocessing and that it is very qualitative. In contrast, unstructured data needs to be converted first into structured data. Based on the type of textual data and the adopted techniques, this conversion is not 100 percent accurate and has the risk of losing some context and details in the process. However, and as demonstrated in this chapter, state-of-the-art NLP methodologies are quite successful in the conversion process whilst maintaining as much detail as possible. Furthermore, we also concluded that NLP models are also able to perform classification tasks based on the Tweet corpora.

References

[1] T. Harbert. *Tapping the power of unstructured data*, February 2022. Online; retrieved on 4-November-2022. Available from: https://mitsloan.mit.edu/ideas-made-to-matter/tapping-power-unstructured-data.

[2] I. E. Team. *Tapping the power of unstructured data*, September 2022. Online; retrieved on 7-November-2022. Available from: https://in.indeed.com/career-advice/career-development/what-is-web-scraping.

[3] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*, 2018. arXiv:1802.02611.

[4] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. *The Cityscapes Dataset for Semantic Urban Scene Understanding*, 2016. arXiv:1604.01685.

[5] J. AI. *GitHub - JaidedAI/EasyOCR: Ready-to-use OCR with 80+ supported languages and all popular writing scripts including Latin, Chinese, Arabic, Devanagari, Cyrillic and etc.* https://github.com/JaidedAI/EasyOCR, oct 13 2022. [Online; accessed 2022-11-14].

[6] R. Smith. *An Overview of the Tesseract OCR Engine*. In ICDAR '07: Proceedings of the Ninth International Conference on Document Analysis and Recognition, pages 629–633, Washington, DC, USA, 2007. IEEE Computer Society. Available from: http://www.google.de/research/pubs/archive/33418.pdf.

[7] F. Morales. *keras-ocr*. https://github.com/faustomorales/keras-ocr, oct 3 2019. [Online; accessed 2022-11-17].

[8] L. V.I. *Binary codes capable of correcting deletions, insertions, and reversals*. Soviet physics doklady, 10:707–710, 1965.

[9] J. Hirschberg and C. D. Manning. *Advances in natural language processing*. Science, 349(6245):261–266, 2015. Available from: https://www.science.org/doi/abs/10.1126/science.aaa8685.

[10] Q. Li, S. Shah, X. Liu, A. Nourbakhsh, and R. Fang. *Tweet topic classification using distributed language representations*. In 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pages 81–88. IEEE, 2016.

[11] G. L. De la Peña Sarracén. *Ensembles of Methods for Tweet Topic Classification*. In IberEval@ SEPLN, pages 15–19, 2017.

[12] K. Halder, A. Akbik, J. Krapac, and R. Vollgraf. *Task-aware representation of sentences for generic text classification.* In Proceedings of the 28th International Conference on Computational Linguistics, pages 3202–3213, 2020.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805, 2018.

5
Geospatial and time analysis in cycling

"Aut viam inveniam aut faciam."

–Hannibal (218 BC)

In this chapter geospatial (time-stamped) analysis techniques are studied in detail. We will provide an overview on how we use geospatial data in safety, performance and fan-engagement in professional cycling. The research question of this chapter can be formulated as: **"Is it possible to use maps, geospatial databases and GPS/workout files for the extraction of racecourse related metadata to improve course safety or to predict race performance or outcome?"**

5.1 What is geospatial (time-stamped) data and analysis?

Geospatial data is defined by IBM as information recorded in conjunction with a geographic indicator of some type [1]. If the geospatial data is time-stamped as well, i.e., a time indication is provided to the information. In professional sports,

[1]https://www.ibm.com/topics/geospatial-data

GPS tracking has gained huge interest over the past few years. This data is extremely valuable to analyze behaviour of opponents, make post-match/race analyses or to help the spectator experience during a race/match.

5.2 Storytelling and performance: Engagement scoring in cyclocross

5.2.1 Use case description

Cyclocross is a cycling discipline in which riders are riding laps on varying surfaces (e.g., sand, mud, forrest roads and tarmac). The lap can contain man-made obstacles such as barriers to hop over, sand pits or scaffold bridges. Races have a fixed duration (i.e., men ride around an hour, women tend to ride around fifty minutes). Within that duration a number of laps of the course are completed and the rider that crosses the finish line first after that number of laps wins the race. The format is rather similar to cross country mountainbiking or criterium road racing, but it is the technicality and the fact that anything can happen during a race that makes the cyclocross discipline interesting to spectate. In cycling, and cyclocross is no exception to this, a lot of the wearable and information technology are relatively well adopted. Most riders register their performance data during races either on a watch or a GPS head unit and most races have an advanced timing system (e.g., MyLaps loops with transponders) to make sure that not a single detail is left unattended. The big races usually have bike-mounted trackers to allow real-time location and performance monitoring of the riders in the race.

The fact that this data is recorded has a lot of possible value for storytelling and performance. Within this use case we examine the possibility and value of collecting GPS data to research which parts of the laps are the most interesting. Based on the stakeholder these insights can range from assisting in the decision-making on where to make the difference as riders, where to put extra cameras as broadcasters or where to stand to cheer on your favourite rider as a fan.

In this use case we will use cyclocross as the discipline to showcase the methodology, but it should be noted that the explained principles can be easily transferred to other disciplines (e.g., road cycling, time trial racing and mountainbiking) and even to other sports (e.g., city marathon running, downhill racing or cross-country skiing and Formula 1 racing).

5.2.2 Data

As mentioned in the use case's description, location data is the main data source that is used for this analysis. The data can be either analysed in real-time or as post-processing. For this use case we will start from workout data of athletes that was retrieved from GPS workout files (e.g., GPX, FIT).

Riders can record a lot of information with their GPS head units and/or watches nowadays. Most riders have power meters on their bikes and they wear heart rate straps to get an insight of how hard they are working during training sessions and races. However, for these analyses, the minimum amount of information that is needed is the time-stamped logging of GPS coordinates (i.e., a latitude and longitude pair). Other performance related data might be valuable to overlay the sector analyses, or to calculate an even more advanced engagement score, but it is not directly used in this version of the segment analysis.

The next important data source are the coordinates of the lap. Ideally, these coordinates are as accurate as possible. In our experiments, we manually created the courses with a route planner that has satellite imagery of the region to ensure that the points are mapped as closely to the "real course". If required, this step could also be further refined by employing specialised people (i.e., land surveyors) to get the most accurate representation of the course.

The final input source of the methodology is the exact definition of the sectors. Once again, this can be defined in multiple ways. For instance, in cyclocross courses, we could subdivide the course features as separate segments (e.g., sand pit, uphill running segment or finishing straight). In practice, this is the most convenient way to get valuable insights. For this experiment, however, and for the ease of evaluation, fixed length sectors were defined for the analyses that were performed.

The input sources and the desired output of the use cases are visualised in Figure 5.1.

5.2.3 Analysis

The methodology can be split in three big sub-steps, as illustrated in Figure 5.1. In this subsection we will elaborate on each of these sub-steps to provide a thorough understanding of the workflow.

5.2.3.1 Lap extraction

The first step is to extract the laps from a workout file of an athlete that contains multiple laps of the course that was ridden during that activity. If we study

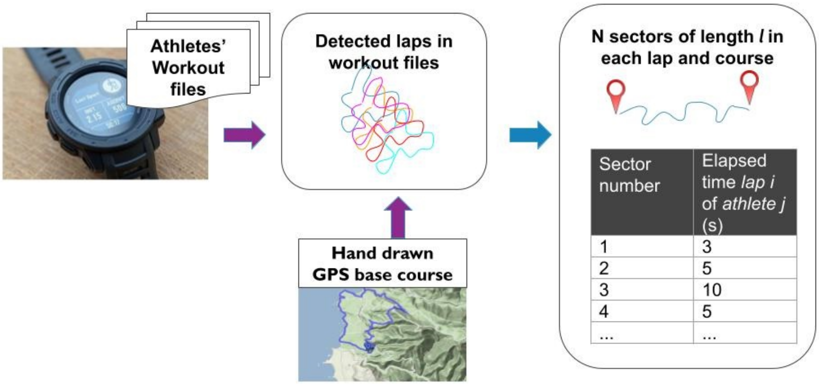


Figure 5.1: Overview of the GPS-driven segmentation principle.

both the patterns of the latitude and longitude traces of a rider's workout we can clearly see the repetitiveness in it (see Figure 5.2).

Fast Fourier Transform is a technique used to convert time-domain signals to the frequency-domain spectrum, which might reveal certain periodicity or rythmicity in the signal. Li et al. [1] used this technique to successfully detect sleep/activity cycles in new-borns at different ages. The authors were able to define which periodicity was most dominant at the varying ages of young kids. This methodology is also often used in music (recognition algorithms) to characterise a music part.

In Figure 5.3, the Fast Fourier Transformation was applied on both the latitude and the longitude time series that was showcased in Figure 5.2. We can clearly observe a peak at five, which is the correct number of laps that the rider completed. This principle has been illustrated by Francis in his excellent blog post "Strava - Automatic Lap Detection" [2] and was successfully adopted in this methodology.

As a next step, once we have found the exact number of laps that are contained in the workout file, the lap boundaries (i.e., start and end of a lap) have to be found. For this task, we implemented a straightforward algorithm that uses a region-based search methodology. As a start we have to trim the workout file at the start and end so that we can assume that the riders are actually "moving" throughout the workout.

The pseudocode to find the split point for the next lap is provided in Algorithm 5.1. The search procedure is always initiated from $start$. For the first lap the $start$ index that was found when the activity starts. For the next few laps (as many times as the number of laps that were found by the Fast Fourier Transform methodology) we set $start$ to the point previously found by the procedure

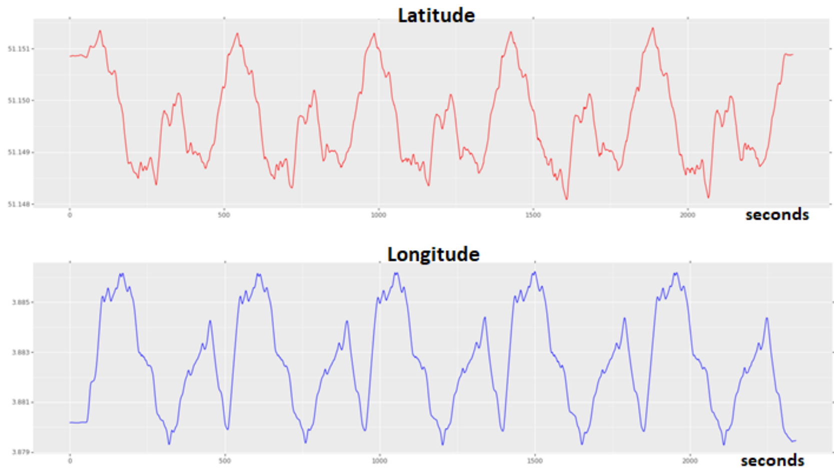


Figure 5.2: Repetitiveness of latitude and longitude coordinates of an athlete's workout file.

(Algorithm 5.1). The $findClosestPoint$ procedure in Line 6 (Alg. 5.1) is a simple proximity-based search method that uses the haversine distance formula (see Listing 5.1 for a Python implementation) to find the closest point within the specified range.

Algorithm 5.1 Region based lap boundary search

1: **procedure** findNextPoint(start, nCoordinates, nLaps)
2: startPoint ← searchStart
3: theoreticalNextLapPoint ← searchStart + nCoordinates/nLaps
4: margin ← 10/100 or 15/100
5: rangeToSearch ← [theoreticalNextLapPoint - margin, ... theoreticalNextLapPoint + margin]
6: nextLapPoint, lapPointDistance ← **findClosestPoint**(startPoint, rangeToSearch)
7: **return** nextLapPoint **if** lapPointDistance <= requiredAccuracyM

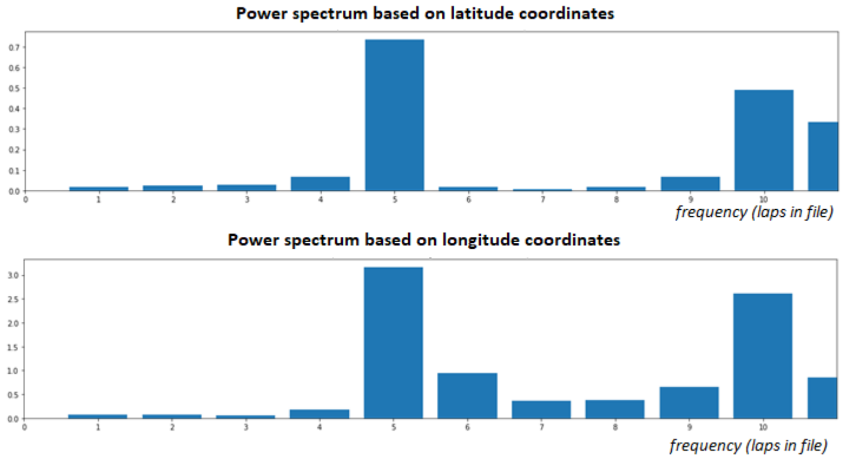


Figure 5.3: Fast Fourier Transform applied on the latitude/longitude time series of Figure 5.2

Listing 5.1: Python implementation of the Haversine formula

```
from math import radians, cos, sin, asin, sqrt

def haversine(lon1, lat1, lon2, lat2):
    """
    Calculate the great circle distance in kilometers between two points on the earth  (specified in decimal degrees)
    """
    # convert decimal degrees to radians
    lon1, lat1, lon2, lat2 = map(radians, [lon1, lat1, lon2, lat2])

    # haversine formula
    dlon = lon2 - lon1
    dlat = lat2 - lat1
    a = sin(dlat/2)**2 + cos(lat1) * cos(lat2) * sin(dlon/2)**2
    c = 2 * asin(sqrt(a))
```

```
    r = 6371 # Radius of earth in kilometers. Use
        ↪ 3956 for miles. Determines return value
        ↪ units.
    return c * r
```

5.2.3.2 GPS sector matching

In the next step, as the number of laps and each their boundaries are available, each point of the riders' workout files is matched with the coordinates of the course. Although this step might sound very straightforward and easy, it is often not the case when dealing with GPS files recorded by different watches and head units. A study of Johansson et al. [3] in which different wearable GPS units were tested during a 56km running race have 0.6% (median) difference in the distance covered (officially measured race distance versus distance reported by the watch). They also conclude that some units are a lot better than others, so this variability has to be considered as well in our use case. Furthermore, 0.6% might seem very minimal, but if we have very short segments (e.g., a sandpit is typically only a hundred meters long), using the raw coordinates from the GPS devices might offer difficulties.

To respond to this "shortcoming" of GPS recordings we implemented a more clever methodology that tries to match the sometimes blurry recorded coordinates with the coordinates of the ground truth (i.e., the hand drawn course). The main formula is displayed in Equation 5.1. As could be observed, the function takes the square root of the sum of where the point is positioned along the sector s ($d_{course,j}$ and $d_{lap,i}$) and the distance between the point on the course ($pt_{lap,i}$) and the point we want to match with the course ($pt_{s,j}$). This action is performed for every point i on the hand-drawn course that might be close to the candidate point (Eq. 5.2 & 5.3).

$$d_{(s_j,pt_i)} = min\left(\sqrt{\left[s_l\left|1-\frac{d_{course,j}}{d_{lap,i}}\right|\right]^2 + dist(pt_{lap,i},pt_{s,j})^2}\right) \tag{5.1}$$

$$\forall i \in \left[pt_{s,j-1}+\frac{1}{4}n, pt_{s,j+1}-\frac{1}{4}n\right] \tag{5.2}$$

$$n \longleftarrow n\_points\left[pt_{s,j-1}, pt_{s,j+1}\right] \tag{5.3}$$

5.2.3.3 Sector analysis and engagement scoring

Now that laps and its boundaries were extracted from the athletes' workout files and we matched every point of an athlete's sectors throughout the laps with hand-drawn lap base course, we can start analysing segment times across athletes and between laps of a single athlete.

Our methodology was first tested during the Belgian national cyclocross championships in 2020 in Antwerp. It was a race that was characterised by the now famous pontoon and a lot of sand sectors. As a test, we analysed the GPS workout files of the U23 category (who rode the course one day prior to the elite men). The workout files of the U23 category were matched against a hand-drawn version of the championship course and 50 meter segments were created. Figure 5.4 shows an overview of the statistics that were extracted from the lap and segment segmentation of a total of 11 riders. Some riders (i.e., 4 riders) could not finish the full races and were pulled from the race by a mechanical defect or were expelled from the course due to the eighty percent rule that exists in cyclocross. For the readers not familiar with the cyclocross discipline, the 80 percent rule ensures an unobstructed course for the race leaders by taking out the riders on the course that are more than 80 percent slower than the race leader's first lap time. In total, we have 58 fully completed laps of the 11 riders who were included in the study. This gives us 58 values for each segment to perform some order statistics on them. If we consult Figure 5.4, mainly one segment rises above the others, which is the twelfth fifty meter segments. If we look where exactly this twelfth sector is located on the course, we can find that this is an off-camber sector that became very tricky due to relatively heavy rainfall in the days prior to the event. Other sectors with relatively high variety in segment times were mostly uphill bursts and sandpit sectors.

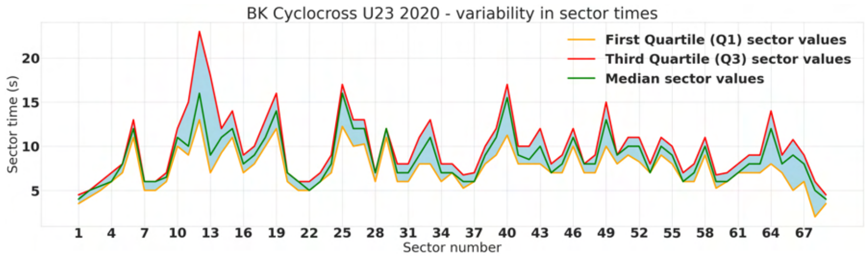


Figure 5.4: First quartile, median and third quartile sector times for the Belgian National Championship course in 2020

Finally, this data can also be transformed in an "engagement score" which uses the individual sector time data to calculate a metric that indicates how attractive (for spectators to watch) or challenging (for riders to complete) a certain segment on the course might be. The formula to calculate attractivity is presented in Equation 5.4.

$$s_i = \ln\left[\frac{|Q1_i - Q3_i|Q2_i}{\max\limits_{\forall j\in[0\ldots n]} t_j - \min\limits_{\forall j\in[0\ldots n]} t_j} Q2_i + \left(1 - \frac{Q2_i}{\max\limits_{\forall j\in[0\ldots n]} t_j - \min\limits_{\forall j\in[0\ldots n]} t_j}\right)\frac{1}{n}\sum_{j=1}^{n} t_j\right] \tag{5.4}$$

Where:

$$\begin{aligned}
Q1_i &= \text{First quartile of sector times for segment i}\\
Q3_i &= \text{Third quartile of sector times for segment i}\\
Q2_i &= \text{Median sector time for segment i}\\
n &= \text{number of segments on the course}\\
t_j &= \text{Median sector time for segment j}\\
s_i &= \text{engagement score for segment i}
\end{aligned}$$

When the engagement formula in Eq. 5.4 is applied to the data of the U23 national championship race we get a graph as presented in Figure 5.5. The attractive sectors were already derived from Figure 5.4, but this formula represents them as a single number that is weighted from 0 to 10 and allow nice visualizations of the data as presented on the research group website [2].

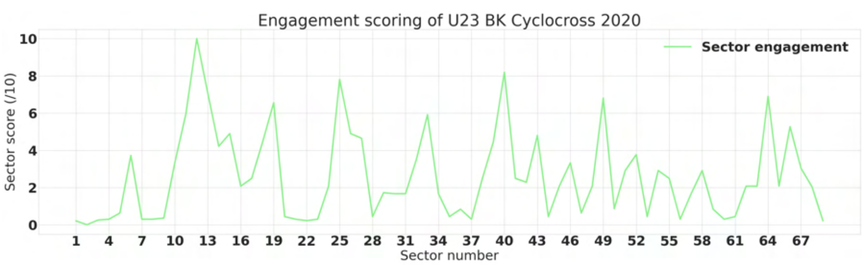


Figure 5.5: Engagement scoring for the Belgian National Championship course in 2020

[2] https://users.ugent.be/~jcdbock/BK/kaart.html

5.2.4 Evaluation

In this study we successfully implemented a methodology that was able to provide stakeholders with attractivity or level of difficulty of segments on a course. This has its applications in both storytelling and performance contexts. However, the study still has room for improvements. Firstly, an obvious improvement is the shift towards variable length sectors. The only adaptation that's required for this modification is the weighting of the segment times with the distance it covers. When this path is chosen, the fact that longer segments will have larger absolute differences as the shorter ones should be also considered in the engagement formula (Equation 5.4). Secondly, and a bit of a more challenging one, is the accuracy of GPS signals. As mentioned previously [3], GPS is not the most accurate, especially when the athletes are competing in dense forests (as it is often the case in cyclocross). The matching of raw GPS workout files with a hand-drawn course (as described in the methodology) is already one aspect that mitigates the shortcoming, but other possible solutions might be the adoption of novel tracking technologies such as Ultra Wide Band (UWB) or differential GPS (DGPS). Thirdly, another potentially useful improvement is the inclusion of metadata such as weather information, surface type or elevation rise/drop in the engagement formula. One of the shortcomings is that the current formula is very weather dependent.

When the weather changes it might make some sectors barely rideable which can result in a change in engagement scores. If weather impact would be included in the formula, this would produce a much more robust and weather-independent scoring mechanism. When weather parameters would be recorded alongside the segment parameters, the impact of changing weather circumstances could also be directly included in the engagement scoring (e.g., usually speaking, a sandpit sector gets less engaging when it is wet). Lastly, and as mentioned previously, this methodology also has a lot of performance analysis possibilities.

5.3 Storytelling : GPS-driven camera selection

This section is a rework of the publication *"GPS-driven camera selection in cyclocross races"* [4].

5.3.1 Introduction

Races are broadcasted on Belgian national television and the action is captured by multiple cameras. The optimal camera stream for a given moment in the race is usually selected by the broadcast director who is monitoring the race footage in the camera truck at the race location. When the race gets very eventful and a lot of action happens simultaneously this can be a rather hectic job. Furthermore, it is not unimaginable that the directors in charge have some kind of subjectivity or preferences for certain riders or sectors. In the current workflow, a lot of video footage is lost, as only the main broadcast is usually kept at their servers. This broadcast is commonly consisting of a single (manually) selected video stream at a time. Other active video streams are not used and very often even not stored. As it will be illustrated in this paper and in further research, this is definitely a missed opportunity because having the raw footage at each moment in the race might be valuable to generate additional race insights.

Data-driven race reporting might offer a solution for the previously mentioned shortcomings. The evolution of wearable technology and the implementation of new wireless standards allows race organisers to track every rider on the cyclocross course in real-time. A study of Hess et al. [5] showed that nowadays accurate GPS tracking is even possible with most of the available smartphones. However, for demanding events such as cyclocross (a lot of direction changes on a small surface) and long road races (high battery demands), a dedicated GPS tracker is used more often. The Quarq Qollector is an example of such a device that tracks GPS location and sensor data such as heart rate, power and cadence. The sensor is usually mounted on the riders' bikes or at the back of their skin-suit and is transmitting its data in real-time to the Quarq servers using 3G cellular data connectivity.

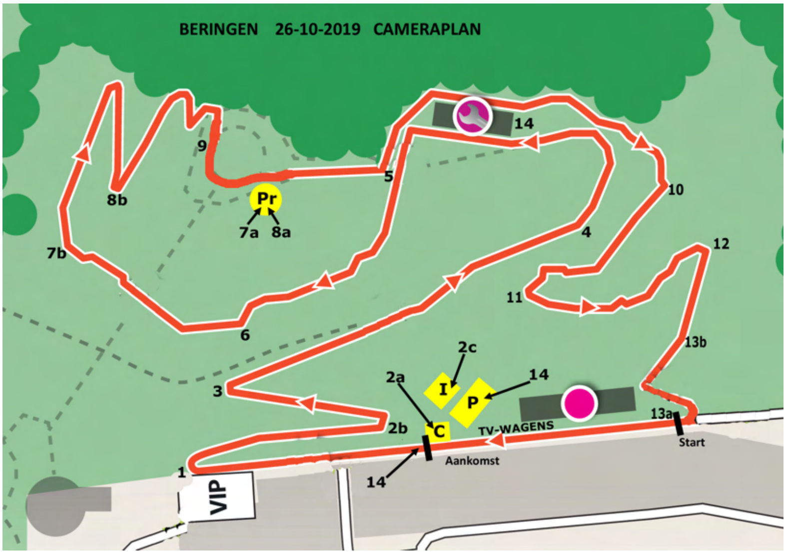


Figure 5.6: Example of the race broadcasters plan of the Beringen cyclocross track. Plan contains the course (red line). Cameras are annotated as black numbers.

Based on the location of a rider we can start searching for the camera on the course that is best used to capture the riders. Cyclocross courses are challenging for riders, but also for the film crew to bring the race to the television viewer. The design of a cyclocross track can be considered as “a fine art” as courses are usually built by a handful of specialists and often ex-professional riders such as Adrie van der Poel, Richard Groenendael and Erwin Vervecken. Planning, building and finalising the perfect track takes weeks to provide riders with a challenging, but safe course. Obstacles such as barriers, sandpits and off-camber sectors make cyclocross interesting to watch both at location and on the television. The broadcast director visits the location a couple of days in advance and accurately plans the camera setup at race-day. Cameras, identified by their index number, are placed on a kind of a topological map of the course (see Figure 5.6 for an example of such a course plan).

The combination of the multiple camera streams, detailed riders locations and carefully planned and documented course and camera location make the cyclocross broadcasting a very interesting use case for automation. As mentioned, camera selection and race monitoring can be very hectic at times. In this paper we present a mechanism that tracks the rider’s location based on either GPS files or real-time data and returns the closest camera to a given rider. Video editors and directors could benefit from this mechanism as streams can be pre-filtered based on the vicinity of riders. Finally, it can also be used to generate a summary of a rider/team across all cameras.

5.3.2 Methodology

As mentioned, both the advancements in technology and the popularity of the sport are arousing the interest of video broadcasters worldwide. Men and women races are usually broadcasted live. In this section we introduce the steps of the methodology to match a rider based on its GPS location to the closest camera.

5.3.2.1 Course digitisation

The first step toward a GPS-assisted camera proximity algorithm is the digitisation and annotation of the race course and cameras positioned along that course. Currently this is a partly manual process in which a portable document format (PDF) and a GPS exchange format (GPX) file are used as input. The GPX file can be either provided by a rider who did practise on the course or it can be drawn with tools such as Komoot, RideWithGPS or Garmin Connect. This GPX file of the course is converted with a Python script to a GeoJSON linestring object. GeoJSON is a very useful standard to store, represent and programmatically access geographic data. The linestring of the course consists of multiple interconnected points (latitude/longitude pairs) of the track (red line, Figure 5.7).

Figure 5.7: Digitized version of the cyclocross course of the Rectavit Series Leuven (2020). Red line is the GeoJSON course linestring. The blue markers are GeoJSON point features and are representing the cameras and their respective identifiers (camera id), which are hidden in this picture.

The locations of the cameras are registered by the broadcasters in a PDF containing both the course and the locations of the various cameras across the track. Every camera on the schematic has an identifier (e.g., Camera 7a), but the logic behind the numbering is not related to its exact location on the track. To make this camera locations programming interface friendly, we manually added the locations to the GeoJSON track file of the previous step. The camera location was stored as a GeoJSON point feature with the identifier of the camera as its property (see Listing 5.2) for the JSON code representing the camera location of camera 2). The camera digitisation is a rather manual procedure, but tools such as GeoJSON.io [3] are making this rather straightforward. Future work will also further focus on

[3] http://geojson.io

Listing 5.2: GeoJSON code representing the "camera 2" coordinates on the course as a point geometry. The feature has its type and camera_id stored as the properties.

```
{
	"type": "Feature",
	"properties": {
		"type": "camera",
		"camera_id": 2,
	},
	"geometry": {
		"type": "Point",
		"coordinates": [
			4.710833430290222,
			50.85269722820937
		]
	}
}
```

the development of a tagging tool that can automatically generate the GeoJSON data.

5.3.2.2 Rider-location processing

Once we have a structured representation of the track and the cameras placed around it we can start looking for riders on it. As mentioned in the introduction, several methods do exist to accurately track and trace riders on the cyclocross course. We divided rider tracking in two separate approaches. In the first approach, GPS eXchange (GPX) files were used to get time-stamped locations of the riders, who recorded their races with their own GPS head units or watches. These kinds of files are usually uploaded to online web applications such as Strava [4], Trainingpeaks [5] or Today's plan [6] for further analysis. Although this is not offering us real-time locations of the riders, it still provides great information for our camera and rider matching algorithm for post-race video analysis and summarisation. Another benefit of using this approach is that there is no need to interface

[4] https://www.strava.com/
[5] https://www.trainingpeaks.com/
[6] https://www.todaysplan.com.au/

with external Application Programming interfaces (APIs) and real-time data storage and management is not an issue.

The second approach to get the riders' locations is by using one of the many connected GPS trackers (e.g., SPOT or Quarq Qollector). These GPS trackers are usually worn on the riders' bodies or fixed to their bikes. Most of the trackers are accessible by APIs. A Python program was written to interface, read and interpret the data from the Quarq qollector API. Each Quarq Qollector device has a tracking id (tid). With the API we were able to periodically (every second) retrieve all the sensor and location data that were recorded by the Quarq Qollector for a list of participating tracking ids.

Both approaches result in the location of the tracked riders for any given moment during the race. An abstraction layer was written to quickly get a location for a specific rider at a given timestamp, independent from the underlying rider tracking technology.

5.3.2.3 Camera rider matching algorithm

Now that race track, camera locations and rider locations are available and converted to a computer-understandable format a camera matching algorithm can be introduced.

As a start, a formal definition of the proximity of a rider to a certain camera should be given. The calculation can be tackled in a couple of different ways. A first possibility might be the calculation of the haversine distance between a rider and all of the available cameras. As mentioned previously, the haversine distance is a formula that is very important for geospatial purposes as it is calculating the distance between two points (using their latitude and longitude coordinates) on a sphere (i.e., the earth) [6]. In some cases, this might be sufficient, but sometimes the track layout of a cyclocross race might not be ideal for this calculation method. Tracks usually consist of lots of tight turns on a compact area. Figure 5.8 illustrates this principle. The yellow circled camera is the closest to the rider (represented by the red X). This might result in a good shot of the rider, based on the orientation and direct visibility of that specific camera at this location.

Another possibility is by projection of both the cameras and the riders on the course's linestring. The proximity of a camera is now determined by the distance along the course from the rider to the projection of the camera (see Figure 5.9). The projection is achieved by finding the index of the point on the course with minimum (haversine) distance to the camera's location. If n is the number of cameras and l the number of course points this approach has a time complexity of $O(ln)$. As mentioned, camera and course data is available prior to the race so this step can be pre-processed for faster real-time querying.

A final optimisation that can be done is the pre-indexing of the course points

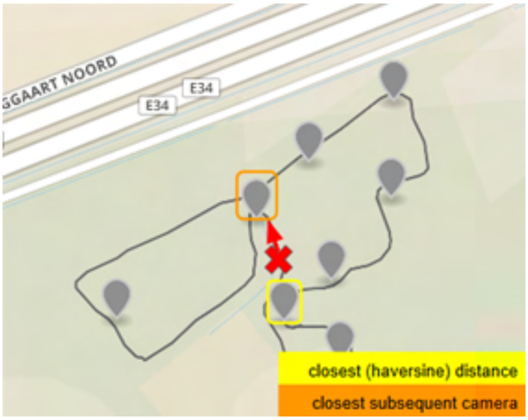


Figure 5.8: Camera setup on a test course. The red X is the current location of the rider following the course in the direction of the red arrow. Yellow circled marker is the absolute closest camera. Orange circled marker is the next camera the rider will visit on the course.

based on the closest cameras (see Figure 5.10). Finding the closest camera for a given rider is reduced to a search in a pre-computed list that is mapping each point on the course to the best/closest camera. Furthermore, it is worth mentioning that labelling which camera is serving which point on the courses can also be done manually (by the responsible video director) prior to the race. Although this process is a manual effort (that can of course be facilitated by a software tool), it will sometimes be more accurate as dense forests, elevation differences or audiences blocking the direct view of the cameramen might cause the closest subsequent camera technique to return the wrong best/closest camera.

Now that we are able to get the closest camera for a given rider's location we also want to get an idea of how far away the camera is from the rider. Figure 5.9 shows the definition of the distance ($d_{r,c}$) between a camera c and the rider r. The distance is not the direct distance between both points, but is again the distance along the course's path. A negative distance means that the camera is behind the rider and positioned earlier on the course. A positive distance is that the rider is approaching this camera.

To further speed up the camera matching algorithm the distances of riders and cameras on the courses were mapped with respect to a fixed point on the course. We chose the start of the course as the reference point (see Figure 5.11). This allowed pre-indexing all camera projection locations elapsed distances w.r.t. the course starting point. Finding the distance between a rider and a camera with given index is now a case of finding the elapsed distance of the rider (projected on the course line) w.r.t the course reference point, look up of the elapsed distance

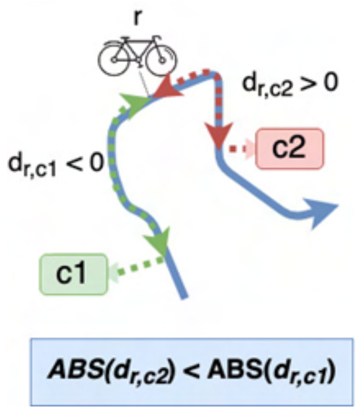


Figure 5.9: Illustration of closest camera along the course principle. Rider r is "snapped" on the blue course line following the direction of the arrow. Cameras c1 and c2 are also projected on the course linestring. Distance d is the distance from rider r to a camera c with a given index, following the course path. Negative distances are cameras behind the rider, positive distances are cameras the rider is approaching. Closest camera is the camera with the smallest absolute distance.

of the camera's elapsed distance (with respect to the course reference point) and subtracting the first distance from the latter.

5.3.3 Results

With the introduced building blocks we can now track riders on the course and find their distances to a given camera. This approach facilitates a number of interesting analyses. A first possibility is the sorting of riders based on their proximity to a certain camera. Figure 5.12 shows an example of such a search for the cyclocross race of Leuven around minute 50 of the race. Solely based on sensor data and the course and camera metadata we can reduce the number of riders that might have been filmed by a certain camera. As shown in Figure 5.12, the total number of riders nearby was limited to two riders. From one side this methodology might give us an idea of how many and who to expect at a given camera at a given moment in the race. On the other side a certain rider could also be followed across the different cameras which is opening the door to track a specific rider across all cameras during the race. This can be definitely useful for teams or for fans who are only interested to see their favourite rider.

With this information in mind, we can start collecting video extracts from different camera streams and compile them in either a rider specific summary

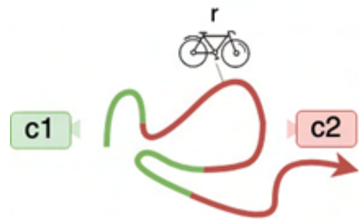


Figure 5.10: Alternative straightforward approach. Only riders are projected on the course. Each point on the course is labelled with the camera that is serving that location on the track. This can be done by the race directors prior to the race or can be the result of the camera on course projection technique. Getting the closest camera is now a case of looking at the camera index of the point of the rider on the course. Distance to the camera can also be pre-processed for each point on the course using either the haversine or the distance across the course metric.

or extract only the parts from a raw camera stream in which riders were nearby. When the streams are available as individual files (e.g., camera_1.mp4, camera_17.mp4 and camera_5.mp4 in Listing 5.3) it is possible to make an m3u playlist which is bundling all the separate clips as one continuous video, without the need to duplicate the data of the original camera streams. Listing 5.3 shows the content of an m3u file, selecting three video extracts from three different camera stream files.

Listing 5.3: Sample .m3u file. An m3u file is a kind of playlist that extracts specific parts from larger video files (in our example the different camera streams) and plays them subsequently.

```
#EXTVLCOPT:start-time=28
#EXTVLCOPT:stop-time=36
camera_1.mp4
#EXTVLCOPT:start-time=500
#EXTVLCOPT:stop-time=550
camera_17.mp4
#EXTVLCOPT:start-time=900
#EXTVLCOPT:stop-time=1100
camera_5.mp4
```

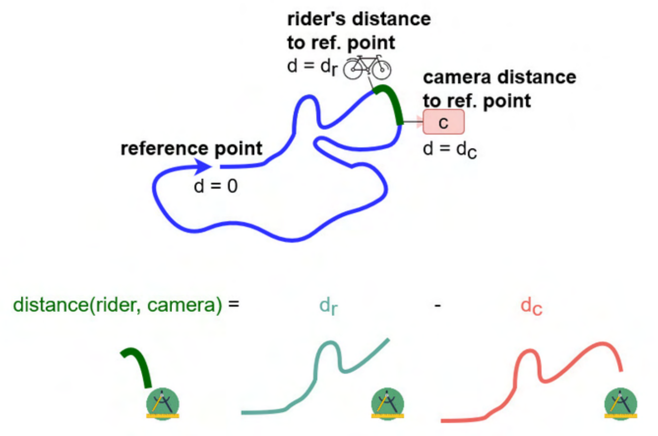
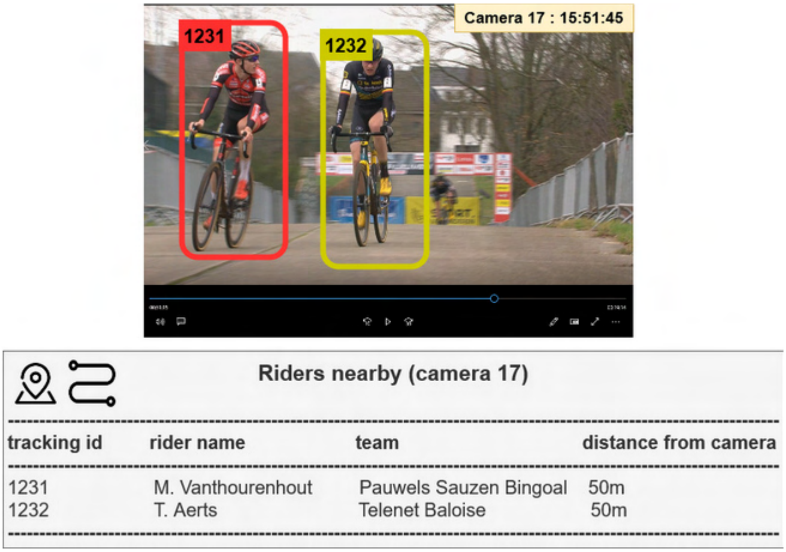


Figure 5.11: Visual illustration how the referencing approach uses distances w.r.t. a reference point. The distance between two points is the result of the difference between distances from both points to the reference point..

5.3.4 Conclusion and future work

The proposed mechanism is not only useful for race broadcast directors, but the filtered streams can also be further processed by video processing algorithms to annotate, index and document the race footage. The camera proximity algorithm can serve as a first filter for the video footage before it is processed by the video processing tools which are also currently being developed within the cyclocross video research project. Techniques such as text recognition, face recognition and pose estimation are used to further annotate the filtered video extracts from the camera streams. The camera proximity algorithm can also further assist the computer vision modules. For instance, if in the example of Figure 5.12, our text recognition detects "Telenet" and the camera proximity algorithm outputs that only one rider of that team is in the neighbourhood of the camera in question (i.e., camera 17), we can easily find out the rider's identity (i.e., T. Aerts).

As mentioned, another aspect we're currently focusing on is the creation of an intuitive camera coverage labelling tool for the race directors. Such a tool should allow them to easily label the locations of the cameras on the race course using a web application on a wearable device. The tool could do the conversion process to the GeoJSON standard automatically, which would be a huge step towards a fully automated camera proximity algorithm.



tracking id	rider name	team	distance from camera
1231	M. Vanthourenhout	Pauwels Sauzen Bingoal	50m
1232	T. Aerts	Telenet Baloise	50m

Figure 5.12: Example output of the camera matching algorithm for a cyclocross race of the 2019-2020 season. Verification on the video stream that riders 1231 and 1232 are visible by the camera. This shows the feasibility of tracking a rider during the race across the available cameras which is facilitating the video searching process.

5.4 Safety: GPS-driven course analysis

This is a partial rework of the submission "Road cycling safety scoring based on geospatial analysis, computer vision and machine learning" [7]

The goal of geospatial analysis is to convert a list of coordinates into a set of actionable safety insights based on algorithms and a geospatial database such as OpenStreetMap (OSM). Different types of races (e.g., hilly, flat, cobbled or mountainous) will require different aspects to focus upon when the safety is analysed. The course type profiling is done based on two methods. The first method calculates the difficulty of a course, the second one predicts how big the first group sprinting for victory will be. Finally, when an idea of the race scenario is provided, we can further focus on potentially dangerous aspects on the course. If there is a sprint, elevation, straightness and the number of turns in the last few kilometres will have a crucial impact on overall course safety. Last, but not least, the analysis also provides a relative risk score of segments of the course based on the information that is available in the geospatial database. In the following subsections we will introduce the previously mentioned analyses that could be performed on a course's GPS file.

5.4.1 Course hardness

The course difficulty is a ranking that is very similar to the stage type classification that is often provided by experts in the route books of stage races. Our model replicates this expert ranking solely based on a GPX file as input source. The first step enriches the GPX file with elevation data from a Digital Elevation Model (DEM) to guarantee elevation consistency across GPX files of different sources with different elevation measurement techniques. Based on the produced list of coordinates with its annotated elevations peak analysis is performed. Our algorithm calculates peaks and valleys of the course (see Figure 5.13, peaks calculated for stage 8 of the 2020 Tour de France). The height of the detected peaks is calculated by the peak prominence principle. Topographic prominence is defined as the minimum height to descend from a peak to reach a point that is higher than the current peak [8]. The higher this prominence value, the more important the peak is. This principle can be further illustrated with a small example: if you were to ride to the summit of Alpe D'huez, in France, you would stand at an elevation of 1840 meters, but you would have climbed only 1117 meters of prominence (starting from le Bourg-d'Oisans). This is because the surrounding Alps are all tall mountains, so you would have started at a much higher elevation than if you had climbed up from sea level. The Scipy signal Python library was used to find peak prominence. With this library the peak finding algorithm can be fine-tuned by defining a minimum required distance between subsequent peaks and a minimum required prominence to be considered as a peak. Within the context of a grand tour (UCI category for multi-day stage races) values were empirically set at 30,000 meters for minimum distance and 40 meters for minimum required prominence values. Valleys were found with the same method, but by inverting the elevation profile of a route.

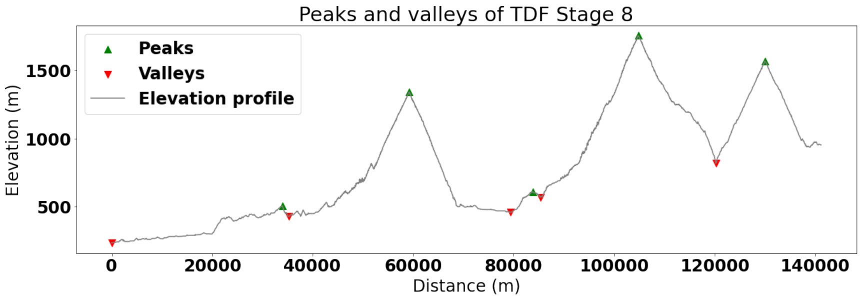


Figure 5.13: Calculated peaks and valleys on the elevation profile of stage 8 of the Tour de France.

With the calculated peaks and valleys, a scoring function can be implemented

Algorithm 5.2 Course difficulty scoring

1: $D_{1,2,\cdots,l} \leftarrow$ accumulating course distance, $D_l =$ the length of the course
2: $E_{1,2,\cdots,l} \leftarrow$ elevation at point by index of the course
3: $P_{1,2,\cdots,n} \leftarrow$ indexes of n detected peaks
4: $V_{1,2,\cdots,m} \leftarrow$ indexes of m detected valleys
5: $score \leftarrow 0$
6: $i \leftarrow 0$
7: $j \leftarrow 0$
8: **if** $P_1 < V_1$ **then**
9: $\quad V.prepend(1)$ ▷ Algorithm expects valley to occur before first peak
10: **while** $i \leq n$ & $j \leq m$ **do**
11: $\quad score \leftarrow score + [E_{P_i} - E_{V_j}] \cdot e^{\frac{D[V_j]}{D_l}}$
12: $\quad i \leftarrow i + 1$
13: $\quad j \leftarrow j + 1$
14: $score \leftarrow \frac{score}{D_l}$ ▷ Scale by course length to make results comparable

to represent the difficulty of a stage based on elevation data. The pseudo-code of this algorithm can be consulted in Algorithm 5.2. The final score is at its core a sum of the differences between subsequent peaks and valleys. As illustrated in line 11 of Algorithm 5.2, the difference between a peak and its preceding valley (i.e. the climb) is exponentially weighted based on ratio of distance completed and total course distance. To make the results comparable with other courses of various lengths, the score is divided by the length of the course (see line 14, Algorithm 5.2).

5.4.2 Bunch sprint prediction

A key step in the geospatial processing pipeline is the calculation of the likelihood of a bunch sprint. Bunch sprints are defined as a bigger group of riders that sprint towards the finish line for the victory. The course hardness already reveals a part of the picture. As studied in the previous subsection, races can be categorised based on the elevation profile. An extra check that can be made is the probability for a bunch sprint. This is performed by a classifier that predicts if a bunch sprint is likely to happen. The classifier uses a combination of properties which are calculated based on the elevation data (e.g., peak features similar to previous method, distribution of grade percentages and distribution of elevations) and the type of race (e.g. World Tour, U23, Pro Continental or Regional races). Race information (race, stage, country and results) was scraped from the UCI website for the seasons 2018-2019. GPX course files were also scraped from the website of organisers and other GPS file sharing websites such as "RouteYou", "Komoot", "RideWithGPS" and "La Flamme Rouge". GPX files and results were fuzzy matched using the Fuzzywuzzy Python package based on the race name, stage number and date. From the results, the number of people that finished in the first group were calculated. People are in the same group if the time gap between them is not more than 2 seconds. The first group can also consist of a single rider if he/she finished solo. The ratio between the number of riders in the first group and the total participants was considered as the ground truth for our training dataset. In agreement with the international cycling federation (UCI), if more than five percent of the total participants were in the front group, the race was labelled as "ending in a bunch sprint". This resulted in a training set of 1241 races that were further split for testing and training (with a train/test split of 5%). To get a balanced dataset, the 640 races ending in a bunch sprint were under sampled such that the training set contained 538 samples of both classes. The best results were obtained with a RandomForrestClassifier (using 5000 estimators, from the sklearn.ensemble package) and resulted in an accuracy score of 83% and an F1-score of 82%.

As mentioned, the combination of the type of the course and the likelihood of a bunch sprint can further optimise our road safety scoring mechanism. For instance, the scoring mechanism will need to focus more on the descents in mountainous stages but will pay more attention at the final kilometres of flat stages when the sprint probability is high.

5.4.3 Turn detection

Turns are defined as direction changes in the road. For cycling safety it is very valuable to detect and localise sharp turns on the course. Especially turns that occur in descents or in the last few kilometres of a race. In this section a mechanism to detect direction changes in roads is presented.

The angle of a turn is defined as the angle between two direction vectors on the course. Direction vectors are constructed by applying linear regression on n subsequent points. This provides a line that is a best fit for these n points [9].

This principle is illustrated in Figure 5.14 where regression lines were calculated for points P_i and P_j (Figure 5.14, step 1). Each regression line was constructed by points P_i and P_j and its two preceding points. This also means that point P_j is three indexes further than point P_i. The second step is projecting two arbitrary points per produced regression equation which can be used to produce two vectors (Figure 5.14, step 2). In the final step (Figure 5.14, step 3), both vectors are used to calculate the angle between both points on the course, providing if

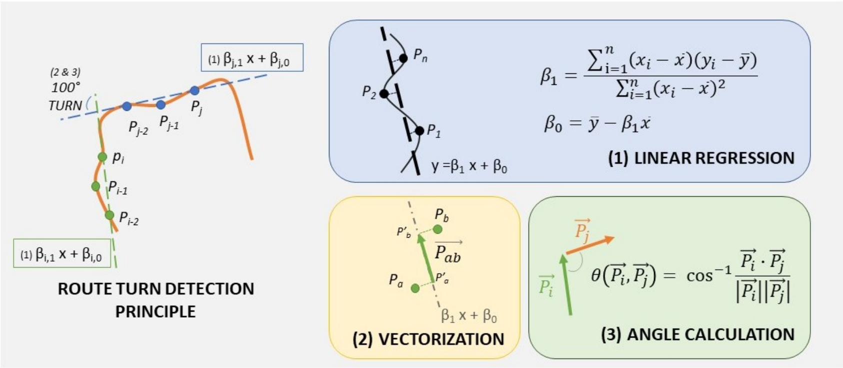


Figure 5.14: Overview of the turn detection mechanism which consists of linear regression and vectorisation of 2 points on the course with subsequent angle calculation based on the vectors.

and how much the road is bent between P_i and P_j. When the route of a race is iterated in such a fashion, fast direction changes can be detected. Slow direction changes, however, are a bit less obvious to detect. The difference between fast and slow turns are illustrated in Figure 5.15. The principles introduced in Figure 5.14 can be directly used to find the fast direction changes. Slow direction changes can be found in a very similar way. The major difference is that for a slow direction change the previously detected turn point (Figure 5.15, b), rather than the subsequent points on the course (Figure 5.15, a), is used as the first vector.

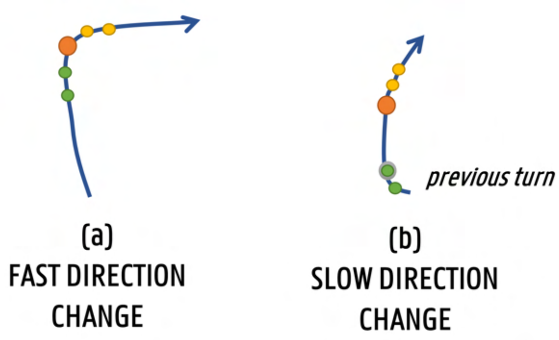


Figure 5.15: Fast direction change with the subsequent points used for turn calculation (a) and slow direction change with current point and previous detected turn used for the slow turn calculation (b). Green points are the points used for the first regression; yellow points are the points for the second regression.

5.4.4 OpenStreetMap course segment analysis

The scoring mechanism exploits the available data from the collaborative geospatial dataset OpenStreetMap (OSM), which has more than a million contributors. Contributions can range from adding the opening hours of a bakery to the creation of a new road. For the scoring methodology we are mainly interested in the road network which is also well documented in OSM.

The first step consists of processing the course path into a list of coordinates that are used in the later analysis. The scoring mechanism requires a "computer-friendly" version of the course. Usually the course is available as a GPX file, but sometimes only a detailed PDF which describes the route in detail and the approximate time the peloton passes at a described point on the course. The latter needs an additional, manual effort to convert what is basically a list of streets and municipalities on the course to an actual list of coordinates (i.e. the course of the path). If required, the list of coordinates can be further trimmed down to the last couple of kilometres or the descents for instance (based on the course type and sprint probability, see sections 5.4.1 and 5.4.2). In the following step, the required metadata for the safety scoring mechanism is calculated and/or gathered. This metadata was partly calculated using algorithms and partly from OpenStreetMap metadata. The calculated metadata is turn and elevation information. Turns are defined as a change in the travelled direction. The angle of a turn is defined as the angle between two direction vectors on the course. Direction vectors are constructed by applying linear regression on a number of n subsequent points. This provides a line that is a best fit for these n points.

Listing 5.4: Python implementation of the Overpass Query (using the overpy Overpass interface package) that retrieves the closest OSM way to the provided latitude and longitude within a certain radius of that point.

```
import overpy

def get_way_data(lat, lon, radius):
    overpass_api = overpy.Overpass(url='https://
        ↪ overpass.kumi.systems/api/interpreter')

    result = overpass_api.query("""
            way
            (around:{},{},{})
            ["highway"~'.*'];
            (
              ._;
```

```
            >;
          );
          out;
      """.format(radius, lat, lon))
    ways = []

    geo_node = [None, None, None, None, None]

    if len(result.ways)>0:
        way = result.ways[0]
        geo_node = [way.id,
                    way.tags.get("highway", None),
                    way.tags.get("maxspeed", None),
                    way.tags.get("junction", None),
                    way.tags.get("lanes", None)
        ]

    return geo_node
```

The input source for the geospatial segment analysis is the OpenStreetMap database. For this part we used the OSM Overpass API, which allows easy querying of the underlying OpenStreetMap database. The query language used to retrieve information is the Overpass Query Language (Overpass QL) and the produced results can be formatted in JSON, CSV or XML. A full explanation of how data is represented and stored within OSM would lead us too far, but its most important data types for our analysis are nodes and ways. A node is used to mark a single geographical point and provide information about that point using tags. Ways are collections of nodes and can have tags to describe the way as well. For the scoring mechanism we search the OSM database for roundabouts, the road type, possible speed restrictions and traffic infrastructure in the neighbourhood of a coordinate (i.e. latitude/longitude pair) on the race course. Code listing 5.4 shows an example Overpass OSM query that searches the closest way near a coordinate. The data for our route score mechanism is provided as tags embedded within the returned way object.

The coordinate pairs are equally distributed in segments that have the same length. In the current version a segment is one kilometre long and Overpass queries are performed every 100 meter on this segment. With the results of the Overpass queries a segment risk score can be calculated. A schematic overview of this methodology is presented in Figure 5.16.

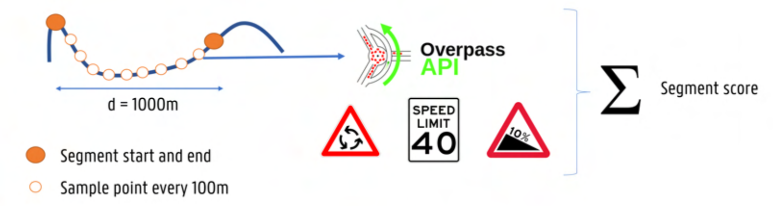


Figure 5.16: Schematic overview of the segment risk estimation along a course.

The segment risk estimation is basically a weighted sum of the different traffic infrastructure elements that can be found in the sampled course points via the Overpass API. Roundabouts are "penalised" with four penalty points. If the maximum allowed speed on the road is lower than 40 kilometres per hour, we also add two penalty points to the overall segment penalty. Furthermore, between every two subsequent points on a segment, the number of changes in the OSM characteristics is counted and added to the overall segment risk score. Finally the overall score is balanced by the elevation difference per segment. The overall score of a segment gets reduced if a segment goes uphill, and the overall score gets increased if the segment goes downhill. For this purpose, the grade of the segment is calculated. Every percent of inclination is worth one penalty point (or penalty reduction if the segment goes uphill).

This relatively simple, yet effective mechanism is able to provide an initial screening of cycling race courses. In Section 6.3, we will further illustrate of how this mechanism is used in the overall course safety screening procedure.

5.5 Conclusion

Geospatial data is arguably becoming one of the most valuable and important data sources of the data available in the sport. Its main strengths are that actions/performance or story telling elements can be directly linked to a location. The weaknesses are actually concealed in its strengths, i.e., the geodata often has to be linked to provide valuable insights. On its own, geodata is not always the most interesting data, but when it is linked with another data source its real strength shows. In Chapter 6, this exact linking of data is further discussed with the help of multiple use cases that use different sources of data to solve a problem. The UCI use case (Chapter 6, Section 6.3) is an example where geodata is used to highlight dangerous points on a course that the cyclists have to cover. Furthermore, in Chapter 6, Section 6.4, information about where a rider is on a cyclocross course is used to provide sensor-driven insights to the TV spectators.

References

[1] X. Li, M. Kane, Y. Zhang, W. Sun, Y. Song, S. Dong, Q. Lin, Q. Zhu, F. Jiang, and H. Zhao. *Circadian rhythm analysis using wearable device data: A novel penalized machine learning approach.* bioRxiv, 2020.

[2] G. Francis. *Strava – Automatic Lap Detection*, Aug 2018. Available from: https://science4performance.com/2018/08/04/strava-automatic-lap-detection/.

[3] R. E. Johansson, S. T. Adolph, J. Swart, and M. I. Lambert. *Accuracy of GPS sport watches in measuring distance in an ultramarathon running race.* International Journal of Sports Science & Coaching, 15(2):212–219, 2020.

[4] J. De Bock, S. Verstockt, et al. *GPS driven camera selection in cyclocross races for automatic rider story generation.* In icSPORTS 2020, the 8th International Conference on Sport Sciences Research and Technology Support, pages 67–74. SCITEPRESS, 2020.

[5] B. Hess, A. Z. Farahani, F. Tschirschnitz, and F. von Reischach. *Evaluation of fine-granular GPS tracking on smartphones.* In Proceedings of the First ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems, pages 33–40, 2012.

[6] P. Ingole and M. K. Nichat. *Landmark based shortest path detection by using Dijkestra Algorithm and Haversine Formula.* International Journal of Engineering Research and Applications (IJERA), 3(3):162–165, 2013.

[7] J. De Bock and S. Verstockt. *Road cycling safety scoring based on geospatial analysis, computer vision and machine learning.* Multimedia Tools and Applications, 82(6):8359–8380, 2023.

[8] A. Kirmse and J. de Ferranti. *Calculating the prominence and isolation of every mountain in the world.* Progress in Physical Geography, 41(6):788–802, 2017.

[9] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to linear regression analysis.* John Wiley & Sons, 2021.

6
Multimodal data analysis in cycling

"Ubi concordia, ibi victoria."

–Publius Syrus (1st century BC)

This chapters brings together each of the individually discussed analysis modules from the previous chapters and shows how to combine them into meaningful sports analysis scenarios in different use cases. The research question for this chapter is: **"Can multimodal data mining improve performance, safety and storytelling in professional cycling?"**.

6.1 What is multimodal data analysis?

As a start, the term "multimodal data" will probably need some clarification. Multimodal data fusion aims to fuse the data of different sources, types and distributions and ultimately represent them in a uniform manner [1]. Within this dissertation, we defined three main data sources: "video, text and (time-stamped) geospatial data". Within this chapter we will introduce some use cases that used multiple from the aforementioned data sources with the goal to unify the data in such a way that it contributes to one of the three domains that were targeted in this dissertation (i.e., safety, performance or storytelling).

6.2 Case study: Track cycling performance analysis

6.2.1 About the project

The Flemish and Belgian cycling federation are catching up rapidly in the track cycling discipline with the popular track nations (e.g., Great Britain, Australia and the Netherlands). Through the years, a good coaching staff and talent detection mechanism have been developed. Within cycling, and track cycling is no exception to this, a lot is planned out in advance. Training sessions, race tactics and pre-race fuelling are just a few examples. However, and in contrast to road cycling for instance, UCI rule 3.2.005 requires that *"any electronic device with display (for instance speedometer or powermeter) must be hidden so that it cannot be read by the riders"*. However, the rules do not stipulate that coaches cannot see the data in real-time. This data is definitely useful and important for coaches in some of the disciplines. In the team pursuit discipline for instance, it is crucial to have an optimal pacing plan and use every athlete in the pursuit train in an optimal way. The aim of this project was to build a software platform that is able to capture, display, store and analyse performance-related sensor data of athletes on the track using the ANT+ signal.

6.2.2 The setup

Modern fitness sensors can transmit data wirelessly and via a standardised protocol from the sensor to the receiving devices. Most off-the-shelf sensors now transmit data via either ANT+ or Bluetooth Low Energy (BLE). The ANT+ network is best suited for our goal as it does not require pairing and it can be received by any device that are listening on the ANT+ network, which operates at the 2457MHz (i.e., 2.4Ghz) radio frequency. This frequency is the same as the WiFi 2.4Ghz frequency, so interference with the ANT+ signal is not impossible. The 2.4Ghz band is so popular for protocol adoption as it uses rather long waves, which offer relatively long signal range through obstacles. Furthermore, ANT+ does not use encryption, as it is designed to be observable by any device in the network, which is actually beneficial for the proposed use-case. To capture the signals that are sent from the sensors we rely on North Pole Engineering's (NPE) WASP devices [1]. The devices basically capture all ANT+ data that's in the neighbourhood and puts it into a local area network via WiFi using multicast messages. The more WASP devices that are positioned along the course (i.e., the cycling velodrome), the better coverage we will have for the area. For each WASP that was able to capture a specific sensor's data, its received signal strength indicator (RSSI) is reported so

[1] https://npe-inc.com/wasp-3/

this also gives us an idea of how close a sensor might be to a specific WASP in our topology. WASP data packets are captured by a C# (.NET) application which we call the data aggregator in our setup. Figure 6.1 shows the overview of our Wireless Cycling Network (WCN) setup.

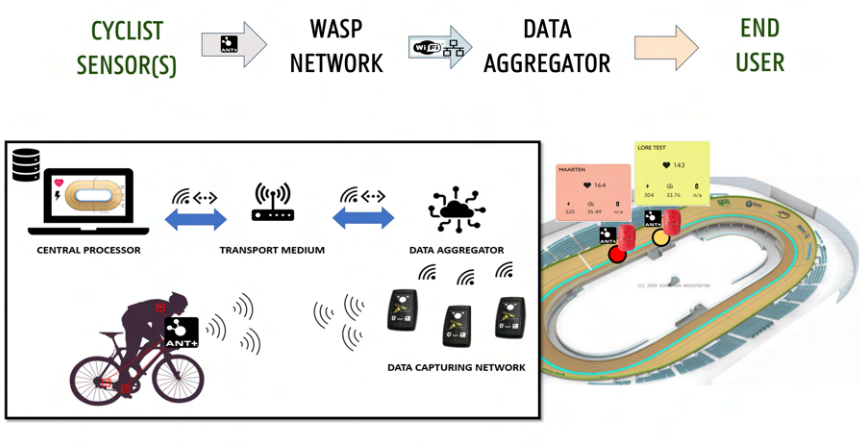


Figure 6.1: Schematic overview of the WCN setup on the cycling track.

6.2.3 Applications with the setup

6.2.3.1 Automatic handsling detection

Remark: this section is a partial rework of the publication "Sensor-Based Performance Monitoring in Track Cycling" [2]

Among the different track cycling disciplines, madison belongs to the relay category. A madison race consists of multiple teams composed of two riders. At all times, one rider of each team is considered to be actively racing and typically located on the lower part of the track. This rider is denoted as the active rider in what follows. The other rider, denoted as the inactive rider, is located on the upper part of the track, riding with much lower speed and waiting for the active rider to catch up. Once the active rider of a team catches up with the corresponding inactive rider, the inactive rider steers down towards the active rider. When the active and inactive rider are located next to each other, the speed of the active rider is transferred to the inactive rider by means of a handsling. This event is often called a change.

When it comes to the racing format, the official madison race distance as defined by the International Cycling Federation (UCI) is 50 km, i.e., 200 laps on a 250 m track. Every tenth lap, the first active rider of a team at the finish line earns five points, while the second, third and fourth rider earn three, two and one points respectively. One exception is the final lap, in which the rewarded points are doubled. In addition, a team can escape the front of the bunch and catch up with the back of the bunch, which is called lap gain. In this case, the team is rewarded with 20 additional points. Of course, if a team loses a lap by getting dropped from the bunch, the team loses 20 points. The goal is to gain as much points as possible by the end of the race as a team.

In current analysis of this discipline, required data is typically extracted from data files generated by a cycling computer. After the data collection, handsling events are manually extracted from the file and the corresponding data is interpreted by a coach. With the help of the data that comes automatically in the WCN platform, this process can be significantly speed up and a lot of the workload that is required from coaches or athletes can be eliminated. With accurate timestamps of when and how long a handsling takes, performance monitoring of madison handslings is the logical next step, objectifying the current feedback provided to the riders and allowing a coach to only perform an in-depth analysis of handslings that require additional attention. As a proof of concept, two training sessions were recorded for a madison rider pair of professional, experienced and beginner level, thus six training sessions in total.

When it comes to track cycling, the bikes have a fixed gear. This causes the cadence data only to be dependent on the speed of the rider and the gear installed on the bike. Because speed is often recorded using GPS signals, which are often of low quality on an indoor track, solely cadence data was used for madison handsling detection in this study. For simplicity, data was recorded from rider pairs using the same gear. Consequently, the absolute values of cadence can be compared in a meaningful way. When different gears are used, cadence values can be converted using a gear ratio chart.

Using the properties of the madison discipline, it can be assumed that the cadence values of the active and inactive rider only intersect when a handsling occurs. However, small deviations from this assumption can easily be captured by a check for cadence values not to cross twice in a limited amount of time. Detection of madison handslings can now easily be implemented by searching for intersections in cadence values (C in Equation 6.1) throughout the training session. Each intersection can be associated with a handsling, where the rider with a downward trend in cadence becomes the inactive rider after the handsling and the rider with an upward trend becomes the active rider. More formally, timestamp t (Equation 6.1) corresponding to the cadence value (C, Eq. 6.1) intersection point can be described as:

$$\exists\, t : C_{t-1}^{inactive} \leq C_{t-1}^{active} \wedge C_{t}^{inactive} \geq C_{t}^{active} \tag{6.1}$$

Detecting madison handslings and assigning them with a fixed duration will capture the entire handsling only when the fixed duration spans the duration of the effective handsling. On average, a duration of five seconds suffices, but in a more ideal case, it is possible to detect the start and end time of a handsling and derive the duration dynamically. Important to see is the increasing cadence of the inactive rider through the intersection, while the cadence of the active rider decreases. Defining an appropriate threshold over the rolling covariance between these two cadence time series allows the extraction of the start and end time of a madison handsling in a rule-based manner.

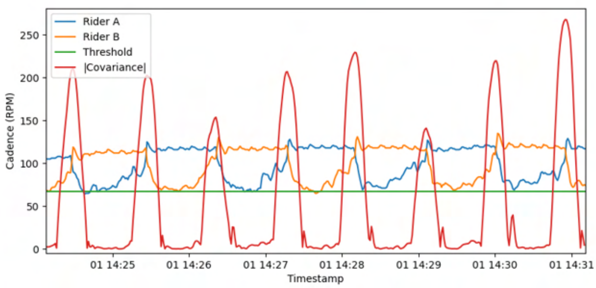


Figure 6.2: Cadence values of both riders during a training session and the corresponding covariance and threshold used to detect handslings.

Finally, patterns in performance data, more specifically cadence data, can be exploited when detecting madison handslings. Calculating the rolling covariance over 25 seconds (performance data is typically recorded at 1 Hz), Figure 6.2 shows how handslings can be detected by setting a threshold at 25% of the maximum rolling covariance value. The intersections of the threshold with the rolling covariance denote the start and end timestamps of the handslings within a training session. This approach yields perfect accuracy for all training sessions where cadence values only cross during a handsling. By filtering out all cases where the cadence of the inactive rider peaks above the cadence of the active rider for at most two seconds and only using intersection points of at least 80 RPM, all false positives were eliminated from the collected training session data. Such filtering is required for example at the very beginning of a training session where riders are riding next to each other at the same speed, before starting the actual madison training session.

Ideally, the average active duration of both riders in a rider pair is approximately equal. Nevertheless, during a training session or race, the duration be-

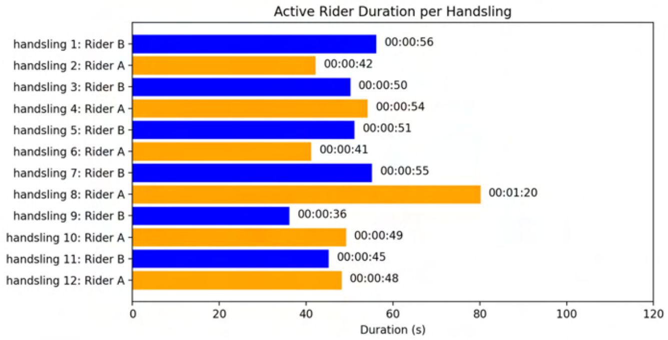


Figure 6.3: Duration as active rider during a madison race simulation.

tween handslings will slightly vary. This might for example be the case during a sprint lap, where the goal is that one rider takes the majority of sprints, due to its sprinting abilities. Thus, race situations should be taken into account when comparing active duration of riders. The influence of a sprint lap is illustrated by handsling 8 in Fig. 6.3. Just before the sprint lap, Rider B accelerates, in order to save energy for the sprint and become active at one lap to go in the most ideal scenario. This leads to a long duration for Rider A as active rider. Finally, generating these timestamps and durations of handslings can yield very useful performance metrics.

Now we have detected the handslings in the data, our focus can be shifted to some statistics based on power data. By calculating the average power of riders in a time range around the intersection timestamp t of cadence values, the exerted power during a handsling can be measured. This is shown in Table 6.1. From the comparison it is clear that Rider C spends less power during the $[t-2s, t+2s]$ intervals, both as an active and inactive rider. Nevertheless, the higher values of Rider B and Rider D might originate from their more explosive rider type, meaning they can more easily reach higher peak power values with less effort. More important is the difference in power values as active rider, compared to being an inactive rider over the entire handsling. Here Rider C achieves similar power values, whereas the values for Rider B and Rider D significantly differ. When using the $[t-5s, t]$ interval indicating the exerted power for the inactive rider just before the handsling, it becomes clear that only for Rider C, this exceeds the average power during the $[t-2s, t+2s]$ interval as inactive rider. This leads to the conclusion that Rider C exerts too much power just before the handsling, not optimally using the gradient of the track to gain speed. Once the handsling effectively takes place, the rider already gained sufficient speed, causing a less efficient transfer of speed from active to inactive rider.

Table 6.1: Comparison of average power for different handsling time intervals.

	Handsling Active	$[t-2s, t+2s]$ Active	Handsling Inactive	$[t-2s, t+2s]$ Inactive	$[t-5s, t]$ Inactive
Rider B	149 W	261 W	249 W	373 W	235 W
Rider C	192 W	199 W	174 W	201 W	271 W
Rider D	160 W	232 W	221 W	289 W	227 W

6.2.3.2 Automatic handsling detection in video

Next to the sensor data handsling detection algorithm (see section 6.2.3.1), we can also find the handsling movements in videos to assess the quality of the handsling. Experienced coaches can easily tell the difference between a well-executed handsling and an inferior one. Currently, their workflow involves a lot of manual browsing in the recorded video footage (if they already record such videos). With our setup, we try to further assist coaches in automatically detecting handslings in video images to provide them with clips of the action, overlaid with the captured data by the WASP setup (see section 6.2.3.1). This methodology exists of two consecutive steps: in the first step we pre-filter the long video sequences with the aforementioned sensor-driven methodology and in the second step we try to find the exact video frames in which the handsling sequence is actually visible in the video. These short clips can then be presented to the coaches.

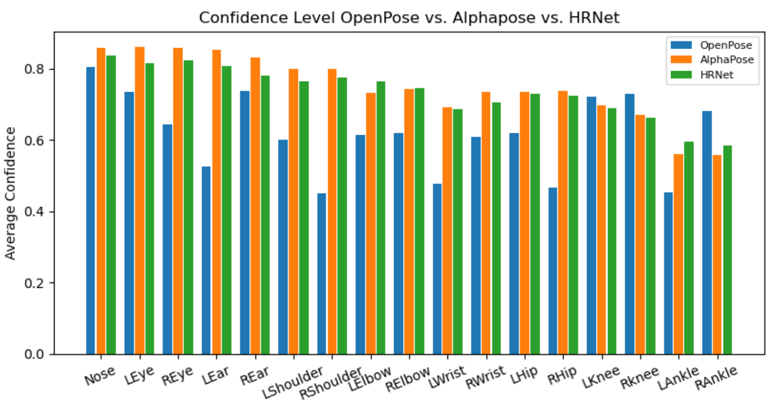


Figure 6.4: Average keypoint confidence comparison between AlphaPose, OpenPose and HRNet.

In order to qualitatively assess the handslings, coaches want to have camera recordings. When a motorbike is riding in front of the riders, it offers them the best visiblity of the handslings, but it is only possible during training sessions. To detect the handslings, we selected and compared three different pose estimators: OpenPose [3], AlphaPose [4] and HRNet [5]. Figure 6.4 shows the average confidence of multiple body joint location predictions for the three selected human pose estimators during a madison training session. These frames were captured using a GoPro action camera at a resolution of 720p and a frame rate of 60 fps. The first conclusion that can be drawn is the under-performance of OpenPose compared to AlphaPose and HRNet. The only exception is the level of confidence for the lower body joints, i.e. knees and ankles. Keeping our goal in mind, the most interesting body joints are shoulders, elbows and wrists, where AlphaPose and HRNet perform significantly better. For the most important body joints (also called key points in the pose estimation model terminology), AlphaPose slightly outperforms HRNet in 67% of the cases. The easy integration of AlphaPose with PoseFlow [6], allows the tracking of skeletons. As this makes it easier to process the obtained keypoints, this feature reinforces the choice for AlphaPose over HRNet.

In the previous paragraph we selected the AlphaPose detector to extract the relevant body joints for the riders performing the handsling motion. The next step consists of an in-depth analysis of the handsling motion. The first step towards such an analysis is the detection of these handslings in the entire video sequence of the training session. An intuitive way to achieve this, is by using the keypoints extracted from the video sequences captured by action cameras attached to the back of a motor or the helmet of the driver. A madison handsling can only occur when the hands of both riders touch. This means, the distance between both hands (i.e., the two dimensional coordinates corresponding to the body joints of the wrists) should be 0. The distance is measured with the Euclidean distance metric. This is a an oversimplification from the reaility, as in 3D this distance can be 0 even without the hands touching each other. It is however true that as soon as hands of two different bodies (i.e., riders) come close that they are actually starting their handsling. Taking into account the inaccuracies coming with pose estimation, a threshold D should be imposed. When the distance between the left wrist of the inactive rider and the right wrist of the active rider falls below this threshold, a handsling should be detected. Note that using only D as threshold to detect a handsling does not suffice. Another consequence of the same inaccuracies, is the risk on false positive handsling detections, because of a single outlying estimation for a body joint location corresponding to a wrist. Therefore, a minimum on the number of consecutive frames for which the distance between the wrists of both riders is smaller than D should be chosen. This parameter can be denoted as T. Note that there is a trade-off when choosing a value for T.

Figure 6.5: Keypoint-based handsling detection using distance between wrist keypoints.

When applying this approach to one of the videos that was captured during a madison training session, it is clear that handslings can potentially be detected with high accuracy. Using the video footage captured from the back of the motor, the value for *D* can be fixed at 100 pixels. Fixing the value for *T* at 20 means that keypoints corresponding to the wrists should be less than 100 pixels apart from each other for at least 20 frames. An improvement of this parameter *D* could be by setting it at a percentage of the poses dimensions as it would rule out potential variations of the distance between the riders and the camera on the motorcycle. Within these 20 frames, a predefined number of frames *O* can contain a distance between both hands larger than *D*. For the experiments the *O* parameter was set at 5. By setting these values, it is possible that more than 5 frames contain values not fulfilling the threshold imposed by *D*, for example when wrists are occluded during the handsling. To mitigate the segmentation of a single handsling, detections within a limited time span can be merged. When defining the average duration of a handsling at 5 seconds and the time between handslings in the range of $[45, 75]$ seconds, it is clear that detections will potentially occur in intervals, each 45 to 75 seconds, for a duration of approximately 5 seconds in total. This is because the recordings are from a single madison team and they change every other 45 to 75 seconds. The result of detecting handslings using all described parameters on video footage captured from the back of the motor is displayed in Figure 6.5a. The analysed video contained 13 handslings, of which all 13 were detected by the described approach. The same parameter values can be used for a training session that was captured from the helmet of the driver. The result for a video containing 10 handslings is shown in Figure 6.5b. As three false positives and only 7 out of 10 handslings are detected in this case, it is clear that the parameter values are susceptible to the angle from which the handslings are captured. This also explains why the camera fixed with a mount on the motorbike outperforms the helmet camera (i.e., the helmet camera is not perfectly still as the driver is always slightly moving his/her head).

As as final step, the data of both the sensor data capturing platform and the video handsling detector can be merged based on the timestamp of both the sensor data and the video clips. The identification of riders in video can either be performed manually, or by positioning the camera at a MyLaps measurement loop. The fusion of both data sources provides the coaches a holistic overview of the handsling that took place.

6.2.3.3 RSSI-based localisation

Another potentially interesting application of the setup is the localisation on the elliptical track based on the RSSI values and how they relate to each other. For every training session, the exact location of each WASP device in the setup is recorded. This principle is illustrated in Figure 6.6, where four WASP devices (W1, W2, W3 & W4) are distributed evenly across the track.

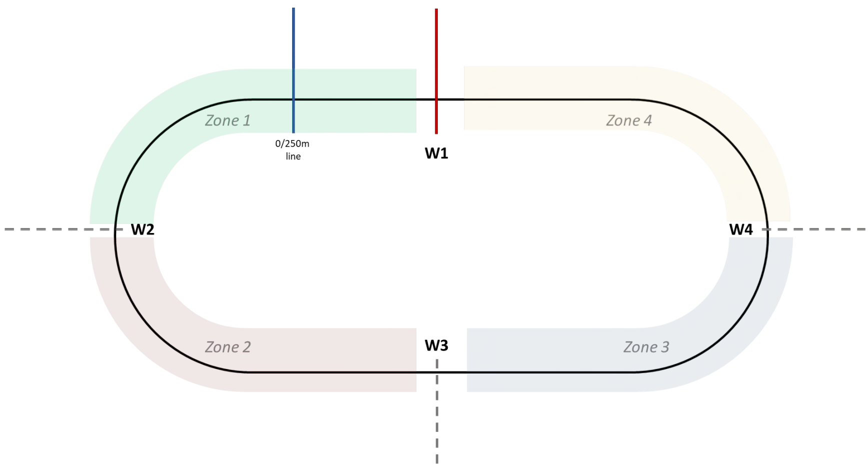


Figure 6.6: Topology of the WASP setup on the cycling track

If we now look at the evolution of RSSI values over time recorded on the 4 WASPs in the network we can clearly see peaks and troughs in RSSI over time. Figure 6.7 shows the RSSI signals of four WASPs during a 70 second recording positioned in the topology illustrated in Figure 6.6. If we use some domain-specific knowledge (i.e., an athlete's sensor RSSI will sequentially peak at W1, then at W2, and so on) within a basic RSSI peak detection algorithm, this technique is quite capable of predicting the zones where an athlete might be in (zones are highlighted on Figure 6.6 as well). In theory, this methodology can be used when the athlete only wears 1 ANT+ emitting sensor, but when he wears multiple (e.g., heart rate, cadence, power and speed), the algorithm can be extended by a majority voting mechanism. In this section we will present the RSSI-based localisation technique and compare it with the timing information of the MyLaps measurement loops.

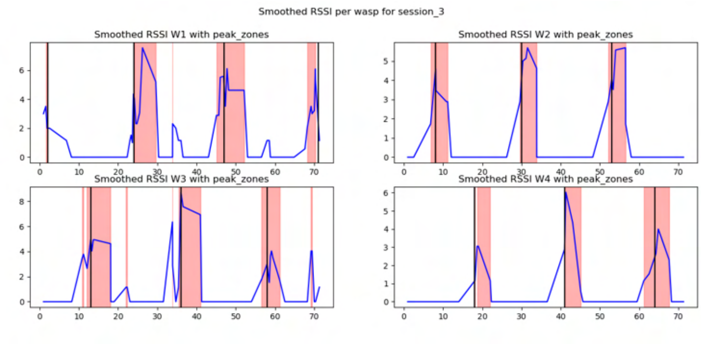


Figure 6.7: RSSI peaks of a sensor captured over time by the different WASPS in our topology

As discussed in Chapter 4, advanced timing systems found their way into (professional) cycling. The Eddy Merckx track where we implemented our WCN platform is no exception to this rule. A total of seven loops are installed on the 250m track. To validate our RSSI localisation findings we conducted an experiment where we positioned a WASP device directly on top of an RFID receptor loop. Our test rider was equipped with both a MyLaps transponder and three ANT+ sensors (i.e., heart rate monitor, power meter and speed sensor). The rider rode 17 laps on the track and his lap times based on RSSI were compared against the MyLaps lap times (which will serve as ground truth for this experiment).

As we can see in the raw results of the experiment presented in Table 6.2, we can observe that the WASP-based lap times have a tendency to over-estimate lap times (i.e., they are generally slower as the ground truth MyLaps laptimes). This is probably related to the transmission and reception capabilities of the antennas in ANT+ sensors in combination with the fact that the ANT+ is not specifically made for this purpose (rather than the UHF RFID transmitters and receptors in the MyLaps hardware). Also, MyLaps transponders are generally closer and their signal path is much more unobstructed to the loops under the track. Increased distance and a blocked pathway (i.e., by other cyclists riding "through" the signal pathway) to the receivers will undoubtedly introduce (minimal) additional delay.

If we take a closer look at the statistics in Tables 6.3 and 6.4 we can see similar trends when we leave out the last lap's outlier or when it was included. The only notable difference is that the outlier increases the relative differences between WASP and MyLaps times significantly, but otherwise all trends are similar. With the outlier removed we can observe an absolute difference of around 666 milliseconds (Table 6.3). This is very acceptable to use in scenarios where no MyLaps timing system is available or for further summarisation of the data (e.g., average power per lap on RSSI based lap timings). In our current setup at the Eddy

Merckx cycling track in Ghent, we integrated MyLaps times directly into the WCN platform as it is permanently installed and gives better accuracy than the RSSI based methodology.

Table 6.2: Comparison of the laptimes of the MyLaps timing system with our RSSI-based methodology

Lap Nr.	MyLaps time (sec.)	WASP RSSI time (sec.)	Difference (sec.)
1	34.939	36.883	-1.944
2	33.929	33.841	0.088
3	33.657	33.841	-0.184
4	30.549	30.751	-0.202
5	22.749	22.904	-0.155
6	18.859	19.457	-0.598
7	20.293	19.458	0.835
8	29.064	28.265	0.799
9	33.652	35.427	-1.775
10	33.759	32.613	1.146
11	32.761	32.918	-0.157
12	31.631	31.476	0.155
13	25.710	26.240	-0.530
14	22.474	22.902	-0.428
15	22.327	22.453	-0.126
16	25.272	26.809	-1.537
17	38.095	28.570	9.525

Table 6.3: Lap time comparison statistics with outlier (i.e., lap 17) included

	MyLaps (sec.)	RSSI (sec.)	Diff. (sec.)	Abs. diff. (sec.)
median	30.549	28.570	-0.157	0.530
average	28.8070	28.518	0.289	1.187

6.2.3.4 Real-time energy monitoring

A final application of the software platform is the real time analysis and feedback of the anaerobic energy that a rider has left during a race or training session. Back in 2012, Skiba et al. [7] came up with a ground-breaking, yet relatively

Table 6.4: Lap time comparison statistics without outlier (i.e., lap 17)

	MyLaps (sec.)	RSSI (sec.)	Diff. (sec.)	Abs. diff. (sec.)
median	29.807	29.508	-0.171	0.479
average	28.227	28.515	-0.288	0.666

straightforward model to monitor "how much gas a rider has left in his/her tank". In Monod and Scherrer [8] defined critical power as *"the maximum rate of work that it can keep up for a very long time without fatigue"*. In other words, this is the threshold at which the body starts to use the anaerobic energy system in such a way that the body cannot maintain its homeostasis (e.g., lactate building up in the muscles). The amount of work above critical power someone can do is very person-specific and is usually measured by a number of all-out efforts for varying duration. In 2021, Lievens et al. [9], set up the hypothesis that this energy reconstitution might not be constant as it was modelled in Skiba's original formula, but rather varying based on the recovery power. A new model that takes this now confirmed hypothesis in consideration is currently being developed and tested and will be included in the WCN core application. For now, we have a basic Skiba W' model implemented in the WCN platform. As the software platform was built with the aim of getting real-time data in a centralised location, this was a relatively easy addition. The only extra parameters that have to be recorded from an athlete were his/her critical power and his/her anaerobic work capacity.

6.2.4 Conclusion

In this use case the potential of using real-time data in track cycling was showcased. Centralising, showing and storing data is the first step in creating usable insights in both safety, performance and storytelling use cases. The data that was gathered in this project is multimodal as it comes from different sources (e.g., sensor data, athlete metadata or weather data). The real-time energy monitoring use case was a nice example how each of the three aforementioned types of data (sensor data, athlete metadata and weather data) was used together to provide an energy reserve for the riders. Most of the use cases with the data focused on the performance aspect, but they can be easily repurposed for story telling. For instance, the handsling insights might be also very valuable for the audience to get an idea of which couple is best attuned to each other.

6.3 Case study: UCI safety analysis platform

6.3.1 About the project

The Union Cycliste Internationale (UCI) is the world's governing cycling body that coordinates all national cycling federations across the world. The big international races (e.g., Omloop Het Nieuwsblad, Strade Bianche and Tour de France) are closely coordinated and monitored by UCI officials and commissaires. Road cycling uses regular paved (and sometimes some unpaved) roads that are designed for cars and everyday commuting. Over the years traffic calming measures made the roads safer for everyday use, but it had quite an impact on professional road cycling. A speed bump for instance might calm the speed of cars and reduce the risk for casual cyclists, but they offer a huge challenge when the bunch approaches that same speed bump in the final of a sprint-finish race.

This important fact is unfortunately also represented in the number of severe crashes that took place the last few years. A great, but unfortunate example of this statement is the crash of the Deceuninck Quickstep rider Yves Lampaert during Milano-Torino in 2020. Within the last ten kilometres, the Belgian classics rider crashed at full speed into a traffic island, causing a collarbone fracture and a couple of weeks out of competition.

These factors have encouraged the UCI to actively research and prioritise rider safety on their agenda. In their 2030 agenda, which they published and distributed on their annual world tour congress in 2022, the key action points of the UCI concerning the riders' safety are discussed in Chapter 5, section 2a [10]. One of those action points is the further development of the safety database that was developed within this project. The database documents the incidents during professional cycling races using a series of automated tools to pre-populate the database with possible crashes. These crashes can be manually reviewed by teams, riders or UCI officials. In the remainder of this section, the work that was performed for the UCI project will be discussed.

6.3.2 Safety analysis design

An overview of our methodology to quantify the safety of cycling racecourses can be consulted in Figure 6.8. The safety scoring mechanism has three big sub-parts: input, processing and output. An extra step is the feedback loop that reports gathered knowledge as input back into the mechanism. As input we use video recordings (provided by dash cameras from the organisers), course GPX files and the knowledge of what is already in the incident database (e.g., a lot of crashes happen on roundabouts). For the processing we use two different analysis tech-

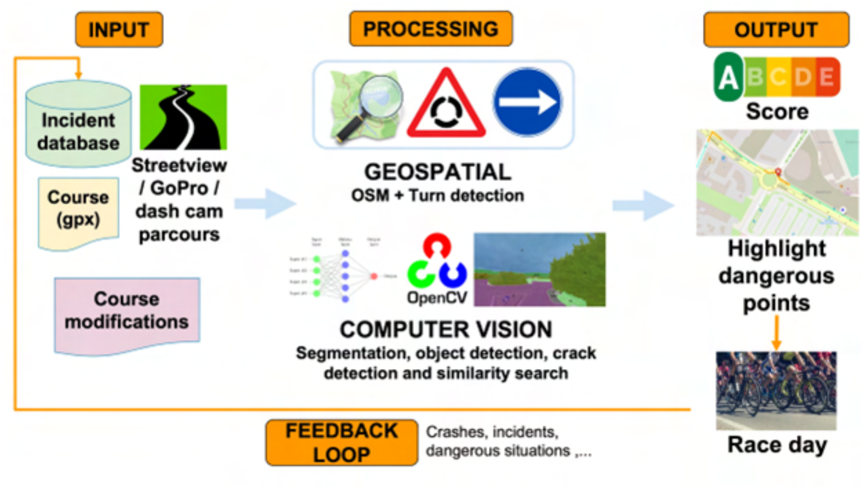


Figure 6.8: The safety methodology for improved rider security.

niques to provide safety highlights to the UCI and organisers. We investigate the provided course files with queries to geospatial databases and calculate metrics such as bunch sprint probabilities. With the provided video recordings that were also geospatially annotated, we use video processing algorithms to visually find dangerous elements on the course (e.g., potholes, poles or traffic islands). Finally we provide the UCI (and organisers) with a detailed report containing the found irregularities from the processing stage.

In the remainder of this project walk-through we will mainly focus on the geospatial and computer vision methodology to highlight risk on race courses (see Figure 6.8 for a full overview of the UCI methodology). The Tweet NLP methodology has already been discussed in Chapter 4, Section 4.4.

6.3.3 Geospatial and computer vision road analysis

6.3.3.1 Introduction

Our historical incident database revealed that traffic calming road infrastructure such as speed bumps, traffic islands and road narrowings have caused a lot of the crashes, especially when the bunch is large, and the speeds are high. Figure 6.9 shows a bar graph of the number of incidents reported per cause for the races between 2019 and 2023. In this graph traffic infrastructure is amongst the main reported causes for an incident. Other obvious ones are upcoming point-of-interest and riders' mistakes. Šegvić et al. [11] implemented a methodology that was able to detect and annotate these types of traffic infrastructure based on detected traffic signs. Another type of road infrastructure that causes disruption in the peloton is a roundabout. Most of the roundabouts are well documented in OpenStreetMap (OSM), which is a crowd-sourced geospatial database. However, the type and the complexity of the roundabout is often not documented in such databases, but this detailed information is useful to get the bigger picture. For instance, a roundabout that has multiple lanes and has an unobstructed entrance will be less dangerous for the peloton than a narrow one with traffic islands before entering the roundabout.

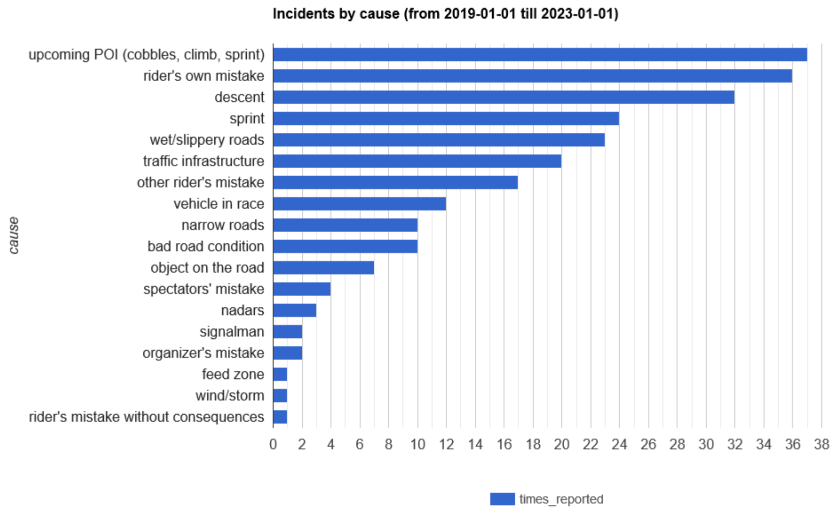


Figure 6.9: Incidents by cause from our UCI incident database (over the seasons 2019-2023).

Luckily, the limited coverage of geospatial databases can be improved by computer vision techniques, as demonstrated by Kurath et al. [12]. Road width and sudden changes in it also have a significant impact on the behaviour and fluidity of the peloton, i.e., sudden and unannounced narrowing can cause incidents in the peloton. Pixelwise semantic road segmentation and monitoring of the calculated road width over time can help to monitor these issues. This challenge is relatively well covered within the computer vision community. Wang et al. [13] [14], for example, provide a state-of-the-art implementation for the computer vision-based road segmentation task.

Another factor that is contributing to increased risk in cycling is the degree of deterioration of the road. Potholes, road cracks or longitudinal deep slits in the road can cause disturbance and crashes in the peloton. The global Road Damage detection challenge [15] is an initiative within the Computer Vision community to detect these kinds of imperfections in the road surfaces. The road damage detection challenge (RDD) is a yearly recurring challenge that offers participants a dataset containing labelled road imperfections (e.g., cracks, potholes and marking damage). Participants are challenged to find the best possible implementation to detect imperfections in the roads. The winning team in 2022, called ShiYu SeaView, used an ensemble model based on YOLO and FasterRCNN models and achieved an average score of 71.6 percent [16].

6.3.3.2 The UCI video course analysis methodology

The course analysis that is performed for the international cycling union is a semi-automated process that requires manual input and reasoning from somebody that analyses and draws conclusions from the computerized results.

An overview of our methodology to quantify the safety of cycling race courses (that builds further on the discussed state-of-the-art) can be seen in Figure 6.8. The input (i.e., course of the race, road modifications, a visual representation of the course and insights from our historical incident database) is very crucial as automatic course safety scoring is only possible if enough data is available and if it was delivered in a digital, computer friendly way. In the next step the analysis of the provided data is performed. The analyses provide a quantitative measurement of the different areas of concern when analysing the safety of a track. The analysis can be divided in two different subcomponents: geospatial analysis and computer vision-based analysis. The results/scores of both subcomponents are combined in the final safety report which is compiled based on the insights provided by the two subcomponents and is discussed with both the international cycling federation and the involved race organiser. It is worth mentioning that this final reporting step is performed by a human that performs an additional check of the raw computerised results and decides if the raised dangerous segments were already mitigated by safety promoting measures of the organisers or if this was just a “problem” on the day of the recording (e.g., car blocking the view, ongoing construction works that will be finished on race day, etc.). In the following paragraphs both the geospatial and computer vision based analyses will be further explained with the help of some examples and insights of actual courses that were analysed.

Geospatial analysis

The goal of the geospatial analysis is to convert a list of coordinates into a set of actionable safety insights based on algorithms and a geospatial database such as OpenStreetMap (OSM). An example of such an insight is the likelihood of a bunch sprint and is currently estimated using a Random Forrest classifier that has been trained on gpx files and race results of 1241 races - resulting in an accuracy score of 83% and an F1-score of 82%.

Figure 6.10: Example of the course viewer that shows the performed geospatial analyses.

An example of a race that was analyzed with our collection of geospatial tools is presented in Figure 6.10. For this race, the "Classica San Sebastian 2022' we can see that our algorithms predicts the race as very hard and not ending in a bunch sprint. Furthermore, we can also see that in the OSM based segment analysis quite some segments exceeded the threshold (the methodology was explained in Chapter 5, in section 5.4.4). The full analysis of this race can be consulted on https://courseviewer.sportsdatascience.be/#/course/62d787138f3375c224795b37. These automatic analyses of the geospatial analysis tool are used to get an idea of how closely the final kilometres should be studied (e.g., is there a downhill in the last few kilometres, does the race will end in a bunch sprint, how fast does the algorithm estimates the riders will finish, etc.). Furthermore, the segments that are above the threshold will be either studied on satellite imagery or even better, when the organiser has provided video recordings of the course, the segments will be also analyzed by our computer vision algorithms.

Computer vision

The computer vision-based analysis starts with the semantic segmentation of the recorded footage into road features using the panoptic_fpn_R_101_3x semantic segmentation model in the Detectron2 model zoo [17]. On the resulting road segments, road deterioration (such as potholes or cracks) are detected using a similar YOLO v5 model as the one discussed in [18]. Furthermore, a traffic infrastructure object detector has been developed to detect zebra crosswalks, manholes, catch basins, traffic signs, low poles and traffic island signalisation poles. The model currently achieves a mean Average Precisions (mAPs) of 17%, 54%, 45%, 83%, 92% and 94% for each of the mentioned classes. The general workflow of the scoring methodology is summarised in Figure 6.11. Our current version of the weighing mechanism is a distance-based implementation. This means that the impact of aspects such as dangerous obstacles, sudden downhills or road narrowing gets bigger as they appear closer to the finish line. It is important to mention that the scale factors are not "set and forget" parameters. Ideally, the factors are modified in such a way that the most important characteristics have the biggest impact on the overall safety penalty.

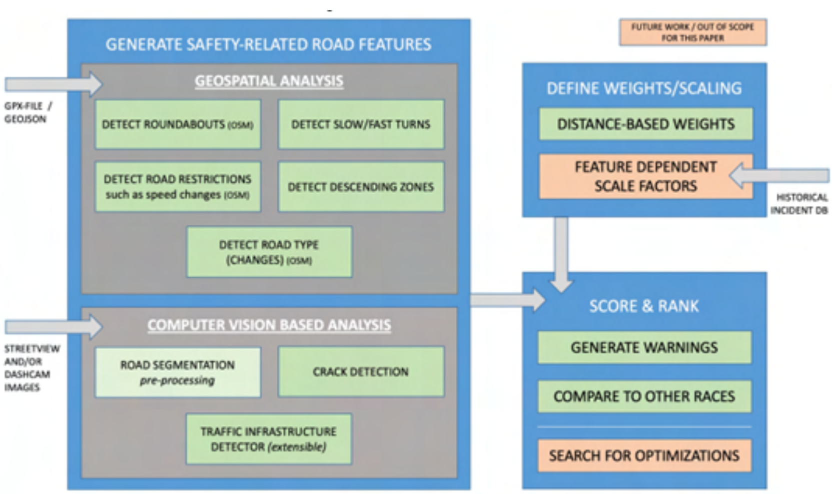


Figure 6.11: Components of the geospatial safety classification and scoring mechanism.

6.3.4 Conclusion

The incident databases serves as collection hub for all safety related incidents that occur in professional cycling racing. By collecting, documenting and reviewing the incidents, the international cycling federation gets a more complete understanding of how and when incidents happen. The reasons of incidents, for instance, can be directly used in the course analysis as it illustrates on which elements should be focused. For now, the incident database is only reviewed and modified by race commissaires, but in future iterations, other stakeholders such as riders, teams or organisers might further enrich the knowledge in the incident database.

The previously introduced geospatial and computer vision techniques become especially useful when they are combined into pre-race safety indication and post-race reporting mechanisms. The route safety score is an important measurement to compare and threshold different racecourses. Additionally, the information about which aspects made the score low or high are also interesting to report and display to the race officials and/or organisers. Furthermore, computer vision and geospatial analysis can also help to further describe the incidents that were reported in the incident database.

6.4 Case study: DAIQUIRI storytelling platform

6.4.1 About the project

The Data & Artificial intelligence for Quantified Reporting in sports (DAIQUIRI) project was a project that was achieved with close collaboration between the university, broadcasters, video capturing companies and specialist AI companies. In DAIQUIRI a media-focused sensor data platform and professional dashboard was developed, allowing content creators to augment live sports experiences. DAIQUIRI targets both real-time augmentation of live TV and near-live story snippet inserts in an interactive set-top-box application layer. The final project demonstrator showcases end-to-end sensor data integration for reporting of hockey and cyclocross. The consortium covers the full value chain bringing together unique expertise in sports event capturing (Videohouse, NEP), enriched sensor data platform (InTheRace, Arinti, imec-IDLab), editorial tooling and storytelling (VRT) and interactive user experiences in a living room environment (Telenet, imec-MICT).

Within DAIQUIRI and the scope of this dissertation there is mainly focused on two aspects. Firstly, video-sensor matching and metadata enrichment to accurately recognise riders and actions. This enrichment process should facilitate automatic video summarisation and race/match analysis. Secondly, research was performed to intelligently analyse streams, summarise them and make them semantically queryable by enhancements and annotations of both real-time and historical data.

6.4.2 Used methodologies

As mentioned in the introduction of the case study/project one of the main goals is to generate useful stories based on the raw streams of data (e.g., video and sensor information). In Chapter 3 in section 3.2.3.5 the image based ride modus detector was discussed in detail. The goal of the introduced computer vision and data analysis methodologies is to use them for story telling purposes. To further explain this intent, the ride modus detector will be further applied onto the DAIQUIRI story telling workflow. The output that the riding modus detector produces is showed in Listing 6.1.

Listing 6.1: Example output of the Ride Modus detector


```
{
    "fence": {
        "x_width": 633,
        "y_height": 257
    },
    "video": {
        "fps": 30,
        "filename": "women_r2.mp4",
        "duration_seconds": 42.0,
        "video_height_pixels": 760,
        "video_width_pixels": 540
    },
    "rider_paths": [
        {
            "pose_id": "6",
            "rider": "BETSEMA",
            "start_frame": 111,
            "last_frame": 169,
            "duration": 58,
            "begin_x": 514.0,
            "begin_y": 1.0,
            "end_x": 282.5,
            "end_y": 233.5,
            "major_ride_modus": "cyclist_riding",
            "mrm_percentage": 100.0,
            "slope": −1.0043196544,
            "direct_dist": 328.0983084382,
            "fit_params": [
                −0.0022645489,
                0.8639414568,
                158.2920890329
            ],
            "fit_error": 5762.5976800665,
            "points": [
                [
                    514.0,
                    1.0
                ],
```

```
                        [
                                514.0,
                                3.0
                        ],
                ...
            ]
        }, ...
    ]
}
```


With this information, the story telling stakeholders within DAIQUIRI can now start to create nice visualizations that can be overlayed on the main race broadcast. In Figure 6.12, such a visualization is showed. In the bottom left graph the ride lines of 3 female cyclocross riders in a sandpit are schematically represented. During this race, all three riders could ride the segment, but we can see the that rider represented by the green dots does not follow the "lines of the corner". This is also further exemplified by the segment speed that is significantly lower for the rider with the green dots. The main image in Figure 6.12 shows how the broadcasters use this information for visualisation in their live broadcast or on the social media platforms afterwards.

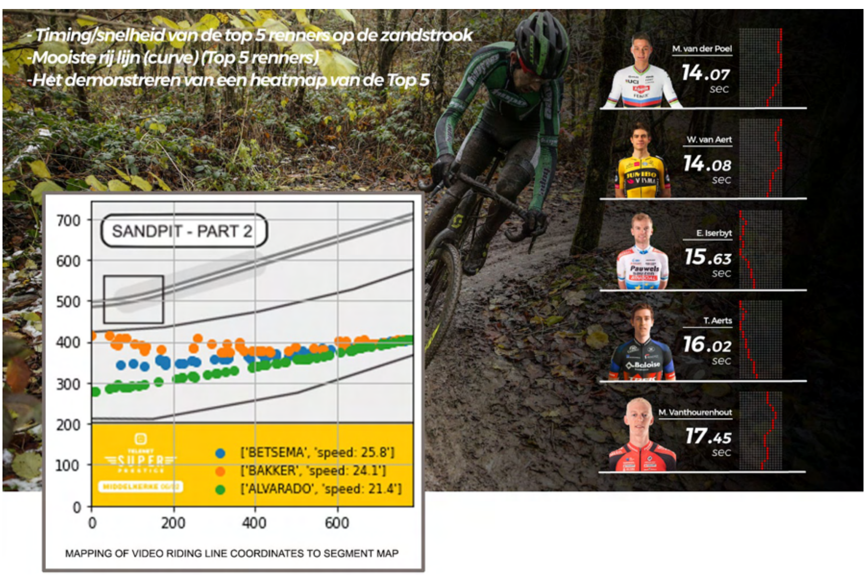


Figure 6.12: Visualization of the ride lines and mode analysis.

This approach is feasible and easy-to-use on race-day. On race-day, riders typically want to be analysed in an as unobtrusive manner as possible. For training purposes, however, we could also equip riders with additional sensors that perform the ride mode analysis in another way. Within the DAIQUIRI project, riders were also equipped with MetaMotion (MBientlab) sensors that are capable of recording high-resolution accelerometer data. Data was captured in real-time with an Android Application that communicates with the sensors and calculates the ride modi with an on-board algorithm. For this experiment multiple algorithms were empirically tested. The final algorithm that was included in the smartphone application uses a peak finding algorithm on the Fourier transformed signals on the accelerometer axis with the biggest range in accelerations measured. It does this with a time window, meaning that every three seconds the accelerometer values are collected and ran through the algorithm. However, this algorithm was not used in the final version due to a number of reasons. The first reason, the obtrusiveness, was already mentioned in the introduction of the methodology. Furthermore, a lot of parameters have to be tuned per accelerometer and rider type (e.g., not every rider produces the same forces when standing on their pedals) and finally, we only have a ride modus for every three seconds worth of data. This in contrast with the video based ride modus analyser which provides instant feedback and doesn't need any additional parameters. The only real challenge in the visual ride mode detector is the linking with the rider identity. But this challenge has been discussed in Section 3.2.2 of Chapter 3. Within DAIQUIRI, this identification task was also further simplified by the MyLaps data that was pulled into the data platform. If camera-based ride modi were analysed on a segment on the course, a MyLaps loop could be installed at the beginning of the segment, which significantly simplifies the rider and team identification of the riders detected by the ride modus algorithm. The only objecting factor against this idea is the rather expensive set-up cost of a MyLaps system.

6.5 Conclusion

In this chapter we brought all elements presented in this dissertation together in a number of use cases that highlighted the power of bringing together different data sources and analysis techniques. The combination of state-of-the-art techniques can produce often unrevealed insights in story-telling, performance analysis and rider safety. Selection and implementation of the correct data analysis techniques is where the real added value is added.

References

[1] J. Gao, P. Li, Z. Chen, and J. Zhang. *A survey on deep learning for multimodal data fusion.* Neural Computation, 32(5):829–864, 2020.

[2] M. Steyaert, J. De Bock, and S. Verstockt. *Sensor-Based Performance Monitoring in Track Cycling.* In International Workshop on Machine Learning and Data Mining for Sports Analytics, pages 167–177. Springer, 2022.

[3] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh. *OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.* CoRR, abs/1812.08008, 2018. Available from: http://arxiv.org/abs/1812.08008, arXiv:1812.08008.

[4] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu. *RMPE: Regional Multi-person Pose Estimation.* In ICCV, 2017.

[5] K. Sun, B. Xiao, D. Liu, and J. Wang. *Deep High-Resolution Representation Learning for Human Pose Estimation.* CoRR, abs/1902.09212, 2019. Available from: http://arxiv.org/abs/1902.09212, arXiv:1902.09212.

[6] Y. Xiu, J. Li, H. Wang, Y. Fang, and C. Lu. *Pose Flow: Efficient Online Pose Tracking*, 2018. arXiv:1802.00977.

[7] P. F. Skiba, W. Chidnok, A. Vanhatalo, and A. M. Jones. *Modeling the expenditure and reconstitution of work capacity above critical power.* Medicine and science in sports and exercise, 44(8):1526–1532, 2012.

[8] H. Monod and J. Scherrer. *The work capacity of a synergic muscular group.* Ergonomics, 8(3):329–338, 1965.

[9] M. Lievens, K. Caen, J. Bourgois, K. Vermeire, and J. Boone. *W'reconstitution accelerates more with decreasing intensity in the heavy versus the moderate intensity domain.* Medicine and science in sports and exercise, 53(6):1276–1284, 2021.

[10] *UCI Agenda 2030*, Nov 2022. Available from: https://www.uci.org/2030-cyclings-agenda/6IvT7HIQybfVcb2kZczQAQ.

[11] S. Šegvić, K. Brkić, Z. Kalafatić, V. Stanisavljević, M. Ševrović, D. Budimir, and I. Dadić. *A computer vision assisted geoinformation inventory for traffic infrastructure.* In 13th International IEEE Conference on Intelligent Transportation Systems, pages 66–73. IEEE, 2010.

[12] S. Kurath, R. D. Gupta, S. Keller, et al. *OSMDeepOD-Object Detection on Orthophotos with and for VGI.* In GI Forum, volume 2, pages 173–188, 2017.

[13] Q. Wang, J. Gao, and X. Li. *Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes.* IEEE Transactions on Image Processing, 28(9):4376–4386, 2019.

[14] Q. Wang, J. Gao, and Y. Yuan. *A joint convolutional neural networks and context transfer for street scenes labeling.* IEEE Transactions on Intelligent Transportation Systems, 19(5):1457–1470, 2017.

[15] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, A. Mraz, T. Kashiyama, and Y. Sekimoto. *Transfer learning-based road damage detection for multiple countries.* arXiv preprint arXiv:2008.13101, 2020.

[16] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, and Y. Sekimoto. *Crowdsensing-based Road Damage Detection Challenge (CRDDC-2022).* arXiv preprint arXiv:2211.11362, 2022.

[17] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár. *Panoptic segmentation.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9404–9413, 2019.

[18] V. Pham, C. Pham, and T. Dang. *Road damage detection and classification with detectron2 and faster r-cnn.* In 2020 IEEE International Conference on Big Data (Big Data), pages 5592–5601. IEEE, 2020.

7
Conclusion

"Orandum est ut sit mens sana in corpore sano."

–Decimus Iunius Iuvenalis (Satire X, 1st century AD)

In this final chapter we will discuss some general findings of the research performed in this dissertation. Finally, we will also point out some future work related to the introduced projects and methodologies.

7.1 Key take-aways

In this dissertation, we discussed how data and several techniques can contribute to a safer, more performant and better televised professional cycling experience. This research has led to some conclusions which will be further discussed in this final chapter.

The first key take-away is that the analysis of **professional cycling is both a multi-disciplinary and multi-modal affair**. For this conclusion, the track cycling use-case (see Chapter 6, section 6.2) can serve as a perfect example. The real-time energy model does not only rely on the input from our real-time sensor platform, but it also relies on input from sports physiologists that model the exhaustion of an athlete. Furthermore, the model and the data capturing can provide the raw values of the training sessions and the athletes' remaining energy levels, but coaches need to interpret and correlate the results with the context (e.g., type of training, periodisation of the athlete, etc.).

Another important conclusion is the fact that data on its own is just data. In an ideal scenario **data should inform the stakeholders, not (automatically) decide what to do**. Throughout this PhD thesis this was especially true in the UCI safety study (Chapter 6, Section 6.3). For this project an analysis methodology was implemented in close collaboration with UCI experts (i.e., commissaires, safety managers and riders) with the ultimate goal to aid them in their decision making whether a certain segment of a course is safe. The tool does not decide if a segment is safe, it just exploits the available data (video and GPS data) to simplify the decision making of the safety manager.

A very important lesson that was learned throughout the course of this research trajectory is the fact that **analyses should never be performed for the sake of performing analyses**. Each analysis should originate from a problem statement or research question. Within business, and sports is even more applicable to this general rule of thumb, time is money. If analyses are not practically applicable then it is a waste of time of both the time and financial investments that were made in the analysis.

Next, a useful remark to make within this dissertation is that **often there is no need to "reinvent the wheel"**. If an existing building block exists it is often more useful to re-purpose them for our specific use case (i.e., professional road cycling). The mixture of existing building blocks, however, is usually what makes the proposed solution unique and even sometimes unseen within the research domain.

Finally, we can conclude our gathered insights with the fact that the **full potential of technology and data within professional cycling has not yet been reached**! As technology, processing capabilities and overall physiological knowledge keep on improving there is still room for improvement and next iterations of the proposed methodologies within this thesis.

This final paragraph brings us seamlessly to the following section where future work, linked with this dissertation, is further discussed.

7.2 Future work

In this section we will discuss some planned future work on some of the use cases that were introduced in this dissertation.

A video-related idea for future work is to make a video fully indexed in such a way that it is easily searchable with advanced queries. An indexed video allows advanced querying of the video and is crucial for more advanced storytelling use-cases. An ultimate end goal of our research is the creation of fully automatic and personalised summaries. This principle is illustrated in Figure 7.1. The idea is to use multiple data sources combined with a preference matrix of the viewer to select the “optimal” summary of the desired duration. To reach this goal, most of the building blocks are already developed. The final steps that need to be performed are the merging of the separate building blocks and the matching of video segments with the preference matrix of a viewer.

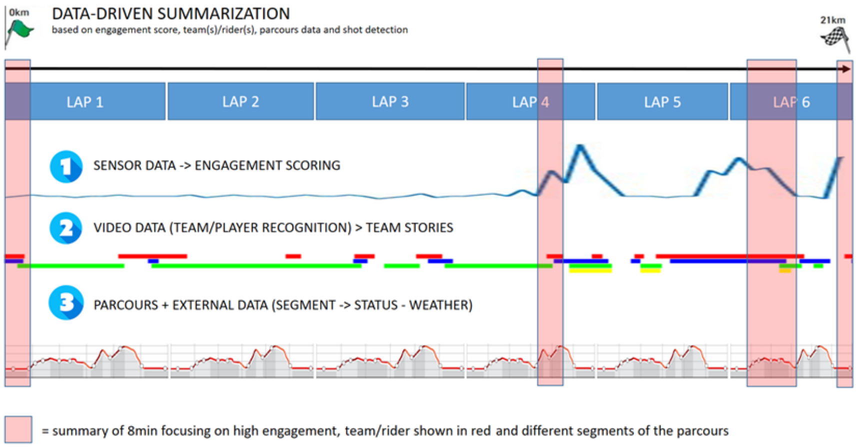


Figure 7.1: Creation of personalised summaries based on engagement, sensor data, team detection and parcours metadata

For this dissertation, most, if not all, of the use cases were targeted towards professional cycling. A next step is making the techniques, methodologies and algorithms available to recreational cyclists as well. The STRADA project, which actually builds further on the gathered knowledge of previous projects such as WCN and DAIQIURI, is an example of such a project where we are re-using our ANT+ and video processing knowledge to create personalised video clips of users passing by an intelligent video tripod.

A final future work idea is the filming workflow for race organisers. In order to perform the proposed safety analysis we need a course GPX file and a video with GPS metadata embedded. In practice, the latter is the most difficult to achieve. For our analyses we ask the uncompressed video files that are recorded with GoPro action cameras with GPS turned on. This is were most of the errors happen. Some organisers forget to turn on GPS recording, or others compress the files anyways. Transferring multiple and uncompressed files is not the most straightforward as well, so in the future a better approach should be adopted. This can be achieved in multiple ways. A first idea is to build a tool that does the merging and compressing whilst keeping the GPS metadata at the client side (i.e., on the organisers' computer). They then can upload a single, smaller video file that has the GPS coordinates embedded. This tool, combined with a detailed "how-to guide" might already solve a lot of the aforementioned issues. Another, more automated solution, is designing our own edge-computing device that performs recording, safety-screening and result processing on the device itself.