Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Curriculum in Ingegneria Biomedica, Elettronica e delle Telecomunicazioni

---

Pattern Recognition for challenging Computer Vision Applications

Ph.D. Dissertation of:
Marina Paolanti

Advisor:
Prof. Primo Zingaretti

Coadvisor:
Prof. Emanuele Frontoni

Curriculum Supervisor:
Prof. Francesco Piazza

XVI edition - new series

Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Curriculum in Ingegneria Biomedica, Elettronica e delle Telecomunicazioni

---

Pattern Recognition for challenging Computer Vision Applications

Ph.D. Dissertation of:
Marina Paolanti

Advisor:
Prof. Primo Zingaretti

Coadvisor:
Prof. Emanuele Frontoni

Curriculum Supervisor:
Prof. Francesco Piazza

XVI edition - new series

Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Facoltà di Ingegneria
Via Brecce Bianche – 60131 Ancona (AN), Italy

Life is not easy for any of us. But what of that?
We must have perseverance and above all confidence in ourselves.
We must believe that we are gifted for something, and that this
thing, at whatever cost, must be attained.

Acknowledgments

"You've got to find what you love. And that is as true for your work as it is for your lovers. Your work is going to fill a large part of your life, and the only way to be truly satisfied is to do what you believe is great work. And the only way to do great work is to love what you do. If you haven't found it yet, keep looking. Don't settle. As with all matters of the heart, you'll know when you find it. And, like any great relationship, it just gets better and better as the years roll on. So keep looking until you find it. Don't settle...

Your time is limited, so don't waste it living someone else's life. Don't be trapped by dogma — which is living with the results of other people's thinking. Don't let the noise of others' opinions drown out your own inner voice. And most important, have the courage to follow your heart and intuition. They somehow already know what you truly want to become. Everything else is secondary..."

I start my acknowledgment by mentioning a man who thought that the people who are crazy enough to think they can change the world, are the ones who do. I wouldn't be here if I didn't feel the truth of all he has said. I have always been confident and firm that this Ph.D. would allow me to achieve my personal goals and it was all I really wanted. By the way, up until then, I had always everything under control and all my life was completely organized. I had a plan for everything and everyone. I was thinking clearly on what I wanted to become. But I had not considered what this choice meant. It is hard to write about what these three years have represented. My whole life has been turned inside out, I have not a plan anymore and when I think about tomorrow I do not find answers to my million questions. But when I hold it up against the prospect of seeing the perfect life project, there's just no comparison. I had amazing experiences: I travelled a lot, I met incredible people and I spent a winter alone in Germany.

For this and much more I have to thank my two advisors Prof. Primo Zingaretti and Prof. Emanuele Frontoni.

I'm thankful to Prof. Primo Zingaretti for never say me "Good work!", for his lessons to make an excellent review and for his suggestions and advices to do valuable scientific papers. I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. He is the most picky Professor that I could find, but I'm really lucky exactly for this.

Every time, his criticism was a cause for improving and for making a great work. Now, I am pretty sure that he is proud of me even if he'll never admit it.

Prof. Emanuele Frontoni deserves my deep gratitude. He is the responsible of everything. Under his guidance I did things I would have never imagined. The enthusiasm he has for his work was contagious and motivational for me, even during tough times in the Ph.D. pursuit. It's hard enough to avoid getting into his activity for his energy, charisma and intelligence. It is also hard to want not to look like him. He has helped me get through challenges that I wouldn't even have considered. If today, I am really satisfied for this Ph.D. and I have not any regrets it is for him. Now, I got to admit: he was right all along and what I've been through is its great merit.

I would like to extend my sincerest thanks and appreciation to Dr. Adriano Mancini for the value added to the work. I have great esteem for him because he is one of the most competent professionals I know.

During these years, I divided the work with persons that could not defined only colleagues. Daniele and Annalisa have become two of my friends. Annalisa is the first one that I met and from the first time we have a particular feeling. I still remember the first day we arrived to VRAI group: we were so nervous and excited about the new experience. It was just like yesterday, instead we grew up together during this Ph.D. We were desperate enough, we have encouraged each other and we had a wonderful time in our trips: Benicassim and the unmentionable taxi, Barcellona and the terrible hostel, Dubai with the sheiks and then Pordenone and Grado with the expensive call to Trenitalia. Every place has some unforgettable moments. I know that we are completely different for some point of view, but we think exactly the same thing on other issues. In our ups and downs, we've always been there for each other. I never forgot the million of messages and the hours of calling. She is a beautiful person and a lovely traveling companion. If I think my opposite I think Daniele. We have different manners, ways of thinking and acting, but he is really important during these years. He taught me a lot and helped me many times. We had an amazing time working together. I hope he understood me like I did for him and I wish him every success with his new work because he deserves it. Anyway, I'm afraid he'll just have to bear with me this strange, otherwise, how will he do without my fashion advices?

Thanks to the others colleagues: VRAI colleagues (Chiara, Michele, Rocco and Rama) because this thesis was very much down to a team effort, the InterviewLab office mate (Paolo and Nicola) for their patience, for putting up with me and for not beating me up when I listen to Tiziano Ferro, and the Manhattan project guys (Laura, Luca, Roberto, Davide and Raffaella) for the fun in our journeys.

However, a remarkable gratefulness should be deserved to the power of: the dinamico Duo Mirco and Roberto. They are the main responsible if I feel really at home in Ancona. The last months have been wonderful and for the first time really entertaining. Mirco is the one that has given me real life lessons. He has challenged me every time to develop my talents, to become self confident and to grow up. He proved himself to be a real friend. He was sitting next to me in the InterviewLab and every time I turned around, he was there to blame me, but he was always right there just ready to listen. I know that I have a lot of work to do to feel more confident, but I will do my best to follow his own advices.

And last, but, certainly not least, It could only be Roberto. Dubrovnik was the turning point in my Ph.D. From that moment, I got to really understand what I am doing. I was so confused about my university path, but he has explained everything to me of this world. He knows, I've always tried to emulate him because I have had the greatest respect for him as a researcher. No one ever lets me talk this long: from the research on any given topic to the deepest literary works towards the fashion trends. We can correlate and extrapolate, and we can give meaning, meaning to each and every single argument. Last years, he wrote in his thesis acknowledgment that "we could converse for hours without never get bored" and it is the perfect definition of our crazy friendship. If today, I really know what I want and I really know what I like to do, it is for him and for this I'll always be grateful. We have designed an ambitious plan and up until now he is always right. So, while we're thinking about that, why don't we just go with it?

I also thank my lifelong friends, my cousins and my uncles, I will not make a list because each of them is important. By the way, it is impossible for me not mentioned Cecilia, Ludovica and Sofia. They know everything about me and they support me each day of this Ph.D. I'm sorry I haven't been around much the last years, but every time their enthusiasm for my work and their support for the challenges that I have to face were one of the reasons of my success. They know that they are part of my life and there's not anything that can change this. They are the best friends I could ever have.

I have also to add that Francesca and Antonio are right. I didn't think it would happen like this. She will always be my role model and that year has left its mark. And to think that Antonio has also forecast this absurd 2017. I'm really starting to wonder if he is a veterinarian.

Daniele is not on the list of people to thank because he is a part of my heart. I have promised him that during this period I would be passionate and dedicated to a belief as he is. So far, I have failed in that task. It is impossible to be like him. He is always the best and I am proud to be his sister.

For my parents, there is nothing I can write that can express my heartful

thinking for their love, support, encouragement and understanding all the time. Without them, I would not have the courage to conquer all the difficulties. If today I have the possibility to chose what I want to do and if I can write this thesis it is for them and for everything that they have done for me Can I only write: "I love you. . . "

Although I never know what will happen I am really lucky for what I have: for the people that surround me and for what I am doing and as Tiziano says "lonely is only a word" I will pull my weight, just like I always have all my life.

Ancona, November 2017

Marina Paolanti

Sommario

La Pattern Recognition è lo studio di come le macchine osservano l'ambiente, imparano a distinguere i pattern di interesse dal loro background e prendono decisioni valide e ragionevoli sulle categorie di modelli. Oggi l'applicazione degli algoritmi e delle tecniche di Pattern Recognition è trasversale. Con i recenti progressi nella computer vision, abbiamo la capacità di estrarre dati multimediali per ottenere informazioni preziose su ciò che sta accadendo nel mondo. La disponibilità di sensori economici e ad alta risoluzione (ad esempio, camere RGB-D) e la condivisione dei dati hanno portato a enormi archivi di documenti digitalizzati (testo, voce, immagini e video).

Partendo da questa premessa, questa tesi affronta il tema dello sviluppo di sistemi di Pattern Recognition per applicazioni reali come la biologia, il retail, la sorveglianza, social media intelligence e i beni culturali. L'obiettivo principale é sviluppare applicazioni di computer vision in cui la Pattern Recognition è il nucleo centrale della loro progettazione, a partire dai metodi generali, che possono essere sfruttati in più campi di ricerca, per poi passare a metodi e tecniche che affrontano problemi specifici. L'attenzione é rivolta alle applicazioni innovative di Pattern Recognition per problemi del mondo reale e della ricerca interdisciplinare, agli studi sperimentali e/o teorici che forniscono nuove intuizioni che fanno avanzare i metodi di Pattern Recognition. L'ambizione finale è quella di stimolare nuove linee di ricerca, soprattutto all'interno di scenari interdisciplinari.

Di fronte a molti tipi di dati, come immagini, dati biologici e traiettorie, una difficoltà fondamentale è trovare rappresentazioni vettoriali rilevanti. Anche se questo problema era stato spesso affrontato dagli esperti del settore, è stato utile apprendere queste rappresentazioni direttamente dai dati, e gli algoritmi di Machine Learning, i metodi statistici e le tecniche di Deep Learning sono stati particolarmente efficaci. Le rappresentazioni si basano quindi su composizioni di semplici unità di elaborazione parametrizzate, la profondità che deriva dal gran numero di tali composizioni. E' auspicabile sviluppare nuove ed efficienti tecniche di rappresentazione dei dati o di apprendimento delle caratteristiche/indicizzazione, in grado di ottenere prestazioni promettenti nelle attività correlate.

L'obiettivo generale di questo lavoro consiste nel presentare una pipeline per selezionare il modello che meglio spiega le osservazioni date e adattarlo al prob-

lema presentato. Per la progettazione del sistema di riconoscimento dei modelli vengono eseguiti i seguenti passaggi: raccolta dati, estrazione delle caratteristiche, approccio di apprendimento personalizzato e analisi e valutazione comparativa.

Le applicazioni proposte aprono un'ampia gamma di opportunità nuove e importanti per la comunità della machine vision. Il nuovo set di dati raccolti e le aree complesse prese in esame rendono la ricerca impegnativa. In effetti, è fondamentale valutare le prestazioni dei metodi più avanzati per dimostrare la loro forza e debolezza e contribuire a identificare le ricerche future per progettare algoritmi più robusti. Per una valutazione completa delle prestazioni, è di grande importanza collezionare un dataset specifico perché i metodi di progettazione che sono adattati a un problema non funzionano correttamente su altri tipi di problemi. Inoltre, la selezione del dataset è necessaria da diversi domini applicativi per offrire all'utente la possibilità di dimostrare l'ampia validità dei metodi.

Un'attenzione particolare è stata rivolta all'esplorazione di modelli e algoritmi di apprendimento su misura e alla loro estensione a più aree di applicazione, oltre che alla ingegnerizzazione di features utili a lavorare con i modelli proposti. I metodi su misura, adottati per lo sviluppo delle applicazioni proposte, hanno dimostrato di essere in grado di estrarre caratteristiche statistiche complesse e di imparare in modo efficiente le loro rappresentazioni, permettendogli di generalizzare bene attraverso una vasta gamma di compiti di visione computerizzata, tra cui la classificazione delle immagini, il riconoscimento del testo e così via.

Abstract

Pattern Recognition is the study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decisions about the patterns categories. Nowadays, the application of Pattern Recognition algorithms and techniques is ubiquitous and transversal. With the recent advances in computer vision, we now have the ability to mine such massive visual data to obtain valuable insight about what is happening in the world. The availability of affordable and high resolution sensors (e.g., RGB-D cameras, microphones and scanners) and data sharing have resulted in huge repositories of digitized documents (text, speech, image and video).

Starting from such a premise, this thesis addresses the topic of developing next generation Pattern Recognition systems for real applications such as Biology, Retail, Surveillance, Social Media Intelligence and Digital Cultural Heritage. The main goal is to develop computer vision applications in which Pattern Recognition is the key core in their design, starting from general methods, that can be exploited in more fields, and then passing to methods and techniques addressing specific problems. The privileged focus is on up-to-date applications of Pattern Recognition techniques to real-world problems, and on interdisciplinary research, experimental and/or theoretical studies yielding new insights that advance Pattern Recognition methods. The final ambition is to spur new research lines, especially within interdisciplinary research scenarios.

Faced with many types of data, such as images, biological data and trajectories, a key difficulty was to find relevant vectorial representations. While this problem had been often handled in an ad-hoc way by domain experts, it has proved useful to learn these representations directly from data, and Machine Learning algorithms, statistical methods and Deep Learning techniques have been particularly successful. The representations are then based on compositions of simple parameterized processing units, the depth coming from the large number of such compositions. It was desirable to develop new, efficient data representation or feature learning/indexing techniques, which can achieve promising performance in the related tasks.

The overarching goal of this work consists of presenting a pipeline to select the model that best explains the given observations; nevertheless, it does not prioritize in memory and time complexity when matching models to obser-

vations. For the Pattern Recognition system design, the following steps are performed: data collection, features extraction, tailored learning approach and comparative analysis and assessment.

The proposed applications open up a wealth of novel and important opportunities for the machine vision community. The newly dataset collected as well as the complex areas taken into exam, make the research challenging. In fact, it is crucial to evaluate the performance of state of the art methods to demonstrate their strength and weakness and help identify future research for designing more robust algorithms. For comprehensive performance evaluation, it is of great importance developing a library and benchmark to gauge the state of the art because the methods design that are tuned to a specific problem do not work properly on other problems. Furthermore, the dataset selection is needed from different application domains in order to offer the user the opportunity to prove the broad validity of methods.

Intensive attention has been drawn to the exploration of tailored learning models and algorithms, and their extension to more application areas. The tailored methods, adopted for the development of the proposed applications, have shown to be capable of extracting complex statistical features and efficiently learning their representations, allowing it to generalize well across a wide variety of computer vision tasks, including image classification, text recognition and so on.

Acronyms

ABC Artificial bee colony

ACL Active Learning

AR Augmented Reality

ANN Artificial Neural Network

ART Assisted Reproductive Technology

BAN Biomarker Association Networks

BRD Biological Reference Dataset

CAD Computer-Aided Detection

CBAL class Based Active Learning

CBIR Content based image retrieval

CH Cultural Heritage

CHMM Coupled Hidden Markov Model

CNN Convolutional Neural Networks

CRF Conditional Random Field

DAE Deep Auto-Encoder

DBN Dynamic Bayesian network

DBM Deep Boltzmann Machine

DBT Digital Breast Tomosynthesis

DCNN Deep Convolutional Neural Network

DNN Deep Neural Network

DLA Differential Language Analysis

DSS Decision Support System

DT Decision Tree

EM Expectation-Maximization

FCM Fuzzy C-Means

FleBoRE Flexible Body Rigid Extremities

GC Granulosa Cell

GMM Gaussian mixture model

HMM Hidden Markov Model

ICL Imbalanced Class Learning

IVF In Vitro Fertilization

KBA Kernel Boundary Alignment

LDA Latent Dirichlet Allocation

PR Pattern Recognition

DCH Digital Cultural Heritage

LSTM Long-Short Term Memory

MaxEnt Maximum Entropy

MHMM Multiresolution Hidden Markov Model

MIP Most Informative Positive

MLP Multi-Layer Perceptron

PPI protein–protein interaction

RBMNs Recursive Bayesian Multinets

RF Random Forest

RFM Recency, frequency and monetary

RNN Recurrent Neural Network

SVM Support Vector Machine

SMI Social Media Intelligence

VB Variational Bayes

Contents

List of Figures

List of Tables

Chapter 1

Introduction.

By the time, they are five years old, most children can easily recognise digits and letters: small and large characters, handwritten, machine printed, or rotated. These may be written on a cluttered background, on crumpled paper or may even be partially occluded. This ability is taken for granted until we face the task of teaching a machine how to do the same.

Pattern recognition (PR) is the study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decisions about the patterns categories. The problem of searching for patterns in data is a fundamental one and has a long and successful history. For instance, the extensive astronomical observations of Tycho Brahe in the 16^{th} century allowed Johannes Kepler to discover the empirical laws of planetary motion, which in turn provided a springboard for the development of classical mechanics. Similarly, the discovery of regularities in atomic spectra played a key role in the development and verification of quantum physics in the early twentieth century. The field of PR is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories. Although almost 50 years of research, the design of a general purpose machine pattern recogniser remains an elusive goal. In most cases, the best pattern recognisers are humans, even if we do not understand how humans recognise patterns.

In [2], Ross highlights the work of Nobel Laureate Herbert Simon whose central finding was that PR is fundamental in most human decision making tasks: the more relevant patterns at your disposal, the better your decisions will be. This is hopeful news to proponents of artificial intelligence, since computers can surely be taught to recognize patterns.

Nowadays, the application of PR algorithms and techniques is ubiquitous and transversal. With the recent advances in computer vision, we have the ability to mine such massive visual data to obtain valuable insight about what is happening in the world. The availability of affordable and high resolution sensors (e.g., RGB-D cameras, microphones and scanners) and data sharing

have resulted in huge repositories of digitized multimedia documents (text, speech, image and video). Need for efficient archiving and retrieval of this data has fostered the development of algorithms in new application domains. Fields that take advantage of PR range from biology, among the first ones to be investigated, to Social Media Intelligence, among the new-comes, just to mention a few of them. Strategies developed to address a kind of problem are often transposed to solve a different one too. However, this interchange is not always obvious.

This work has the aim of developing next generation PR systems for real applications such as Biology, Retail, Surveillance, Social Media Intelligence and Digital Cultural Heritage. Its main scope is to bring together PR methods for computer vision applications in order to give a landscape of techniques that can be successfully applied and also to show how such techniques should be adapted to each particular domain. The advances in learning and recognizing patterns are allowing a point of view in the definition and development of more efficient and effective frameworks. In the light of this, it is important to apply and to exploit the most recent advancements in learning and recognition algorithms in different applications. In fact, the main goal is to develop computer vision applications in which PR is the key core in their design, starting from general methods, that can be exploited in more fields, and then passing to methods and techniques addressing specific problems. The privileged focus is on up-to-date applications of PR techniques to real-world problems, and on interdisciplinary research, experimental and/or theoretical studies yielding new insights that advance PR methods. The final ambition is to spur new research lines, especially within interdisciplinary research scenarios.

To make PR techniques tailored for the aforementioned challenging applications, considerations such as computational complexity reduction, hardware implementation, software optimization, and strategies for parallelizing solutions must be observed. The overarching goal of this work consists of presenting a pipeline to select the model that best explains the given observations; nevertheless, it does not prioritize in memory and time complexity when matching models to observations. For the PR system design the following steps are performed:

1. *Data collection*: training and testing data are collected. It is important to acknowledge an adequately large and representative set of samples. In fact, it is crucial to evaluate the performance of state of the art methods to demonstrate their strength and weakness and help identify future research for designing more robust algorithms. For comprehensive performance evaluation, it is critical to collect representative datasets. While much progress has been made in recent years with efforts on sharing code and datasets, it is of great importance to develop a library and bench-

mark to gauge the state of the art. The design of benchmark involves several issues. First of all, it needs to be done objectively in order not to give any method unfair advantages. It is important to take into account the specificity of the concrete situation. In fact, the methods design that are tuned to a specific problem do not work properly on other problems. The choice of problems can also be different for real applications. In case of real applications the problems are often cluttered with so many details so that it is often quite tedious, but it is much more valuable for the users. So a delicate balance has to be struck in the design of benchmarks and the choice of datasets. Furthermore, the dataset selection is needed from different application domains in order to offer the user the opportunity to prove the broad validity of methods. Despite the effort of some authors to augment their evaluation with additional datasets, a standardized and widely adopted evaluation methodologies for the aforementioned applications do not yet exists. To this end, five newly challenging datasets are specifically designed for the described task. In fact, each described application has involved the construction of one dataset, used as input. Thus, the learning methods described in this thesis are evaluated on the proposed datasets: Biological Reference Dataset (BRD), sCREEN, Top View Person Re-identification (TVPR), GfK Verein dataset and Eye-Tracking dataset.

2. *Features Extraction*: is an essential pre-processing step to PR problems. The feature extraction process usually consists of three phases: feature construction (features are constructed from linear or non-linear combination of raw features), feature selection (done using techniques like relevancy ranking of individual features) and feature reduction (used to reduce the features number especially when too many features are selected compared to the number of feature vectors).

3. *Tailored Learning Approach*: models are adapted following these criteria: domain dependence and prior information, definition of design criteria, parametric vs. non-parametric models, handling of missing features, computational complexity. Types of models are: supervised, semi-supervised, unsupervised, decision-theoretic or statistical, neural, and hybrid. Using these models, it can be investigated how close is to the true underlying the patterns.

4. *Comparative Analysis and Assessment*: various details have to be investigated such as domain dependence and prior information, computational cost and feasibility, discriminative features, similar values for similar patterns, different values for different patterns, invariant features with respect to translation, rotation and scale, robust features with respect to

Figure 1.1: Pipeline for the development of the challenging computer vision applications

occlusion, distortion, deformation, and variations in environment. Then, performance with training sample are estimated, performance with future data are predicted and problems of overfitting and generalization are evaluated.

In Figure 1.1 these four steps are schematically represented.

The introduction Chapter of the present work is aimed at providing a wide overview about the concept of *PR*, starting from the importance of data, regardless the type of data dealing with. The starting point is that each domain has its own kind of data, like compartments that cannot mesh up and support each other and the original research contributions of the thesis address the broad challenges faced in PR. The multidisciplinary nature of the presented work arose from the above mentioned assumptions.

With the second Chapter, a review the state of art about methods and applications is provided that, more than others, have demonstrated of being suitable for the development of new kinds of interaction and to make the proposed PR applications challenging.

The use cases faced during the research activity are collected in Chapter 3. It represents the summa of different experiences, activities and best practices which can help to contribute, compared with the state of art, to better solving various real-world problems of the computer vision applications in different research areas. Each use case have been divided, to facilitate the reader, into a hierarchical scheme: *Scenario*, *Challenging application description*, *Dataset Description* and *Performance Evaluation and Results*. In this way, the benefits that the different approaches (in different domains), brought to the users can be better highlighted. Chapter 4, besides arguing over the possibilities that the proposed paradigm opens up in different topics, summarizes also the challenges, the open issues and the limitations that require further investigations. From this, the future works are outlined and described in the Chapter 5, together with concluding remarks.

Chapter 2

State of the art and Perspectives of Pattern Recognition Applications.

In recent years, there has been a growing demand for computer vision applications due to the significant improvement in the processing technology, network subsystems and availability of large storage systems. The emerging applications are not only challenging but also computationally more demanding. This visual data demand has led a significant interest in the research community to develop methods to archive, query and retrieve this data based on their content. Automatic (machine) recognition, description, classification, and grouping of patterns are important problems in a variety of engineering and scientific disciplines such as biology, medicine, marketing, computer vision and artificial intelligence. Such systems employ PR methods developed over the years.

In this context, the term PR methods refer to their applicability in feature extraction, feature clustering, generation of database indices, and determining similarity in content of the query and database elements. The PR system design essentially involves the following aspects: definition of pattern classes, sensing environment, pattern representation, feature extraction and selection, cluster analysis, classifier design and learning, selection of training and test samples, and performance evaluation. The problem domain dictates the choice of sensors, pre-processing techniques, representation scheme, and decision making model [3].

In many rising applications, it is clear that no single approach for classification is "optimal" and that multiple methods and approaches have to be used. Consequently, combining several sensing techniques and classifiers is now a commonly used practice in PR.

It is generally agreed that a well-defined and sufficiently constrained recognition problem (small intraclass variations and large interclass variations) will lead to a compact pattern representation and a simple decision making strategy. Learning from a set of examples (training set) is an important and desired attribute of most PR systems. At present, the best known PR approaches for developing a vision system are:

- Machine Learning;
- Statistical Methods;
- Deep Learning.

These models are not necessarily independent and sometimes the same PR method exists with different interpretations.

Although the multidisciplinary nature of this work brought to uncover the potential of PR applications among different domains, as stated in the introduction section, the major efforts were undertaken towards five challenging domains: biology, retail, surveillance, social media intelligent and Digital Cultural Heritage (DCH). A systematic literature review has been performed in order to understand the research issues related to the use of PR for computer vision applications, and more, to understand if and how PR methods and techniques could help the creation of applications in these fields. In the following, besides a brief overview about the PR methods and techniques, there will be a specific focus on the state of the art in the five chosen domain. In particular, for each research field are analysed the methods and techniques, also main paths that most approaches follow are summarized and their contributions are pointed out. The reviewed approaches are categorized and compared from multiple perspectives, including methodology, function and analyse the pros and cons of each category.

2.1 Background

PR is concerned with the design and development of systems that recognize patterns in data. The purpose of a PR program is to analyse and describe a scene in real world which is useful for the accomplishment of some task. The real world observations are gathered through sensors and a PR system classifies these observations.

Over the years, several definitions of PR has been given. Watanabe [4] defines a pattern "As opposite of a chaos; it is an entity, vaguely defined, that could be given a name". Duda and Hart [5] described PR as a field concerned with machine recognition of meaningful regularities in noisy or complex environments. For Jain et al. [3], PR is a general term to describe a wide range of problems such as recognition, description, classification, and grouping of patterns. Pavlidis in his book affirmed that "the word pattern is derived from the same root as the word patron and, in his original use, means something which is set up as a perfect example to be imitated. Thus PR means the identification of the ideal which a given object was made after" [6]. In [7], PR is a classification of input data via extraction important features from a lot of

noisy data. "A problem of estimating density functions in a high-dimensional space and dividing the space into the regions of categories of classes" is the PR for Fukunaga [8]. Schalkoff defined PR as "The science that concerns the description or classification (recognition) of measurements" [9]. PR is referred to the prediction of the unknown nature of an observation, a discrete quantity such as black or white, one or zero, sick or healthy, real or fake. There are several ways in which an algorithm can model a problem based on its interactions with the experience or environment or input data. The primary goal of PR is supervised or unsupervised classification and *Machine learning* has become one of the main-stays of information technology.

In *supervised learning*, a target function is learned and it is used to predict the values of a discrete class attribute as approved or not-approved, for instance when there are label examples of two or more classes (e.g. disease vs healthy). Machine learning algorithms make predictions on a given set of sample whereas supervised learning algorithms search for patterns within the value labels assigned to data points. These algorithms consist of an outcome variable which is to be predicted from, a given set of predictors i.e. independent variables. By using these variables set, it is possible to generate a function that map input to desired outputs. The training process continues until the model achieves a level of accuracy on the training data. Examples of supervised learning algorithms are: Support Vector Machine (SVM) [10, 11, 12], Random Forest (RF) [13], Decision Tree [14], Neural Networks [15], kNearest Neighbors (kNN) [16], Naïve Bayes (NB) [17] and Artificial Neural Network (ANN) [18]. The supervised learning predominantly is divided into two phases: *Training* (learning a model using the training data) and *Testing* (testing the model using unrevealed test data to appraise the model accuracy). The algorithms learn from their past experience and try to capture the best possible knowledge to make accurate business decisions.

Unsupervised learning (clustering) is a difficult problem for many reasons such as effective similarity measures, criterion functions, algorithms and initial conditions [19]. The unsupervised algorithms are used when the samples are not labeled. Thus, it is an approach of learning where instances are automatically placed into meaningful groups based on their similarity and the classifier is designed by deducing existing patterns or cluster in the training datasets. Generally, clustering algorithms can be categorized into partitioning methods, hierarchical methods, density-based methods, grid-based methods, and model-based methods [20]. The partitioning methods are divided into two major subcategories, the centroid and the medoids algorithms. The centroid algorithms represent each cluster by using the gravity centre of the instances. The medoid algorithms represent each cluster by means of the instances closest to the gravity centre. The most known centroid algorithm is the k-means [20].

The k-means algorithm partitions the data set into k subsets such that all points in a given subset are closest to the same centre. Traditional clustering approaches generate partitions and in a partition, each pattern belongs to one and only one cluster. Fuzzy clustering extends this notion to associate each pattern with every cluster using a membership function [21]. Larger membership values indicate higher confidence in the assignment of the pattern to the cluster. One widely used algorithm is the Fuzzy C-Means (FCM) algorithm [22], which is based on k-means. FCM finds the most characteristic point in each cluster, which can be considered as the "center" of the cluster and, then, the grade of membership for each instance in the clusters. Other soft clustering algorithms have been developed and most of them are based on the Expectation-Maximization (EM) algorithm. They assume an underlying probability model with parameters that describe the probability that an instance belongs to a certain cluster [23]. The hierarchical methods group data instances into a tree of clusters. There are two major methods under this category. One is the agglomerative method, which forms the clusters in a bottom-up fashion until all data instances belong to the same cluster. The other is the divisive method, which splits up the data set into smaller cluster in a top-down fashion until each cluster contains only one instance [24, 25, 26, 27]. Both divisive algorithms and agglomerative algorithms can be represented by dendrograms and are known for their quick termination [19]. Some of the hierarchical clustering algorithms are: Balanced Iterative Reducing and Clustering using Hierarchies – BIRCH [28], Clustering Using REpresentatives – CURE [29] and CHAMELEON [30]. Density-based clustering algorithms try to find clusters based on density of data points in a region. The key idea of density-based clustering is that for each instance of a cluster the neighborhood of a given radius (Eps) has to contain at least a minimum number of instances (MinPts). The common density-based clustering algorithms is the DBSCAN [31]. Grid-based clustering algorithms first quantize the clustering space into a finite number of cells (hyper-rectangles) and then perform the required operations on the quantized space. Cells that contain more than a certain number of points are treated as dense and the dense cells are connected to form the clusters. Well-known grid-based clustering algorithms are: STatistical INformation Grid-based method– STING [32], WaveCluster [33], and CLustering In QUEst–CLIQUE [34]. AutoClass is based on the Bayesian approach. Starting from a random initialization of the parameters, it incrementally adjusts them in an attempt to find their maximum likelihood estimates [35]. Furthermore, in [36], it is assumed that there is a hidden variable in addition to the observed or predictive attributes. This unobserved variable reflects the cluster membership for every case in the dataset. The data-clustering problem is even an example of supervised learning from incomplete data due to the

existence of such a hidden variable [37]. Their approach for learning is called Recursive Bayesian Multinets (RBMNs). A model based method is the SOM net, which can be thought of as two layers neural network [38]. Combining multiple clustering algorithms is a challenging problem than combining multiple classifiers. Cluster ensembles can be formed in different ways, such as the use of a number of different clustering techniques (either deliberately or arbitrarily selected), the use of a single technique many times with different initial conditions and the use of different partial subsets of features or patterns [39].

Other machine learning approaches are: Semi-supervised learning, Reinforcement learning, Transduction and Learning to learn [40].

In *Semi-supervised learning*, the training dataset contains both labeled and unlabelled data. The classifier is train to learn the patterns to classify, predict and label the data. The algorithm is trained to map action to situation so that the reward or feedback signal is maximised in the *Reinforcement learning*. The classifier is not programmed directly to choose the action, but instead trained to find the most rewarding actions by trial and error. The *Transduction* attempts to predict the output based on training data, training label, and test data. It has common point with supervised learning, but it does not develop an explicit classifier. Finally in the *Learning to learn* approach, the classifier is trained to learn from the bias that is induced during previous stages.

It is important to specify that feature selection methods, such as Principal Component Analysis (PCA) [41], Linear Discriminant Analysis [42], and wrapper methods [43], seek to reduce the dimensionality of data sets, identify informative features, and remove irrelevant features, to avoid overfitting and underfitting the learned model [44].

A subfield of Machine Learning is the Active Learning (ACL) sometimes called "query learning" or "optimal experimental design" in the statistics literature). It allows choosing the data from which it learns to be "curious" if it will perform better with fewer training [45]. In past years, a combination of Active Learning (ACL) and Imbalanced Class Learning (ICL) was used to address past problems to develop a more efficient feature selection process and address the imbalance problem in datasets. To address this issue, oversampling or undersampling techniques, are used. However, these approaches have their own drawbacks. Oversampling the minority class leads to overfitting, whereas undersampling the majority class leads to underfitting [46].

Statistical methods and estimation theories have been commonly used in PR for a long time. They are classical approaches of PR which was found out during a long developing process and are based on the feature vector distributing which getting from probability and statistical model. In order to build robust machine learning algorithms, is necessary that generative models are capable of capturing various aspects of the data at the same time. These models should be

simple, but capable of adapting to the data. In machine learning community, these models are defined flexible models and are minimally structured probability models with a large number of parameters that can adapt so as to explain the input data. The famous generative models are Dynamic Bayesian network (DBN) which model sequential data. The Hidden Markov models (HMMs) are the most known DBN is the discrete-time, i.e. Markov model in which the states are not directly observable (Figure 2.1). Instead, each state is characterized by a probability distribution function. The HMMs are composed by a sequence of hidden variables $s = \{s_k\}_{k=1}^{k}$ and by a sequence of visible variables $o = \{o_k\}_{k=1}^{k}$. It is also composed of other elements such as:

1. Q, $|Q| = L$ the finite set of (hidden) states;

2. A transition matrix $A = a^{mn}, 1 \leq m, n \leq L$ representing the probability of moving from state m to state n,

$$a^{mn} = P(s_{k+1} = n | s_k = m), 1 \leq n, m \leq L, \tag{2.1}$$

   with $a^{mn} \geq 0$, $\sum_{n=1}^{L} a^{mn} = 1$, and where s_k denotes the state occupied by the model at index k. The index k depends on the context and it indicates a time index if the considered sequence is thought as generated by a temporal stochastic process, or it is considered a site index if the sequence is atemporal, with its spatial structure regulated by a Markovian process.

3. An emission matrix $B = \{b^m(v)\}$, that indicates the probability of emission of symbol $v \in V$ when the system state is m. The emission matrix of the hidden Markov model is usually discrete or gaussian, in the first case B is a multinomial distribution, in the second case is normal-distributed.

4. $\pi = \{\pi^m\}$, the initial state probability distribution,

$$\pi^m = P(s_1 = m), 1 \leq m \leq L \tag{2.2}$$

   with $\pi^m \geq 0$ and $\sum_{n=1}^{L} \pi^m = 1$.

Given an HMMλ, the probability of a particular sequence of visible symbols o under this model, i.e., $P(o|\lambda)$ has to be determinate. This problem can be solved efficiently using the forward-backward procedure [110]. This procedure calculates the forward variables $\alpha_k(m)$, $k = 1....N$. These variables represent the probability of the partial observation sequence $o_1 o_2 .... o_k$ and the state s_k, given the model λ, formally

$$\alpha_k(m) = P(o_1 o_2 .... o_k, s_k = m | \lambda). \tag{2.3}$$

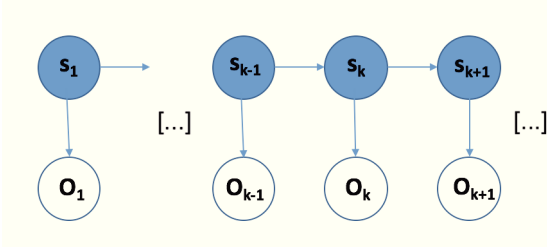


Figure 2.1: An HMM $\lambda =< A, B, \pi >$.

It has also to be determined the most likely sequence of hidden states values $s = \{s_k\}_{k=1}^{N}$ that led to a particular observation o. This problem can be solved efficiently using the Viterbi algorithm [110].

However, the Markovian framework makes strong restrictive assumptions about the system generating the signal that, it is a single process having a small number of states and an extremely limited state memory. The single-process model is often inappropriate for vision and speech applications, resulting in low ceilings on model performance. Coupled Hidden Markov Models (CHMM) provide an efficient way to resolve many of these problems, and offer superior training speeds, model likelihoods, and robustness to initial conditions [47].

There are other statistical approaches which could be used in PR like Latent Dirichlet Allocation (LDA). It was used by Duric and Song [48] to separate the entities in a review document from the subjective expressions that describe those entities in terms of polarities. LDA are generative models that allow documents to be explained by unobserved (latent) topics.

To advance traditional machine learning and artificial intelligence approaches, *Deep Learning* techniques have recently been exploited in a series of multimedia tasks, such as multimedia content analysis and understanding, retrieval, compression, and transmission, natural language processing (NLP), information retrieval, and image analysis[49, 50, 51]. Deep Neural Networks (DNNs) have become a crucial technology in the field of multimedia community. For example, the neural networks Deep Boltzmann Machine (DBM) and Deep Auto-Encoder (DAE) have been widely used for multimodal learning and cross-modal retrieval. The Convolutional Neural Networks (CNN) and their variants have become the basic tools for building deep representations to perceive and understand multimodal information, such as images and audios. Recurrent Neural Networks (RNN) or Long-Short Term Memory (LSTM) can be used for sequence modeling and prediction for high-level semantic data like natural language [52]. The significantly lowered cost of computing hardware, the increased chip processing abilities (e.g., GPU units), and recent advances in machine learning and signal/information processing research are three important rea-

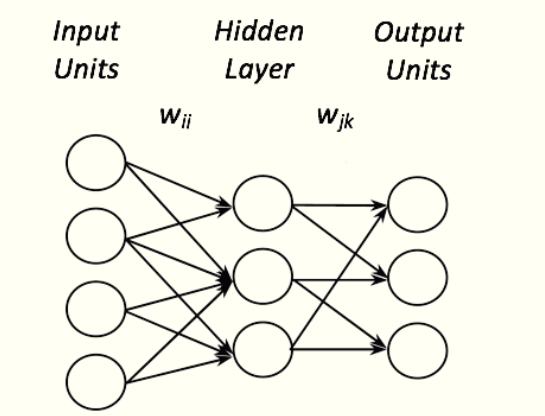


Figure 2.2: Feed-forward ANN

sons for the popularity of deep learning today. The concept of deep learning originated from the study on artificial neural networks (ANNs) [53]. In fact, in the past decades, ANNs have become an active research area [54, 55, 56, 57, 58].

Zhang provided an overview of existing work in Artificial Neural Networks (ANNs) [59]. A multi-layer neural network consists of neurons, which are large number of units, joined together in a pattern of connections. Units in a net are usually segregated into three classes: *input units*(receive information to be processed), *output units* (where the results of the processing are found) and *units* (in between known as hidden units). Feed-forward ANNs allow signals to travel one way only, from input to output (Figure 2.2).

ANN depends upon three fundamental aspects, input and activation functions of the unit, network architecture and the weight of each input connection. There are several algorithms with which a network can be trained [60]. The shallow architectures commonly used Gaussian mixture models (GMMs) and hidden Markov models (HMMs), linear or non-linear dynamical systems, conditional random fields (CRFs), maximum entropy (MaxEnt) models, SVMs, logistic regression, kernel regression, and multi-layer perceptron (MLP) neural network with a single hidden layer including extreme learning machine. A common property to the mentioned learning models is the relatively simple architecture that consists of only one layer responsible for transforming the raw input signals or features into a problem-specific feature space, which may be unobservable [61]. Feed-forward neural networks or MLP (with many hidden layers) are indeed a good example of the models with a deep architecture. A well-known algorithm for learning the networks weights is the backpropagation. However, only backpropagation did not work well in practice for learning networks with more than a small number of hidden layers [62, 63].

The deep architectures are classified into: generative, discriminative, and

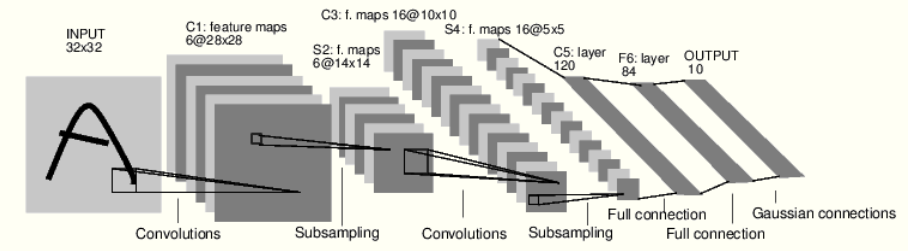


Figure 2.3: The LeNet architecture consists of two sets of convolutional, activation, and pooling layers, followed by a fully-connected layer, activation, another fully-connected, and finally a softmax classifier (image from [1]).

hybrid categories [61]. The generative deep architectures characterize the high-order correlation properties of the observed or visible data for pattern analysis or synthesis purposes, and/or the joint statistical distributions of the visible data and their associated classes. Instead, the discriminative deep architectures are intended to provide discriminative power for pattern classification, by characterizing the posterior distributions of classes conditioned on the visible data. Finally, the Hybrid deep architectures, have the aim to discriminate but are assisted with the outcomes of generative architectures via better optimization or/and regularization, or when discriminative criteria are used to learn the parameters in any of the deep generative models.

A representative deep architecture is the RNN [64]. RNNs are not widely used because they are extremely difficult to train properly due to the "vanishing gradient" problem. More recently, Bengio et al. [65] and Sutskever [66] have explored optimization methods in training generative RNNs that modify stochastic gradient descent and show how these modifications can outperform Hessian-free optimization methods [67].

An example of discriminative deep architecture is CNN, with each module consisting of a convolutional layer and a pooling layer (Figure 2.3). The convolutional layer shares many weights, and the pooling layer subsamples the output of the convolutional layer and reduces the data rate from the layer below. The CNNs are highly effective in computer vision and image recognition [1, 68, 69, 70, 71]. Recently, CNNs are also found suitable for speech recognition, obviously with appropriate changes compared with the [72, 73, 74, 75, 76].

For the third category, the term "hybrid" is referred to the deep architecture that comprises or makes use of both generative and discriminative model components. Generally, the generative component is exploited to help with discrimination which is the final goal of the hybrid architecture [53, 77, 78, 79].

2.2 Biology

The history of relations between biology and PR is long and complex. The biological research has seen an explosive expansion of big data [80, 81, 82] that provides enormous opportunities for translational research such as precision medicine [83] if the new challenges of data mining and knowledge discovery can be addressed. Machine learning is a powerful tool for interpreting biological data. Approaches such as k-Nearest-Neighbor [16, 84] and random forest [85] algorithms are widely used in biological data preprocessing. Combining local knowledge discovery in global summaries and individual datasets of the results crossing datasets [86]. However, the extraction of biological data are time consuming and expensive due to the challenges of implementing experimental procedures that can produce unexpected phenomena and several computational challenges to extract and analyse this data. Biological datasets have high dimensionality, but the cases of interest (e.g., disease states) are relatively rare. The following sections provide an overview of the PR recent works in biology. In particular in SubSec 2.2.1 summarizes the PR methods applied to biological problem, instead SubSec 2.2.2 describes the recent applications in which these techniques are applied.

2.2.1 Algorithms and Approaches

Table 2.1 summarizes the PR techniques and specific approaches applied to biological datasets. It also describes the main advantages and disadvantages of each approach, as well as cites some recent examples from the literature. The supervised algorithms have been applied for the prediction of gene ontology and gene expression profiles across different environmental and experimental conditions [46]. Among the unsupervised algorithms, K-means clustering and hierarchical clustering have been widely used in biological datasets. For example, chromatic data has been used with unsupervised learning algorithms for annotating the genomes to identify novel groups of functional elements. In biology, the semi-supervised algorithms fall between supervised and unsupervised, especially for cases when only a small portion of the samples is labeled. These algorithms have been used to identify functional relationships between genes and transcription factor binding sites. They are widely applied for gene-finding approaches where the entire genome is the unlabeled set and only a collection of genes is annotated. Tentative labels are given after a first pass and the algorithm iterates to improve the learning model. Since retrieving good biological data can take months to years, sometimes when experiments are done, researchers seek specific cases having a low incidence rate. This entails that many biological data are naturally imbalanced. For example, among the 27,000 mouse genes, an experiment may observe only about 100 whose DNA methy-

lation was changed within the experimental settings [87]. Therefore, collecting data on such changes is a time consuming, multi-step process, and, naturally, results in a class imbalance problem. It is important to choose the most informative instances and features for building a good classifier without running more extensive experiments to obtain more rare instances. Therefore, both ACL and ICL methods have applications in biology. In [88], the authors used the Most Informative Positive (MIP) ACL method to find p53 mutants (mutated p53 is responsible for half of human cancers). In their ACL approach, they train the classifier by using positive instances that pass a given score (which ranks all unlabeled instances) and include negative instances in the training set only if there are too few positive instances. In [89], it is proposed a different study uses ACL techniques to annotate digital histopathology data. This method, class Based Active Learning (CBAL), uses a mathematical model that calculates the cost of building a training set with a certain size and class ratio. In recent years, the application of Deep Learning techniques to biological data sets have increased substantially [90]. Many of these works have focused on biomedical imaging [91], but a significant number of studies have focused on genomic data [92]. These tasks include protein structure prediction, protein classification, and gene expression regulation. Such applications are characterized by the computation of hundreds to thousands of predetermined features, such as motifs, which are input to a deep learning network. Some recent works have used deep learning networks to generate relevant motifs using convolution layers on windowed sequence data, such as in the Deep-Bind method [93]. Other approaches have used a one-hot encoding input to a convolution layer, where each sequence window of size W is represented by a W×4 array indicating which bases (A, G, C, T) are present in the sequence window, such as in the DeepMotif method [94].

2.2.2 Applications

In [107], authors have generated and investigated multiple types of feature representations in order to improve the text classification task performance of protein–protein interaction (PPI). They have explored features such as protein–protein interaction trigger keywords, protein named entities and the advanced ways of incorporating Natural Language Processing (NLP) output, in addition to the traditional ones (domain-independent bag-of-words approach and the term weighting methods). The experimental results have shown that both the advanced way of using NLP output and the integration of bag-of-words and NLP output improved the performance of text classification.

An automated chromosome karyotyping scheme using a two-layer classification platform is developed in [108]. The authors have assumed that by selecting

Table 2.1: PR approaches for biological datasets, along with their function, advantages, disadvantages, and recent examples.

PR Approach	**Function**	**Advantages**	**Disadvantages**
Supervised Learning (e.g., SVM [95], RF [96])	Learn a model discriminating one class of biological phenomena from one or more other classes.	Precise model with predictive and interpretative properties.	Requires equally large number of examples from each class.
Unsupervised Learning (e.g., K-means [97], hierarchical clustering [98])	Learn a model descriptive of the biological phenomena in the data.	Does not require class labels on data.	Sensitive to similarity measure; results difficult to interpret.
Semi-supervised Learning (e.g., transduction [99])	Learn model from mixture of labeled and unlabeled data.	Utilize all available data; typically outperforms use just labeled data.	Sensitive to errors in propagating class labels from labeled to unlabeled data.
Feature Selection (e.g., PCA [100], LDA [101], wrapper [102])	Reduce large number of features to fewer, more informative features.	Improves efficiency and accuracy of learning.	Sensitive to feature evaluation metric; may discard informative features.
Active Learning (e.g., uncertainty sampling [103], most informative instance [104])	Identify most informative instances to label for accurate model learning.	Reduces number of examples needed to learn model; reduces burden on human expert and experiment cost.	May focus learner on outliers rather than prominent classes.
Imbalanced class Learning (e.g., minority over-sampling [105], boosting [106])	Learn in the presence of large skew in the number of examples of each class.	Learn with relatively few examples of biological phenomenon of interest.	May underfit or overfit data depending on bias toward minority class.
Deep Learning (DeepBind [87], DeepMotif [88])	Learns complex representations of concepts in the data.	General purpose and high accuracy.	Sensitive to parameter choices; long training times.

most effective feature sets and adaptively optimizing classifiers for the different groups of chromosomes with similar image characteristics, they can reduce the complexity of automated karyotyping scheme and improve its performance and robustness. For this purpose, they assembled an image database involving 6900 chromosomes and implemented a genetic algorithm to optimize the topology of multi-feature based artificial neural networks (ANN). In the first layer of the scheme, a single ANN was employed to classify 24 chromosomes into seven

classes. In the second layer, seven ANNs were adaptively optimized for seven classes to identify individual chromosomes. The scheme was optimized and evaluated using a "training–testing–validation" method.

In [109], the authors presented an original method for quantification and classification of erythrocytes in stained thin blood films infected with Plasmodium falciparum. Their approach was composed of three main phases: a preprocessing step, which corrects luminance differences. A segmentation step that uses the normalized RGB color space for classifying pixels either as erythrocyte or background followed by an Inclusion-Tree representation that structured the pixel information into objects, from which erythrocytes were found. Finally, a two step classification process identifies infected erythrocytes and differentiates the infection stage, using a trained bank of classifiers. Additionally, user intervention was allowed when the approach cannot make a proper decision. Four hundred fifty malaria images were used for training and evaluating the method.

An approach based on the concept of Biomarker Association Networks (BAN) for cancer classification is proposed in [110]. The BAN was modeled as a neural network, which can capture the associations between the biomarkers by minimizing an energy function. Based on the BAN, a cancer classification approach is developed. The derived BAN were observed to be significantly different for different cancer classes, which help reveal the underlying deviant biomarker association patterns responsible for different cancer types.

In [111], the authors have presented a computer-aided detection (CAD) system for mammographic masses that uses a mutual information-based template matching scheme with selected templates. They have presented principles of template matching with mutual information for mammography before. An implementation of those principles in a complete CAD system is proposed. The system, through an automatic optimization process, chose the most useful templates (mammographic regions of interest) using a large database of previously collected and annotated mammograms. Through this process, the knowledge about the task of detecting masses in mammograms is incorporated in the system. Then, they have evaluated whether the system developed for screen-film mammograms could be successfully applied not only to other mammograms but also to digital breast tomosynthesis (DBT) reconstructed slices without adding any DBT cases for training. Since mutual information is known to be a robust inter-modality image similarity measure, it had high potential of transferring knowledge between modalities in the context of the mass detection task.

Kostopoulos et al. proposed in [112], features that evaluate pictorial differences between melanocytic nevus (mole) and melanoma lesions by computer-based analysis of plain photography images and to design a cross-platform, tunable, decision support system (DSS) to discriminate with high accuracy

moles from melanomas in different publicly available image databases were proposed by the authors.

Llobet et al. in [113] have tested a method for the analysis of transrectal ultrasound images aimed at computer-aided diagnosis of prostate cancer. Two classifiers based on k-NN and HMMs were compared. The diagnostic capacity of the system was tested by means of a set of experiments where humans with varying degrees of experience classified a set of ultrasound images with and without the aid of the computer-aided system.

A content based image retrieval system (CBIR) for a database of parasite specimen images was proposed in [114]. Unlike most content based image retrieval systems, where the database consists of objects that vary widely in shape and size, the objects in the database were uniform. and were characterized by flexible body shapes, but with fairly rigid ends. They have defined such shapes to be FleBoRE (Flexible Body Rigid Extremities) objects, and have presented a shape model for this class of objects. The authors have also defined similarity functions to compute the degree of likeness between two FleBoRE objects and developed automated methods to extract them from specimen images. The system has been tested with a collection of parasite images from the Harold W. Manter Laboratory for Parasitology.

In [115], a DCNN approach and the general framework for recognition of objects in a real-time scenario and in an egocentric perspective are proposed. The window of interest was built on the basis of visual attention map computed over gaze fixations measured by a glass-worn eye-tracker. The application of this set-up was an interactive user-friendly environment for upper-limb amputees. Vision has to help the subject to control his worn neuro-prosthesis in case of a small amount of remaining muscles when the EMG control becomes inefficient. The recognition results on a specifically recorded corpus of 151 videos with simple geometrical objects have shown the mean Average Precision (mAP) of 64,6% and the computational time at the generalization lower than a time of a visual fixation on the object of interest.

Blaiotta et al. [116] have presented a variational Bayes (VB) approach for image segmentation. The aim was to show that VB provides a framework for generalising existing segmentation algorithms that rely on an expectation–maximisation formulation, while increasing their robustness and computational stability. They have also shown how optimal model complexity can be automatically determined in a variational setting, as opposed to machine learning frameworks which are intrinsically prone to overfitting. Finally, they have demonstrated how suitable intensity priors, that can be used in combination with this algorithm, can be learned from large imaging data sets by adopting an empirical Bayes approach [116].

2.3 Social Media Intelligence (SMI)

SMI represents the stack of technology solutions and methods used to monitor social media including social conversations and emerging trends. With the wide variety of social media channels available, there is a huge amount of data available. The challenge comes in accessing that data and transforming it into something that is usable and actionable. This intelligence is analysed and used to create meaningful content and make business decisions across many disciplines. SMI includes monitoring of content, such as messages or images posted, and other data, which are generated when someone uses a social media networking site. This information involves person-to-person, person-to-group, group-to-group, and includes interactions that are private and public. Methods may include manually reviewing content as it is posted in public or private groups or pages; reviewing the results of searches and queries of users; reviewing the activities or types of content users post; "scraping" (extracting the content of a web page) and replicating content in ways that are directly accessible to the person gathering social media intelligence. Audience sentiment is often the first place people look when implementing a social listening tool, because it's the quickest way to get a broad, high level overview of what the feelings are of a given subject within the community conversation. Text mining in social media data and visual sentiment classification in social media data as analytic approaches are by now widely used for revealing new insights in social media data. In texts from blogs, comments or posts, the sentiments and opinions of users on certain topics, products, or persons are often mined.

2.3.1 Algorithms and Approaches

Sentiment analysis aims at the detection of polarity and can be achieved in many different ways. Approaches for sentiment analysis can be differentiated with respect to the used methods and data sources. From a methodological perspective, we can distinguish between knowledge-based techniques and statistical methods [117]. Knowledge-based techniques, such as WordNet Affect [118] and SentiWordNet [119], rely on semantic knowledge resources to determine the sentiment. For example, in textual sentiment analysis, the sentiment of text is classified based on the presence of effective words from a lexicon. These methods are popular because of their easy application and accessibility, but their validity depends on a comprehensive knowledge base and rich knowledge representation. Statistical methods are trained with the aid of annotated corpora to identify the sentiment. These powerful methods are widely applied in research, but their performance depends on a sufficiently large training corpus [120]. Statistical approaches like HMM and LDA were used by Duric and Song [48] to separate the entities in a review document from the subjective

expressions that describe those entities in terms of polarities. SVMs were used by Li and Li [121] as a sentiment polarity classifier. Unlike the binary classification problem, they argued that opinion subjectivity and expresser credibility should also be taken into consideration. They have proposed a framework that provided a compact numeric summarization of opinions on micro-blogs platforms. The authors have identified and have extracted the topics mentioned in the opinions associated with the queries of users, and then classified the opinions using SVM. They have worked on tweets for their experiment. While in former times shallow feature representations such as bag-of-words combined with support vector machines have been the mainstream in textual sentiment analysis, deep learning methods are becoming increasingly popular in recent years. In [122], the authors use a Convolutional Neural Network (CNN) to extract sentence features and perform sentiment analysis of Twitter messages. An ensemble system to detect the sentiment of a text document from a dataset of IMDB movie reviews is built in [123]. CNNs have also been applied to visual sentiment analysis. A deep CNN model called DeepSentiBank is trained to classify visual sentiment concepts by Chen [124]. A visual sentiment prediction framework is introduced in [125]. It performs transfer learning from a pre-trained CNN with millions of parameters. With respect to the underlying data sources, sentiment analysis approaches can be divided into unimodal and multimodal [126]. While unimodal approaches consider only one data source, multimodal models take several types of data sources into account when determining the sentiment. In [127] the authors employ both images and text to predict sentiment by fine-tuning a CNN for image sentiment analysis and by training a paragraph vector model for textual sentiment analysis. In [128], the authors employ deep learning to analyze the sentiment of Chinese microblogs from both textual and visual content.

Table 2.2 summarizes the PR techniques and specific approaches for the analysis of Social Media contents.

2.3.2 Applications

In this section, some of the possible applications of PR in SMI are listed. There are a huge number of companies, large and small, that have opinion mining and sentiment analysis as part of their goals. However, these companies are not mentioned because the industrial landscape tends to change quite rapidly, so that lists risk falling out of date rather quickly.

In [131], the authors have focused on developing an incremental face recognition method for Twitter application. In particular, they have proposed a data-independent feature extraction method via binarization of a Gabor filter. Subsequently, the dimension of Gabor representation is reduced considering

Table 2.2: PR approaches for the analysis of Social Media contents with their function, advantages, disadvantages, and recent works.

PR Approach	Function	Advantages	Disadvantages
Supervised Learning (e.g., SVM and NB [129])	Depends on the existence of labeled training documents.	Better performance than the unsupervised methods.	Requires large amounts of labeled training data that are very expensive.
unsupervised Learning (e.g., HMM [48], LDA [130])	Allows documents to be explained by unobserved (latent) topics.	Easy to have unlabelled data.	Very sensitive to noise.
Knowledge-based techniques (e.g., WordNet Affect [118], SentiWordNet [119])	Rely on semantic knowledge resources to determine the sentiment.	Easy application and accessibility.	Their validity depends on a comprehensive knowledge base and rich knowledge representation.
Deep Learning (e.g., CNN [123])	Allow networks to have fewer weights and they are given a very effective tool, convolutions for image processing.	They use to need a lot of training data.	High computational cost.

various orientations at different grid positions. Finally, an incremental neural network is applied to learn the reduced Gabor features. The method is applied to an application which notified photograph uploading to related users without having their ID being identified.

The problem of domain adaptation for sentiment classifiers is studied by Glorot et al. in [132]. The authors have proposed a deep learning approach which learns to extract a meaningful representation for each review in an unsupervised fashion. Sentiment classifiers trained with this high-level feature representation have outperformed state-of-the-art methods on a benchmark composed of reviews of 4 types of Amazon products.

In [133], the authors have introduce a Sentiment Treebank. It included fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presented challenges for sentiment compositionality. They have introduced the Recursive Neural Tensor Network to address them. When trained on the treebank, the model outperformed all previous methods on several metrics. It also captured the effects of negation and its scope at various tree levels for both positive and negative phrases.

A way for feature extraction from text is proposed in [134]. Given a training corpus with hand-annotated sentiment polarity labels the authors have trained a DCNN on it. However, instead of using it as a classifier they have used the values from its hidden layer as features for a much more advanced classifier,

which gave superior accuracy.

In [135], authors presented machine learning approaches with regard to sentiment analysis in blog, review and forum texts found on the World Wide Web and written in English, Dutch and French. They have trained from a set of example sentences or statements that were manually annotated as positive, negative or neutral with regard to a certain entity. The main goal was to understand the feelings that people express with regard to certain consumption products. They have learned and evaluated classification models that could be configured in a cascaded pipeline.

Read et al. in [136] proposed a source of training data based on the language used in conjunction with emoticons in Usenet newsgroups. Training a classifier using this data has provided a breadth of features that, while it did not perform to the state-of-the-art, could function independent of domain, topic and time.

In [137], the authors defined document similarity metrics to enable online clustering of media to events. They explored techniques for learning multi-feature similarity metrics for social media documents in a principled manner and they have evaluated these techniques on large-scale, real-world datasets of event images from Flickr. The performances evaluation suggested that their approach identified events, and their associated social media documents.

The analysis of sentiment is of great importance even for determining the dynamics of health behaviours. In this context, the authors of [138] have collected a dataset of online social media users to measure the spatio-temporal sentiment towards a new vaccine. They have validated the proposed approach by identifying a strong correlation between sentiments expressed online and CDC-estimated vaccination rates by region. Analysis of the network of opinionated users showed that information flows more often between users who share the same sentiments and less often between users who do not share the same sentiments than expected by chance alone. They have also found that most communities were dominated by either positive or negative sentiments towards this vaccine.

In [139], Schwartz et al. analysed 700 million words, phrases, and topic instances collected from the Facebook messages of 75,000 volunteers, who also took standard personality tests, and found striking variations in language with personality, gender, and age. The analyses are useful especially on psychosocial processes yielding results that faced valid (e.g., subjects living in high elevations talk about the mountains). Their technique was the differential language analysis (DLA) based on three key characteristics. It is: Open-vocabulary (classified as a type of open-vocabulary approach), Discriminating (it found key linguistic features that distinguished psychological and demographic attributes, using stringent significance tests), Simple (it used simple, fast, and readily accepted statistical techniques).

Sentiment classification techniques were incorporated into the domain of mining reviews from travel blogs in [140]. In particular, the authors in this paper compared three supervised machine learning algorithms of Naïve Bayes, SVM and the character based N-gram model for sentiment classification of the reviews on travel blogs for seven popular travel destinations in the US and Europe. Empirical findings indicated that the SVM and N-gram approaches outperformed the Naïve Bayes approach, and that when training datasets had a large number of reviews, all three approaches reached accuracies of at least 80%.

2.4 Video Surveillance

Video surveillance is one of the most active research area in PR and computer vision. The main goal is to efficiently extract information from a huge amount of videos collected by surveillance cameras by automatically detecting, tracking and recognising objects of interest, and understanding and analysing their activities [141]. After September 11, 2001, preempting terrorist acts, and providing for the security of citizens at home and abroad have become top priorities not only for the United States but for many other nations around the world. For this purpose, a huge amount of information needs to be captured, processed, interpreted and then analysed. Video surveillance has several applications both in public and private environments, such as crime prevention, homeland security, traffic control, accident prediction and detection, and monitoring patients, elderly and children at home. These applications require monitoring indoor and outdoor scenes of airports, train stations, highways, parking lots, stores, shopping malls and offices. The increasing availability of sensors such as RGB-D cameras, of computer based technologies and the growing need for safety and security in public space have attracted interest in surveillance applications. As a result, advanced video-based surveillance systems have been developed by research groups from academia and industry alike. In broad terms, advanced video-based surveillance could be described as intelligent video processing designed to assist security personnel by providing reliable real-time alerts and to support efficient video analysis for forensics investigations.

The following sections present recent theoretical and practical advances in the broad area of PR techniques for advanced surveillance applications.

2.4.1 Algorithms and Approaches

The advances in computer vision, as well as machine learning and deep learning techniques in the recent years, have ameliorated the expedition towards surveillance and as a result, a plethora of algorithms for the automatic analysis of the

video sequences have been proposed. For video surveillance, machine learning is most applicable in the area of video analytics. The value and possibilities of applying machine learning to video surveillance footage are endless. The most obvious benefits would be the ability for cameras to learn what they are looking at and then react to anomalies or produce detailed reports on the type of activity they observe. For instance, a camera in the mall may learn what areas visitors typically congregate in and could alert security if a crowd develops in an area that is usually quiet. Data acquired through machine learning in video surveillance could ultimately be more valuable for marketing and operational use than for security. Machine learning techniques were found to provide valuable findings, with the main benefits being: high precision and accuracy in assigning broad categories to text data, the ability to identify and visualise discrete patterns from a large amount of data in scenario extraction, and the ability to make inferences about likely future outcomes based on models developed from existing data. Machine Learning algorithms have been used in three different steps; learning the color transformation among different cameras, creating a more discriminative signature and tuning the distance metric among samples [142].

However, there is exceedingly rich information and knowledge embedded in all those videos. With the recent advances in computer vision, we now have the ability to mine such massive visual data to obtain valuable insight about what is happening in the world. Due to the remarkable successes of deep learning techniques, we are able to boost video analysis performance significantly and initiate new research directions to analyze video content. For example, CNNs have demonstrated superiority on modeling high-level visual concepts, while RNNs have shown promise in modeling temporal dynamics in videos. Deep video analytics, or video analytics with deep learning, is becoming an emerging research area in the field of PR.

These strengths and the disadvantages of the PR techniqes employed in surveillance will be summarized in Table 2.3 turn with examples from the literature provided.

2.4.2 Applications

An automatic method to detect abnormal crowd density by using texture analysis and learning is presented by Wu in [149]. This task is important for the intelligent surveillance system in public places. By using the perspective projection model, a series of multi-resolution image cells were generated to make better density estimation in the crowded scene. The cell size was normalized to obtain a uniform representation of texture features. The authors have applied a technique of searching the extrema in the Harris-Laplacian space for

Table 2.3: PR approaches for Surveillance with their function, advantages, disadvantages, and recent works.

PR Approach	Function	Advantages	Disadvantages
Supervised Learning (e.g., [143, 144])	Depends on the existence of labeled training multimedia data.	High precision and accuracy.	Requires large amounts of labeled training videos/images that are very expensive.
Semi-Supervised Learning (e.g., [145])	Try to solve a supervised learning approach using labeled data augmented by unlabeled data.	Overcoming one of the problems of supervised learning which is having not enough labeled data.	Amplifies noise in labelled data.
Unsupervised Learning (e.g., [146, 147])	Datasets are assigned to segments, without the clusters being known.	There is no extensive prior knowledge of area required, but you must be able identify and label classes after the classification.	The user has to spend time interpreting and label the classes following the classification.
Deep Learning (e.g., [148])	Allow networks to have fewer weights and they are given a very effective tool, convolutions for image and video processing.	They use to need a lot of training data.	High computational cost.

diminishing the instability of texture feature measurements. The texture feature vectors were extracted from each input image cell and the SVM classifier was used to solve the regression problem of calculating the crowd density. Finally, based on the estimated density vectors, the SVM was employed again to solve the classification problem of detecting abnormal density distribution. The experiments were performed on real crowd videos.

Another work for the anomaly detection is the one proposed by Ahmed et al. [150]. In this paper, the authors have used two different datasets: pictures of a highway in Quebec taken by a network of webcams and IP traffic statistics from the Abilene network. These were examples in demonstrating the applicability of two machine learning algorithms to network anomaly detection: One-Class Neighbour Machine and the recursive Kernel-based Online Anomaly Detection algorithms.

In [151], in a visual surveillance task, the authors have described a real-time

computer vision and machine learning system for modeling and recognizing human behaviors. The system had the aim to detect when interactions between people occurred and it had to classify the type of interaction. Examples of interaction behaviours included following another person, altering one's path to meet another, and so on. The system combined top-down with bottom-up information in a closed feedback loop, with both components employing a statistical Bayesian approach. They have proposed and compared two different learning architectures such as HMMs and CHMMs for modeling behaviors and interactions. Finally, a synthetic "Alife-style" training system was used to develop flexible prior models for recognizing human interactions.

A framework for multi-camera video surveillance is proposed in [152]. The framework consisted of three phases: detection, representation, and recognition. The detection phase fused video streams from multiple cameras for extracting motion trajectories from video. The representation phase summarized raw trajectory data to construct hierarchical, invariant, and content-rich descriptions of the motion events. Finally, the recognition phase was with event classification and identification on the data descriptors. For effective recognition, they have performed a sequence-alignment kernel function to sequence data learning for identifying suspicious events. They have shown that when the positive training instances (i.e., suspicious events) were outnumbered by the negative training instances (benign events), then SVMs (or any other learning methods) suffered a high incidence of errors. To deal this problem, they suggested the kernel boundary alignment (KBA) algorithm to work with the sequence-alignment kernel.

Another work in the context of appearance-based person recognition was proposed in [153]. Pairwise dissimilarity profiles (functions of spatial location) between categories have learned and these were adapted into nearest neighbor classification. The aim was to better handle the ambiguities and to improve the scalability of classifiers to larger number of categories. To this end, the authors have introduced a dissimilarity distance measure and linearly or nonlinearly combine it with direct distances.

A deep network architecture was proposed in [154] in pedestrian detection. The authors have proposed in this work the interaction between feature extraction, deformation handling, occlusion handling, and classification since these components are learned and designed individually or sequentially. they have formulated these four components into a joint deep learning framework and propose a deep network architecture. They evaluated the performance on the largest Caltech benchmark dataset.

A deep learning approach was proposed for classifying human actions even in [155] without using any prior knowledge. The first step, based on the extension of CNN to 3D, automatically learned spatio-temporal features. An RNN

is then trained to classify each sequence considering the temporal evolution of the learned features for each timestep. The experiments were performed on the KTH dataset.

The recent topic to generate interest in surveillance is person re-identification (re-id) in image and video archives. People re-id aims to answer questions such as "Where have I seen this person before?" [156], or "Where has he gone after being caught on this surveillance camera?". The role of PR has become a central component in the re-id applications due to the learning aspects of this area.

Xiong et al. [157] have proposed the use, and have evaluated the performance, of four alternatives for re-id classification: regularized Pairwise Constrained Component Analysis, kernel Local Fisher Discriminant Analysis, Marginal Fisher Analysis and a ranking ensemble voting scheme, used in conjunction with different sizes of sets of histogram-based features and linear, $\chi 2^2$ and RBF-$\chi 2^2$ kernels.

In [158] the authors have proposed a more general way that can learn a similarity metric from image pixels directly. By using a "siamese" deep neural network, the method can jointly learn the color feature, texture feature and metric in a unified framework. The network had a symmetry structure with two sub-networks which were connected by a cosine layer. Each sub-network includes two convolutional layers and a fully connected layer. To deal with the big variations of person images, binomial deviance is used to evaluate the cost of similarities and labels, which was proved to be robust to outliers.

In [159], a supervised learning framework to generate compact and bit-scalable hashing codes directly from raw images has been proposed. Zhang et al. have posed hashing learning as a problem of regularized similarity learning.They have organized the training images into a batch of triplet samples, each sample containing two images with the same label and one with a different label. With these triplet samples, the margin between the matched pairs and the mismatched pairs in the Hamming space has been maximized. The DCNN was utilized to train the model in an end-to-end fashion, where discriminative image features and hash functions are simultaneously optimized.

2.5 Intelligent Retail Environment

The task of finding patterns in data coming from intelligent retail environment is not new. Traditionally, it was the objective of business analysts, who generally use statistical techniques. The aim of this activity, however, has recently changed. Widespread use of computer vision and integrated systems able to monitor shoppers in a shop has created large electronic databases that store business transactions. Retailers capture millions of sales transactions

through their point-of-sale terminals. Transactions can be analyzed to identify buying patterns of individual customers as well as customer groups, and sales patterns of different stores. Currently, artificial intelligence is pervading retail environments and technologies able to supply humans with supplementary knowledge for making better decisions have been conceived [160], [161]. The shop is becoming a digital ambient with the artificial intelligence that allows environments to be sensitive and adaptive to the human presence [162], [163]. Generally, computer vision and image processing demonstrated to hold great potential for retail practice and research [164]. Integrated systems able to monitor shoppers in intelligent retail environments have been developed with the aim of learning shopper skills [165], [166]. Data gathered from sensors installed in the shop are used with the aim of evaluating the *attraction* (the level of attraction that the store is creating on consumers), the *attention* (the time consumers spend in front of brand display) and the *action* (the number of consumers that enter in the store and interact with merchandise). These factors have changed the way that customers behaviour data are analyzed and given rise to data mining, which integrates machine learning, statistical analysis and visualization techniques, with the intuition and knowledge of the business analyst, to discover meaningful and interesting patterns in business data.

2.5.1 Algorithms and Approaches

Customer data consists of four dimensions [171, 172]:

- *Customer Identification*: is referred to as customer acquisition. This phase involves targeting the population who are most likely to become customers or most profitable to the company. Moreover, it involves analyzing customers who are being lost to the competition and how they can be won back [172]. Elements for customer identification include target customer analysis and customer segmentation. Target customer analysis involves seeking the profitable segments of customers through analysis of customers' underlying characteristics, whereas customer segmentation involves the subdivision of an entire customer base into smaller customer groups or segments, consisting of customers who are relatively similar within each specific segment [173].

- *Customer Attraction*: follows the customer identification phase. After identifying the segments of potential customers, retailers can direct effort and resources into attracting the target customer segments. An element of customer attraction is direct marketing that is a promotion process which motivates customers to place orders through various channels [174].

- *Customer Retention*: customer satisfaction, which refers to the compar-

Table 2.4: PR approaches for customer management, along with their function, advantages, disadvantages, and examples.

PR Approach	Function	Advantages	Disadvantages
Association (e.g. statistics and apriori algorithms [167])	Establishing relationships between items which exist together in a given record.	Greatly compress the candidate item sets and the size of the frequent item sets, and obtain good performance.	Requires many database scan.
Classification (e.g., neural networks, DT and if-then-else rules [168])	Building a model to predict future customer behaviours through classifying database records into a number of predefined classes based on certain criteria.	Ability to implicitly detect complex nonlinear relationships between dependent and independent variables.	Proneness to overfitting.
Clustering (e.g. neural networks and discrimination analysis) [168])	Segmenting a heterogeneous population into a number of more homogenous clusters.	Easy to implement.	Need to define many channels.
Forecasting (e.g. neural networks and survival analysis [168])	Estimating the future value based on a record's patterns.	The projections rely on the strength of past data.	Some forecasting methods may use the same data but deliver widely different forecasts.
Regression (e.g. linear regression and logistic regression [169])	Statistical estimation technique used to map each data object to a real value provide prediction value.	Large amounts of potential predictor variables management, fine-tuning the model to choose the best predictor variables from the available options.	Overfitting the Model.
Sequence discovery (e.g. statistics and set theory [170])	Identification of associations or patterns over time.	Maximize either the precision or the recall and limit the degradation of the other criterion.	

ison of customers' expectations with his or her perception of being satisfied, is the essential condition for retaining customers [172]. As such, elements of customer retention include one-to-one marketing, loyalty programs and complaints management. One-to-one marketing refers to personalized marketing campaigns which are supported by analysing, detecting and predicting changes in customer behaviours [175]. Thus, customer profiling, recommender systems or replenishment systems are related to one-to-one marketing. Loyalty programs involve campaigns or supporting activities which aim at maintaining a long term relationship with customers. Specifically, churn analysis, credit scoring, service quality or satisfaction form part of loyalty programs.

- *Customer Development*: involves expansion of transaction intensity, transaction value and individual customer profitability. Elements of customer development include customer lifetime value analysis, up/cross selling and market basket analysis. Customer lifetime value analysis is defined as the prediction of the total net income a company can expect from a customer [176]. Up/Cross selling refers to promotion activities which aim at augmenting the number of associated or closely related services that a customer uses within a firm [177]. Market basket analysis aims at maximizing the customer transaction intensity and value by revealing regularities in the purchase behaviour of customers [178].

These four dimensions can be seen as a closed cycle of a customer management system [179]. PR techniques, therefore, can help to accomplish such a goal by extracting or detecting hidden customer characteristics and behaviours from large databases, building a model from data [180]. Each PR technique can perform one or more of the following types of data modelling: association, classification, clustering, forecasting, regression and sequence discovery. The listed models cover the generally mentioned PR models in various articles [167]. There are numerous machine learning techniques available for each type of data mining model. The choices of PR methods should be based on the data characteristics and business requirements [180]. Some widely used PR algorithms in retail fields are: association rule, DT, Genetic algorithm, Neural networks, kNN and Linear/logistic regression.

A combination of PR models is often required to support or forecast the effects of a CRM strategy. In such a situation, the classification of data mining models mentioned in the article will be based on the major CRM issues that the article would like to address. For instance, in the case of up/cross selling programs, customers can be segmented into clusters before an association model is applied to each cluster. In such cases, the up/cross selling program would be classified as being supported by an association model because relationships be-

tween products are the major concern; in the case of direct marketing, a certain portion of customers may be segmented into clusters to form the initial classes of the classification model. The direct marketing program would be classified as being supported by classification as prediction of customers' behaviour is the major concern.

Table 2.4 summarizes the PR techniques and specific approaches in customer data management.

2.5.2 Applications

In [181], the authors report on metrics and classification methodologies that have been applied to large scale topographic data, that afford a systematic classification of certain retail spaces potentially at the national coverage. By analysing the form, composition, extent and patterns of buildings within retail spaces, together with their degree of centrality and levels of access, they have demonstrated that it was possible to classify different types of retail space. Three methodologies used for classification (Boolean, fuzzy logic and Bayesian modelling) were compared and were evaluated through comparison with known locations of various retail types as a way of assessing the validity of these approaches.

Jambulingam et al. have developed organization clusters as intangible resources based on entrepreneurial orientation to classify organizations within a retailing industry. They have tested if the entrepreneurial orientations of the resultant groups within the pharmacy industry were related to their perception of the environment, organizational factors, and performance outcomes, using retail pharmacy industry data [182].

In [183], the authors have tried to predict online consumer repurchase intentions. They have chosen a hybrid approach with a combination of machine learning techniques and artificial bee colony (ABC) algorithm. They have performed a classification process and they have evaluated the performance of DT, AdaBoost, RF, SVM and neural network for predicting consumer purchase intention.

In [184], the main purpose was to predict product churn, defined as a significant increase in interpurchase time. In this way, it is possible to identify products which deserve additional marketing attention. The authors have worked with data from a European low-cost food and non-food retailer, from which they have extracted product characteristics based on one year of transaction, product, customer and payment data. The resulting feature set consisted of recency, frequency and monetary (RFM) values, price details, payment details, etc.

An accurate, real-time and fully automated system that relies on computer

vision and deep learning to identify returning customers and comprehensively filter sales and customer information was developed by Song et al. in [185]. The system depended on local computation resources to perform customer detection, tracking and feature extraction. Whereas age and gender estimation, and customer recognition was performed on cloud computing resources to compensate for the limited computation resource at local store and to accommodate multiple stores requirements.

Instead, the authors in [186] have presented a DSS for retail pricing and revenue optimization of retail products. They have used a regression tree/random forest-based machine learning algorithm to predict weekly demand. Price, discounts, holidays, inventory and other regional factors have been incorporated in decision making. The demand-price interdependencies were quantified and integrated into an integer linear programming model for optimal price allocation. This methodology has been implemented on offline retailing of expensive products which generally follow high variation in demand. The expected revenue has been optimized by branch & bound and branch & cut method, followed by root node analysis. Furthermore, the solution was optimized by heuristic methods.

EMOMETRIC, an intelligent trolley that can track customer's emotion and provide a customer behavioural insight through IoT integrated data intelligence running on Apache Spark Cluster, is proposed in [187]. This system has used model based face and emotion tracking under real use case conditions. The method adapted QoS enabled secure MQTT protocol to collect the data by the big data No-sql storage system. The authors have inferred that this technique is not only fast and accurate but also illumination and pose invariant.

In [188], Abrams et al. have reconsidered the problem of recovering exogenous variation from an endogenous regressor. Two-stage least squares recovers the exogenous variation through presuming the existence of an instrumental variable. They have assumed that the regressor was a mixture of exogenous and endogenous observations–say as the result of a temporary natural experiment. In this light, they have proposed an alternative two-stage method based on nonparametrically estimating a mixture model to recover a subset of the exogenous observations. The authors have demonstrated that the method recovered exogenous observations in simulation and can be used to find pricing experiments hidden in grocery store scanner data.

Shaohui et al. [189] have considered the basic requirements for a promotional DSS, i.e. reliance on operational (store-level) data only, the ability to predict sales as a function of prices and the inclusion of other promotional variables affecting the category. The model delivered an optimizing promotional schedule at Stock-Keeping-Unit (SKU) level which maximized multi-period category level profit under the constraints of business rules typically applied in practice.

Firstly, they have developed a high dimensional distributed lag demand model which integrated both cross-SKU competitive promotion information and cross-period promotional influences. then, based on the demand model, they have built a nonlinear integer programming model to maximize the retailer's category profits over a planning horizon under constraints that modeled important business rules. The output of the model provided optimized prices, display and feature advertising planning together with sales and profit forecasts.

In [190], the authors have predicted the Optimal price of smart-phones by combining the results of sentiment analysis and artificial neural network models. Firstly, they have used the sentimental analysis to convert the non-uniform and unclean data to linear numerical values that exactly represent the content of the inputs. The review section of the product is the major area which required sentimental analysis where it needed to interpret and reciprocate the feeling of the user in numerical role. Then, smartphones data were used to train the neural network.

2.6 Digital Cultural Heritage

The digital component is now pervasive in CH research and practice. From the pioneering and sporadic applications of the last quarter of the 20^{th} century, it has grown to become an essential for any CH investigation or management project. This development has led to the creation of a new interdisciplinary domain, the so-called Digital Heritage sector, and the need to support, structure, and manage the use of facilities, resources, services and applications no less than in other research domains, producing long-lasting infrastructures. Such infrastructures may address specific aspects of DCH, such as visualization or documentation, or cover the full range of heritage-related activities, from investigation to conservation, management, education and communication.

Originally digital heritage technology often comprised an exercise in the application of computer techniques originally conceived for other goals; the specific needs of such use has now dictated new requirements and has led to the development of new tools and methods tailored for heritage applications. With the vast expansion of digital contemporary painting collections, automatic theme stylization has grown in demand in both academic and commercial fields. The recent interest in deep neural networks has provided powerful visual features that achieve state-of-the-art results in various visual classification tasks.

The following Subsection briefly describes the algorithms and approach in this research area and some of the built applications.

2.6.1 Algorithms and Approaches

The application of learning methods for content-based curation and dissemination of cultural heritage data offers unique advantages for physical sites at risk of damage. In recent years, innovative techniques from computing, computer vision, image and natural language processing to analyse images and enable semantic are used in this context. Outputs can be multimedia and automated reports of the state of repair of cultural artifacts as well as real-time, elucidating comments for site visitors.

PR approaches that include machine learning and statistical classification are applied with the aim of assisting preservation endeavours. They incorporate multimodal data analysis, and content-based augmented data retrieval. The suitability of machine learning and semantic technologies for the documentation of cultural heritage is demonstrated in [191]. Furthermore, the applicability of deep learning for digital documentation of cultural heritage has been reported [192].

Table 2.5 summarizes the PR techniques and specific approaches in DCH domain.

Table 2.5: PR approaches for DCH, along with their function, advantages, disadvantages, and recent examples.

PR Approach	Function	Advantages	Disadvantages
Supervised Learning (e.g., [193, 194])	The algorithm will then learn the relationship between the images and their associated numbers, and apply that learned relationship to classify completely new images (without labels) that the machine has not seen before.	Precise model with predictive and interpretative properties.	Requires equally large number of examples from each class.
Unsupervised Learning (e.g., [195, 196])	Learn powerful features in an unsupervised way and.	Does not require class labels on data.	Sensitive to similarity measure; results difficult to interpret.
Semi-supervised Learning (e.g., [197])	Learn model from mixture of labeled and unlabeled data.	Utilize all available data; typically outperforms use just labeled data.	Sensitive to errors in propagating class labels from labeled to unlabeled data.
Deep Learning (e.g., [198, 199])	Learns complex representations of concepts in the data.	General purpose and high accuracy.	Sensitive to parameter choices; long training times.

2.6.2 Applications

The artistic content of historical manuscripts is challenging in terms of automatic text extraction, picture segmentation and retrieval by similarity. Grana et al. have addressed the problem of automatic extraction of meaningful pictures, distinguishing them from handwritten text and floral and abstract decorations [200]. They have proposed a solution that firstly employed a circular statistics description of a directional histogram in order to extract text. Then visual descriptors were computed over the pictorial regions of the page: the semantic content was distinguished from the decorative parts using color histograms and a texture feature called Gradient Spatial Dependency Matrix. The feature vectors were processed using an embedding procedure which allows increased performance in later SVM classification.

In [201], the authors have addressed the problem of identifying artistic styles in paintings, and have suggested a compact binary representation of the paintings. They have tried to recognize the style of paintings using features extracted from a deep network. The features suggested in the paper have shown excellent classification results on a large scale collection of paintings.

The automatic images classification from 50 different cultural events is performed in [202]. The proposed solution was based on the combination of visual features extracted from CNN with temporal information using a hierarchical classifier scheme. The authors have also proposed a late fusion strategy that trained a separate low-level SVM on each of the extracted neural codes. The class predictions of the low-level SVMs formed the input to a higher level SVM, which gave the final event scores.

In [203], the authors have presented an approach to discover the characteristic features that determined an artist's touch. By training PigeoNET, a CNN, on a large collection of digitized artworks to perform the task of automatic artist attribution, the network was encouraged to discover artist-specific visual features. The trained network was capable of attributing previously unseen artworks to the actual artists with an accuracy of more than 70%. The trained network also provided fine-grained information about the artist-specific characteristics of spatial regions within the artworks.

Wavelet transforms and machine learning tools can be used to assist works of art in the stylistic analysis of paintings. In this context, Jafarpour et al. have used image processing and machine learning techniques, to tackle two stylometry problems with a dataset collected in the Van Gogh Museum and the Kroller-Muller Museum in the Netherlands and consisting of high resolution scans of paintings by Vincent van Gogh [204]. They have shown how modeling style as a hidden variable, controlling the behaviour of the image observables, such as brushstrokes, color patterns, etc, can improve the accuracy of the style analyzer to a significant extent. They have used a dual-tree complex

wavelet transform that is shift invariant, to capture quantitatively the effects observable in the image. Next, using Hidden Markov Trees, an extension of Hidden Markov Variables, combined with the expectation maximization algorithm, they have extracted the style parameters from the noisy observables. Then, using standard machine learning methods, they classified the extracted features.

In [205], the authors have used telemetry data from Oztoc, an open-ended exploratory tabletop exhibit in which visitors embody the roles of engineers who are tasked with attracting and cataloging newly discovered aquatic creatures by building working electronic circuits. This data was used to build HMMs to devise an automated scheme of identifying when a visitor is behaving productively or unproductively. Evaluation of HMM was shown to effectively discern when visitors were productively and unproductively engaging with the exhibit. Using a Markov model, they have identified common patterns of visitor movement from unproductive to productive states to shed light on how visitors struggle and the moves they made to overcome these struggles. These findings offer considerable promise for understanding how learners productively and unproductively persevere in open-ended exploratory environments and the potential for developing real time supports to help facilitators know how and when to best engage with visitors.

In [206], Polatkan et al. have demonstrated that supervised machine learning on features derived from hidden-Markov-tree-modeling of the paintings' wavelet coefficients has the potential to distinguish copies from originals in the dataset proposed. In fact, they have provided a ground truth data set in which originals and copies were known and image acquisition conditions are uniform.

An approach to automatically classify digital pictures of paintings by artistic genre is described in [207]. The authors' evaluation used variable resolution painting data gathered across Internet sources rather than solely using professional high-resolution data. They have also included a comparison to existing feature extraction and classification methods as well as an analysis of their approach across classifiers and feature vectors.

Another work that focused on classifying works of seven different artists, by using a multi-class SVM with state-of-the-art features is proposed by Blessing [208]. Even in this case, machine learning has good potential to classify artworks.

In [209], Li et al. have addressed the learning-based characterization of fine art painting styles. They have compared the painting styles of artists. To profile the style of an artist, a mixture of stochastic models is estimated using training images. The two-dimensional (2D) multiresolution hidden Markov model (MHMM) was used in the experiment. These models formed an artist's distinct digital signature. The 2D MHMM analyzed relatively large regions

in an image, which in turn makes it more likely to capture properties of the painting strokes. The mixtures of 2D MHMMs established for artists can be further used to classify paintings and compare paintings or artists. They have implemented and tested the system using high-resolution digital photographs of some of China's most renowned artists.

2.7 Pattern Recognition Applications

Table 2.6 summarizes the ten interesting applications described for each studied domain in this Chapter and it introduces the challenging thesis applications that will be the focus of Chapter 3.

Taking into account the state of art discussed in this chapter, Figure 2.4 highlights in red the step of the pipeline depicted in Figure 1.1 that represent the efforts devoted in this thesis for the contribution in each field.

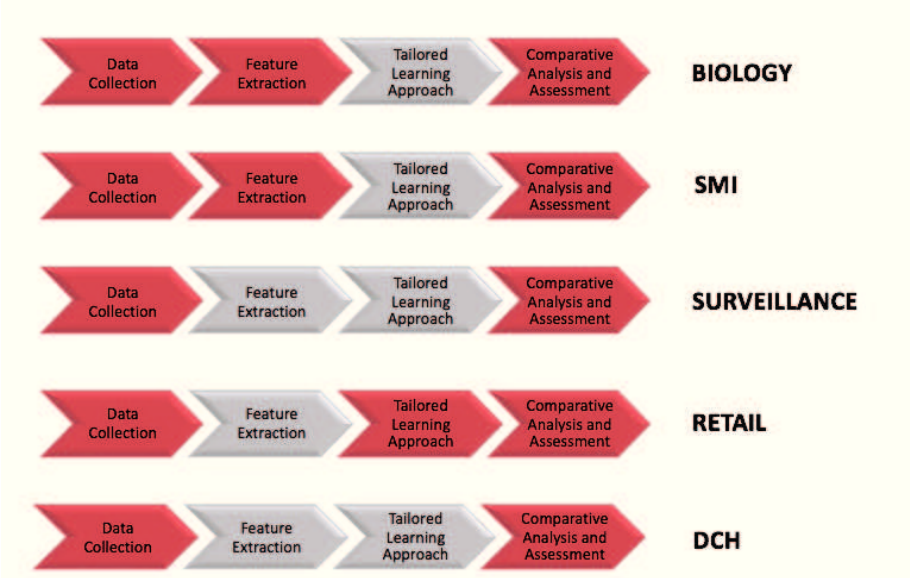


Figure 2.4: Pipelines in which are highlighted the main contributions of this thesis for each domain taken into exam.

In particular, the applications of PR techniques are devoted to real-world problems yielding new insights that advance PR methods. In each research field, extensive efforts are devoted to collecting training and testing data and five newly challenging datasets are specifically designed for the described task. Finally, various details have to be investigated such as domain dependence and prior information, computational cost and feasibility, discriminative features, similar values for similar patterns, different values for different patterns, in-

variant features with respect to translation, rotation and scale, robust features with respect to occlusion, distortion, deformation, and variations in environment. Then, performances with training sample are estimated, performances with future data are predicted and problems of overfitting and generalization have been evaluated.

More details are available in the following chapters.

Table 2.6: PR Applications.

Problem Domain	**Application**	**Input Pattern**	**Pattern Classes**	**Method**	**Reference**
Biology 1	PPI Automatic detection	PPI domain-dependent features	text classification	Classification	[107]
Biology 2	Automated chromosome karyotyping	Chromosome image features	Chromosome classification	Classification	[108]
Biology 3	Parasitemia visual quantification	Erythrocytes images	Automatic identification of infected erythrocytes	Classification	[109]
Biology 4	Cancer Classification	Spectra peaks	Cancer type or subtype	Classification	[110]
Biology 5	CAD system for mammographic masses	screening mammograms and digital breast tomosynthesis	Detection of breast masses	Classification	[111]
Biology 6	PR-system designed by Dermofit images	Digital plain photography images	Comparison of melanomas lesion	Classification	[112]
Biology 7	CAD for prostate cancer	Ultrasound images	Detection of prostate carcinoma	HMMs	[113]
Biology 8	CBIR for a database of parasite specimen images	Parasite specimen images	Identification to general category of "species group"	Querying methods	[114]
Biology 9	Application in assistance to Neuroprostheses	Video and eye-tracking data	Real time object recognition from ego-centric videos	DCNN	[115]
Biology 10	Medical image segmentation	Multispectral data	Brain tissue classification	VB approach	[116]
Biology Thesis Application	**Assisted Reproductive Technology**	**GCs**	**Oocytes for successful pregnancy**	**Classification**	[210], [211]
SMI 1	Face recognition for Twitter application	2D images	Face recognition	Classification	[131]
SMI 2	Domain adaptation for sentiment classifiers	Reviews of Amazon products	Sentiment Analysis	Deep Learning	[132]
SMI 3	Semantic Compositionality	Corpus of movie review	Sentiment Classification	Deep Learning	[133]
SMI 4	Utterance-Level Multimodal Sentiment Analysis	Short video fragments	Sentiment Analysis	Deep Learning	[134]
SMI 5	Sentiment analysis in multilingual Web texts	Movie reviews	Sentiment Analysis	Classification	[135]
SMI 6	Emoticons to reduce Dependency ifor Sentiment Classification	Text marked-up with emoticons	Sentiment classification	Classification	[136]

SMI 7	Event Identification in Social Media	Social media documents	Event identification	Clustering	[137]
SMI 8	Vaccination Sentiments with Online Social Media	Tweets	Sentiment analysis	Vaccination sentiments	[138]
SMI 9	Automatic Lexical Analysis	Words, phrases, and topic instances	DLA	Language of Social Media	[139]
SMI 10	Sentiment classification of online reviews to travel destinations	Reviews	Sentiment classification	Classification	[140]
SMI Thesis Application	**Sentiment Analysis of brand-related pictures**	**Visual and textual features**	**people brand sentiment**	**Deep Learning**	[212]
Surveillance 1	Crowd density estimation	Crowd videos	Abnormal crowd density detection	Classification	[149]
Surveillance 2	Network anomaly detection	Pictures, source IP address, destination IP address, source port number and destination port number	Anomaly detection	Classification	[150]
Surveillance 3	Computer Vision System for Modeling Human Interactions	Pedestrian Images	Human Interaction Detection	HMMs, CHMMs	[151]
Surveillance 4	Framework for multi-camera video surveillance	Video	Suspicious events identification	Classification	[152]
Surveillance 5	Appearance Recognition in Visual Surveillance	Images	Handling the ambiguities	Classification	[153]
Surveillance 6	Joint Deep Learning for Pedestrian Detection	Image	Pedestrian Detection	Deep Learning	[154]
Surveillance 7	Human action Recognition	Action Videos	Human actions classification	Deep Learning	[155]
Surveillance 8	Person re-id	Videos and Images	Person recognition	Classification	[157]
Surveillance 9	Deep Metric Learning for re-id	Pedestrian images	Person recognition	Deep Learning	[158]
Surveillance 10	Framework to generate compact and bit-scalable hashing codes from raw image	Images	Image retrieval and persogn re-id	Deep Learning	[159]
Surveillance Thesis Application	**Top-view re-id**	**Colour and anthropometric features**	**Surveillance Systems**	**Classification**	[213, 214]

Retail 1	Retail Spaces	Topographic Data	Automatic Classification of Retail Spaces	Classification	[181]
Retail 2	Strategic entrepreneurial activities	Entrepreneurial orientation	Organizations Classification	Clustering	[182]
Retail 3	Online consumer repurchase intention	Consumer Personal Characteristics and Online shopping malls attributes	Online consumer repurchase intentions prediction	Classification	[183]
Retail 4	Social products behaviour	Recency, frequency and monetary (RFM) values, price details, payment details	Product churn prediction	Classification	[184]
Retail 5	Customer recognition system	Images	Returning customers Identification	Deep Learning	[185]
Retail 6	DSS for retail pricing and revenue optimization of retail products	Sales data	Demand prediction	Classification	[186]
Retail 7	EMOMETRIC	Body images volumes	Customer behavioural insight	Classification	[187]
Retail 8	Exogenous variation from an endogenous regressor	Retail scanner data	Pricing experiments Identification	Unsupervised Machine Learning Algorithm	[188]
Retail 9	Promotional DSS	Grocery and drug chain data	Sales forecasts	Classification	[189]
Retail 10					
Retail Thesis Application	**Intelligent Retail Environment**	**Trajectories**	**Users behaviour in store**	**HMMs and Clustering**	[215, 216]
DCH 1	Segmentation of digitalized historical manuscripts	Digitalized images	Automatic segmentation of text and decorations	Classification	[200]
DCH 2	Classification of Artistic Styles using Binarized Features	Unique digitized paintings with variable resolution	Artistic Style Classification	Deep Learning	[201]
DCH 3	Cultural Event Recognition	Instagram images	Image classification	Deep Learning	[202]
DCH 4	Learning to recognize artists by their artworks	Digital photographic reproductions of artworks	Discovery of the Artist's Style	Deep Learning	[203]

DCH 5	Stylistic analysis of paintings	High resolution color scans of the paintings	Style analysis	Classification	[204]
DCH 6	Modeling Visitor Behavior in a Game-Based Engineering Museum Exhibit	Visitors' interactions	Patterns of exploration identification	HMM	[205]
DCH 7	Detection of forgery in paintings	Paintings	Image Classification	Classification	[206]
DCH 8	Classifying Paintings by Artistic Genre	Paintings	Digital pictures of painting Classification	Classification	[207]
DCH 9	Identification of Art Paintings	set of paintings	Art identification	Classification	[208]
DCH 10	Digital imagery of ancient paintings by mixtures of stochastic models	Photos	Paintings Classification	MHMM	[209]
DCH Thesis Application	**AR users behaviour**	**eye trajectory**	**museum users behaviour**	**HMMs**	[217, 218]

Chapter 3

Use cases and Results on challenging Computer Vision Applications.

As outlined in the previous chapter, PR deals with designing and developing algorithms based on empirical data. PR has the ability to adapt to new circumstances and detect and extrapolate pattern. The aim of the research activity conducted during the years of the Ph.D. studies by the author, is mainly devoted to the exploitation of cutting edge scientific methodologies for the solution of problems of relevant interest. In particular, it is oriented to bring together applications of PR in order to test a wide landscape of techniques that can be successfully applied and also to show how such techniques should be adapted to different particular domains. This for two main reasons: first of all, even if the main activities where conducted into the retail domain, it was not sufficient to understand the limits and benefits of this paradigm only by exploring a single topic. Secondly, different domains have different needs, so that to understand the potential of PR, it was fundamental to face with different scenario and, eventually, to outline points of contact. Following PR technologies to address the increasingly complex challenges as the main thread, it is selected a set of experiences carried out in different fields; at the center of this investigation there are of course real scenarios with intelligent applications developed in five domains, with research issues explored in terms of their design, implementation, integration, and deployment. The key challenge is to bring PR into reality with applications, taking advantage of the implicit or explicit human-machine interactions. Among real-world applications there are: **Biology** applied to the prediction of the best oocyte for a pregnacy successfully for the Assisted Reproductive technology (ART); **Retail** where the trajectories of shopping carts and baskets are collected and analysed for modelling and forecasting customers behaviour; **Surveillance** to determine if different instances or images of the same person, recorded in different moments, belong to the same subject, i.e. the re-identification (re-id) process; **Social Media intelligence**

for evaluating the goldmine of information coming from social network and finally **Digital Cultural Heritage** where eye-tracking data of people observing a painting are collected and analysed for optimazing Augmented Reality application (AR).

3.1 Biology

In recent years, machine learning, deep learning and data mining techniques have been successfully applied in several medical and biomedical domains, such as the prognosis and diagnosis of cancers, the detection of tumors, and other complex diseases [219, 220, 221, 222].

In medical fields, reproductive technologies are achieving increasing importance. Every year, a number of couples (10–15%) reveal problems connected to infertility and a growing number of them seeks the help of ART laboratories [223]. The main goal of ART is the ability to obtain a large number of competent oocytes. Oocytes are specialized cells produced in the ovary during oogenesis. Their competence is the ability to be fertilized and to develop as an embryo [224]. The oogenesis has characterized and regulated by molecular mechanisms that are not only controlled within the oocyte itself, but also through a complex molecular cross-talk between the oocyte and the surrounding somatic cells [225]. For this reason, oogenesis can also be considered as a continuous and highly integrated process, which includes both the developing oocyte and the surrounding somatic cells. Currently, in ART implantation, there is not a criteria for the oocyte choice. Its selection is based on the morphological features of the cytoplasm, polar body and cumulus cells. However, all these criteria for grading and screening oocytes are subjective and controversial, and seem to not be related to the intrinsic competence of the oocyte [226, 227]. Over the past couple of decade, even with the improvement, the success rates for procedures treating infertility and leading to live births is still less than optimal. From 2011, European data for pregnancy rate was 33% with the pregnancy rate per embryo thaw cycle being 21%, as reported by the European IVF-Monitoring Consortium and the European Society of Human Reproduction and Embryology showed the non-donor embryo transfer (ET) [228].

In the last few years, in medicine, for studying cells, tissues, and biological fluids has been widely applied vibrational spectroscopy. In particular, in the reproductive field, Fourier Transform Infrared Microspectroscopy (FTIRM) is a powerful technique to study the composition and the macromolecular chemistry of cells and tissues, providing the biochemical composition of female gametes. Chemical cartograms are generated by infrared mapping and imaging techniques, based on peak height, integrated areas under specific bands or band ra-

tios, giving a semi-quantitative evaluation of sample biocomponents [229, 230].

The identification of the best quality oocyte is very important because it will increase overall pregnancy rates and the accurate prediction of embryos, that will successfully implant, decreases the risks of abortion and the complications related to it. In addition to, the wrong oocyte can have unintended effects like the risk of multiple gestations, that have many complications such as low birth weight, physical deficits, developmental delays, and costs.

3.1.1 Automatic classification of human oocytes in Assisted Reproductive technology.

In order to assure an appropriate chance of successful implantation, to minimize any complications during the pregnancy, it has been applied for the first time FTIRM for the assessment of oocyte quality by the study of Granulosa Cells (GCs) collected along with the oocytes during oocyte aspiration, as it is routinely done in ART. FTIRM enables the collection of a comprehensive set of macromolecular data for characterizing GCs, by studying the position and relative intensities of the vibrational bands composing IR absorption spectra [231]. The number of collected samples allowed the discrimination of the vibrational biomarkers of a good quality oocyte from the ones of poor quality oocytes, and also from those related to the intrinsic spectroscopic variability, that is a consequence of sample heterogeneity. Moreover, it is performed a series of feature selection procedures to identify new spectral biomarker signatures, measuring a large set of spectral information, and then employed data analysis procedures to select the most informative. Spectral biomarkers evidencing changes related to concentration, distribution and structure of several biomolecules of interest (lipids, proteins, carbohydrates and nucleic acids) will be taken into account and linked to cellular metabolism, DNA methylation defect, apoptosis/autophagy process, oxidative stress status and trascriptomic profiling. All these spectral evidences will be validated by conventional genomic and proteomic tools, statistically related and classified as biomarkers of good quality oocyte on the basis of clinical results/outcomes obtained during IVF (In Vitro Fertilization) procedure of the corresponding oocyte. 17 spectral biomarkers related to the concentration, distribution and structure of lipids, proteins, carbohydrates and nucleic acids, as indicators of cellular metabolism, DNA methylation, apoptosis/autophagy, oxidative stress, etc.. are taken into account. The GCs samples were classified into 4 groups:

- **A**: GCs from good quality oocytes (clinical pregnancy);
- **B**: GCs from bad quality oocytes (due to fertilization failure);
- **C**: GCs from bad quality oocytes (due to embryo failure);

- **D**: GCs from bad quality oocytes (due to implantation failure).

These groups can be summarized in two groups regarding the characteristics of the GCs cells:

- Group **A**, GCs from good quality oocytes;
- Group **NOT A**, GCs from bad quality oocytes.

The spectral biomarkers of the above described two groups were used to create a robust biological reference dataset (BRD), which was retrospectively developed by analysing GCs from oocytes of known quality and validated by clinical outcomes obtained during the ART procedure of the corresponding oocyte.

Regarding the lack of a criteria in ART and considering the previous experience in the developing data standardization system in medicine and the secure sharing between them [232], machine learning techniques are used for predicting oocyte quality for a pregnancy successfully. It is reported the results of a comprehensive evaluation of 4 state-of-the-art classifiers, such as Support Vector Machine (SVM) [10], and k-Nnearest Neighbor (kNN) [5, 233], Decision Tree (DT) [234] and Random Forest(RF) [13].

The approach for oocytes quality evaluation can be described in the following steps:

- FTIR spectroscopy on GCs collected along with the oocytes during oocyte aspiration;
- BRD Development and Features Set;
- Classification Model.

The proposed approach has been schematically depicted in Figure 3.17.

The main novelties are: i) the first machine learning approach for learning and classifying oocytes quality; ii) the proposal of a significant set of biofeatures able to achieve relevant classification performances; iii) the first comparison between different classifiers to evaluate a family of machine learning algorithms able to obtain a significant results in this field.

Considering the important contribution and the innovation of this work in ART, these biological aspects and other related issues are patent pending.

3.1.2 Biological Reference Dataset

In Assisted Reproductive Technology(ART), biomarkers of GCs associated with individual oocytes are used to indicate which embryos have the best chance of implanting in the uterus and completing gestation. The Biological Reference

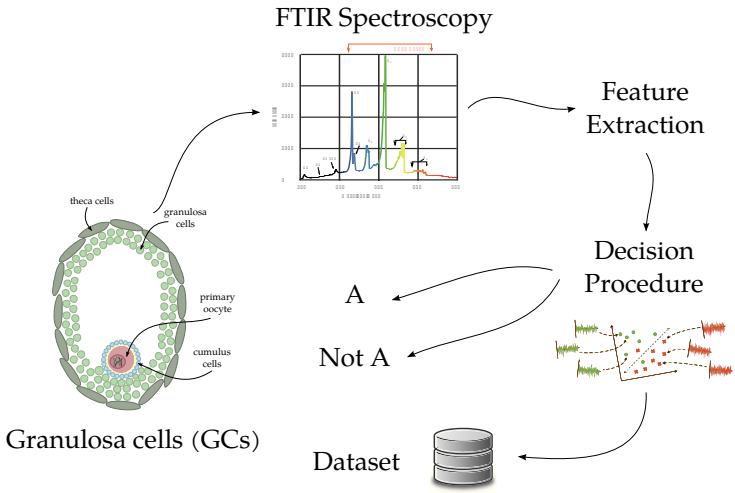


Figure 3.1: General Schema

Dataset (BRD) is built to predict the best oocyte for a successful pregnancy by analyzing vibrational data with multivariate statistical methods.

Firstly, oocyte quality was correlated with specific changes of concentration, distribution and structure of several biomolecules (lipids, proteins, carbohydrates and nucleic acids) in Granulosa Cells (GCs) by using Fourier Transform Infrared Microspectroscopy (FTIM) (Table 3.1).

A meaningful number of GCs spectral markers are obtained by FTIRM spectroscopy for oocyte quality evaluation in patients requiring ART treatment due to fertility, embryo implantation and pregnancy failure. This method comprises 17 spectral biomarkers related to the concentration, distribution and structure of lipids, proteins, carbohydrates and nucleic acids, as indicators of cellular metabolism, DNA methylation, apoptosis/autophagy, oxidative stress, etc..

Table 3.2 describes 17 spectral biomarkers. These are used as classification process features.

The dataset is composed by 2614 instances:

- Group **A**, GCs from good quality oocytes: 1176 instances;
- Group **NOT A**, GCs from bad quality oocytes: 1438 instances.

3.1.3 Performance evaluation and Results

To evaluate the dataset, a 10-fold cross-validation has been applied to ensure the robustness of performance estimate [235]. The performance of different classifiers and feature sets was evaluated in terms of precision, recall and F1-score using weighted macro-averaging over 10 folds. The information about

Table 3.1: Biomolecules used for the construction of BRD

Name	Wavenumber range (cm-1)	Chemical significance	Biological significance
Lipids	2990-2899	CH2/CH3 asymmetric and symmetric stretching	Aliphatic chains (lipids)
Amide I (AI)	1723-1591	Peptide bond's C=O stretching	Proteins
Amide II (AI)	1591-1481	Peptide bond's N-H bending and C-N stretching	Proteins
Proteins	1723-1481	Sum of AI and AII	Totality of proteins
Ph1	1273-1191	P02- asymmetric stretching	Phosphate groups
Ph2	1137-1022	P02- asymmetric stretching	Nucleic acids-specific phosphate groups
COO	1765-1723	COO stretching	Ester groups in lipid molecules
1460	1480-1426	CH3 bending	CH3 terminal group of aliphatic chains,and of protein's lateral chains
1400	1426-1362	CH2 bending	CH2 groups of aliphatic chains and,protein's lateral chains
CH	3027-2995	=CH stretching	Signal of unsaturation
CH2	2946-2899	CH2 stretching	CH2 groups of aliphatic chains
CH3	2992-2948	CH3 stretching	CH3 terminal group of aliphatic chains
970	984-946	DNA sugar-phosphate skeletal vibrations	Amount of DNA
Cell	2990-2899 + 1775-1191	Sum of vibrations indicated above	Cell mass

Table 3.2: Biological features used for oocytes classification

Ratio's Name	Biological significance
Lipids/Cell	Amount of lipids in the cell
Proteins/Cell	Amount of proteins in the cell
AI/AII	Comparison secondary structure of proteins
Lipids/Proteins	Amount of lipids compared to proteins
Ph1/Cell	Degree of phosphorylation in the cell
Ph1/Proteins	Degree of phosphorylation of proteins
COO/Cell	Amount of ester groups in the cell
COO/Lipids	Degree of esterification of lipids
1400/Proteins	Amount of CH2-rich aminoacids related to the protein content
1460/Proteins	Amount of CH3-rich aminoacids related to the protein content
1400/1460	In aliphatic chains: length of the chain; in proteins: indicator of CH2-rich aminoacids
1460/Lipids	Amount of CH3 groups in lipids
CH/Cell	Rate of unsaturation in the cell
CH/Lipids	Unsaturation levels of fatty acids
CH/CH3	Unsaturation levels of fatty acids
CH2/Lipids	Length of aliphatic chains
CH2/CH3	Length of aliphatic chains

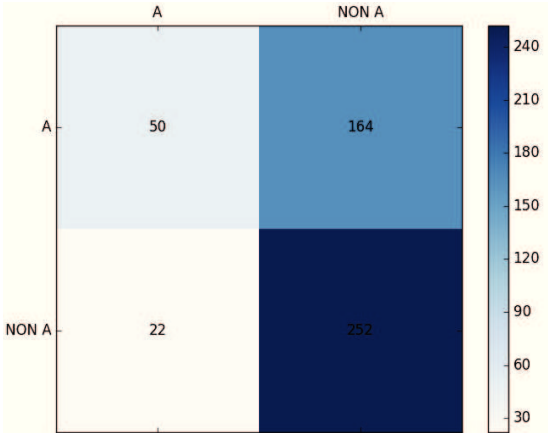


Figure 3.2: Confusion Matrix - SVM

actual and predicted classifications done by a oocytes classification system is depicted by confusion matrix [236]. The classification process is performed with kNN, SVM, Decision Tree and Random Forest classifiers. A cross-validation is carried using weighted macro-averaging over 10 folds. The task is solved using a SVM with a quadratic degree of the polynomial kernel function. For kNN classifier has been chosen "minkowski" as metric distance and "n neighbors = 5". Table 3.10 reports the classification results for the classifiers used in system prediction. Results are presented in terms of precision, recall and F-measure.

Figure 3.2 shows the confusion matrix for SVM classifier, Figure 3.3 depicts the confusion matrix for kNN classifier, Figure 3.4 represents the confusion matrix for kNN classifier and Figure 3.5 reports the confusion matrix for RF classifier.

Results show the effectiveness and suitability of the proposed approach. In fact, even if there are a few data, due to the low successful rate of pregnancy, this approach is very promising. The values of precision, recall and F-measure

Table 3.3: Classification results.

SVM	**Precision**	**Recall**	**F1-score**	**Support**
A	0.69	0.23	0.35	214
NOT A	0.61	0.92	0.73	214
AVG/Total	0.64	0.62	0.56	488
kNN	**Precision**	**Recall**	**F1-score**	**Support**
A	0.58	0.57	0.57	214
NOT A	0.67	0.68	0.67	214
AVG/Total	0.63	0.63	0.63	488
Decision Tree	**Precision**	**Recall**	**F1-score**	**Support**
A	0.64	0.66	0.65	214
NOT A	0.73	0.71	0.72	214
AVG/Total	0.69	0.69	0.69	488
Random Forest	**Precision**	**Recall**	**F1-score**	**Support**
A	0.67	0.73	0.70	214
NOT A	0.77	0.72	0.74	214
AVG/Total	**0.73**	**0.72**	**0.72**	**488**

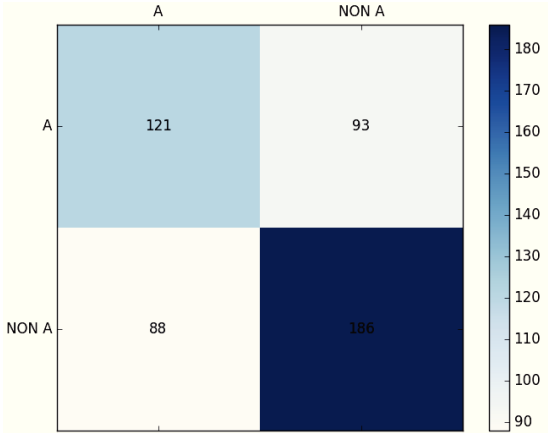


Figure 3.3: Confusion Matrix - kNN

are good.

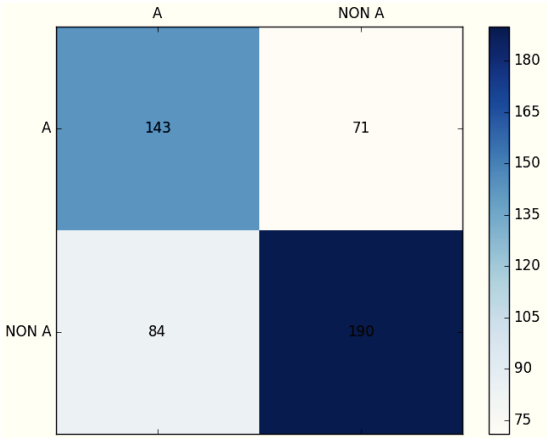


Figure 3.4: Confusion Matrix - Decision Tree

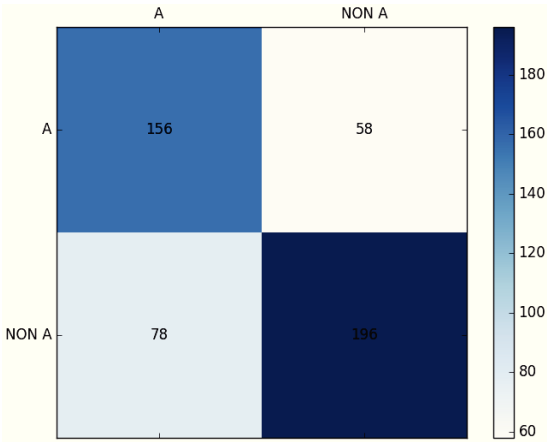


Figure 3.5: Confusion Matrix - Random Forest

3.2 Social Media Intelligence

The advent of Social Media has enabled everyone with a smartphone, tablet or computer to easily create and share their ideas, opinions and contents with millions of other people around the world. Recent years have witnessed the explosive popularity of image-sharing services such as Instagram[1] and Flickr[2]. These images do not only reflect people social lives, but also express their opinions about products and brands. Social media pictures represent a rich source of knowledge for companies to understand consumers' opinions [237]. The multitude of pictures makes a manual approach infeasible and increases

[1] www.instagram.com
[2] www.flickr.com

the attractiveness of automated sentiment analysis [238], [239].

In the past, companies conducted consumer surveys for this purpose. Although well-designed surveys can provide high quality estimations, they can be time-consuming and costly, especially if a large volume of survey data is gathered [240]. In contrast, social media pictures are available in real time and at low costs and represent an active feedback, which is of importance not only to companies developing products, but also to their rivals and potential consumers [241]. Algorithms to identify sentiment are crucial for understanding consumer behaviour and are widely applicable to many domains, such as retail [213], behaviour targeting [217], and viral marketing [125].

Sentiment analysis is the task of evaluating this goldmine of information. It retrieves opinions about certain products and classifies them as positive, negative, or neutral. Existing research papers [242], [243], have focused on sentiment analysis of textual postings such as reviews in shopping platforms and comments in discussion boards. However, with the increasing popularity of social networks and image sharing platforms [244], [245] more and more opinions are expressed by pictures. Several researchers have now started to propose solutions for the sentiment analysis of visual content. However, a multitude of consumers' pictures does not only include visual elements, but also textual elements. For example, people take pictures of advertisement posters or insert text into photos with the aid of photo editing software. In order to estimate the overall sentiment of a picture, it is essential to not only judge the sentiment of the visual elements but also to understand the meaning of the included text. While a picture showing a cosmetic product next to a cute rabbit might be positive, the same picture containing the words "animal testing" might be negative.

An approach to estimate the overall sentiment of a picture based on both visual and textual information is introduced. While many studies have performed sentiment analysis, most existing methods focus on either only textual content or only visual content. This is the first approach to consider visual and textual information in pictures at the same time. The sentiment of a picture is identified by a machine learning classifier based on visual and textual features extracted from two specially trained Deep Convolutional Neural Networks (DCNNs). The visual feature extractor is based on the VGG16 network architecture [246] and it is trained by fine-tuning a model pretrained on the ImageNet dataset [70]. While the visual feature extractor is applied to the whole image, the textual feature extractor detects and recognizes texts before extracting features. The textual feature extractor is based on the DCNN architecture proposed by [247] and is created by fine-tuning a model which has been previously trained on synthesized social media images. Based on these features, six state-of-the-art classifiers, namely kNearest Neighbors (kNN) [16], [84], Support Vector Ma-

chine (SVM) [10], Decision Tree (DT) [234], Random Forest (RF) [13], Naïve Bayes (NB) [17] and Artificial Neural Network (ANN) [18], [248], are compared to recognize the overall sentiment of the images.

The approach has been applied to a newly collected dataset "GfK Verein Dataset" of consumer-generated pictures from Instagram which show commercial products. This dataset comprises 4200 images containing visual and textual elements. In contrast to many existing datasets, the true sentiment is not automatically judged by the accompanying texts or hash-tags but has been manually estimated by human annotators, thus providing a more precise dataset. The application of the proposed approach to this dataset yields good results in terms of precision, recall and F1-score and demonstrates the effectiveness of the proposed approach.

3.2.1 Visual and Textual Analysis of brand-related social media pictures

In this section, the joint visual and textual sentiment analysis framework is introduced as well as the dataset used for evaluation. The framework is depicted in Figure 3.6 and comprises three main components: the visual feature extractor, the textual feature extractor, and the overall sentiment classifier. Especially trained DCNNs for visual and textual feature extraction are used. The visual and textual features are fused and fed into the overall sentiment classifier. Common machine learning algorithms are compared for the overall sentiment classification.

The framework is comprehensively evaluated on the "GfK Verein Dataset", a proprietary dataset collected for this work[3]. The details of the data collection and ground truth labeling are discussed in subsection 3.2.2.

Visual feature extractor

The visual feature extractor aims at providing information about the visual sentiment of a picture and is therefore trained with image labels indicating the visual sentiment of the images. The training is performed by fine-tuning a VGG16 net [246] that has been pre-trained on the ImageNet dataset [70] to classify images into 1000 categories. A fine-tune is performed by cutting off the final classification layer (fc8) and replacing it by a fully connected layer with 3 outputs (one for each sentiment class). In addition, the learning rate multipliers are increased for that layer so that it learns more aggressively than all the other layers. Finally, loss and accuracy layers are adapted to take input from the new fc8 layer. Since the image classifier serves as feature extractor, the

[3]This work is done in GfK Verein `http://www.gfk-verein.org`

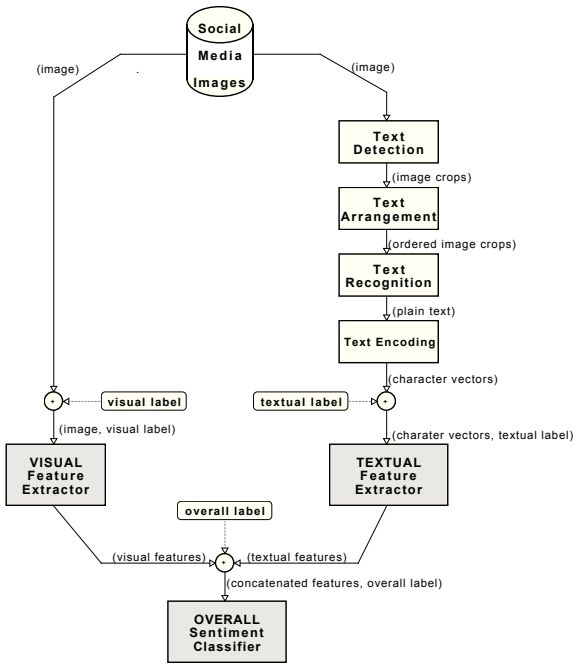


Figure 3.6: Training pipeline flow

output of the next to last fc7 layer is passed to the overall sentiment classifier. The image feature extractor is implemented using standard Caffe[4] tools.

Textual feature extractor

The goal of the textual feature extractor is to provide information about the textual sentiment of a picture. It is therefore trained with image labels indicating the textual sentiment of the images. The textual feature extractor consists of multiple components. The central component is a character-level DCNN with an architecture as described in [247], which has been extended by one additional convolution layer. The extra convolution layer, inserted before the last pooling layer, has a kernel size of 3 and produces 256 features. The textual feature extractor was trained in two phases: first training a base model on synthesized social media images and then fine-tuning that base model on the new dataset. In order to generate training data for the base model, accompanying captions from brand-related social media pictures were inserted into social media pictures in varying fonts, font-sizes, colors and slight rotations. Since the text is embedded in the picture as pixels, the text has to be transformed to characters before it can be processed by the character-level DCNN.

[4]http://caffe.berkeleyvision.org/

Table 3.4: Number of features

Model	Layer	Number of features
image	fc7	4096
text	ip4	1024

The following steps are performed:

1. *Text Detection*: individual text boxes are detected in an image with the TextBoxes Caffe model [249].
2. *Text Arrangement*: detected text boxes are put in order based on a left-to-right, top-to-bottom policy, thus forming logical lines.
3. *Text Recognition*: each text box is processed by the OCR model [250] to transcribe the text of the box.
4. *Text Encoding*: the recognized text is encoded into one-hot vectors based on the alphabet of the character-level DCNN.

The textual features of the next to last layer of the character-level DCNN are passed to the final sentiment classifier.

Overall Sentiment Classifier

On the basis of the visual and textual features, the overall sentiment classifier aims at estimating the overall sentiment of an image. For this purpose, it is trained with labels indicating the overall sentiment of the images. The number of visual and textual features is illustrated in Table 3.4.

Based on the fused features, six state-of-the art classifiers, namely kNN, SVM, DT, RF, NB and ANN are used to recognize the overall sentiment of the images and compared with respect to precision, recall and F1-score.

3.2.2 GfK Verein Dataset

This work, to the best of author knowledge, is the first study on sentiment analysis of brand-related pictures on Instagram. Instagram provides a rich repository of images and captions that are associated with users' sentiments. A visual and textual sentiment dataset is constructed from the pictures on Instagram. The captions of the Instagram posts are utilised to pre-select images that have detectable sentiment content about well-known brands from the industry of fast moving consumer goods. Typically, the image captions indicate the users' sentiment for the uploaded images. The "GfK Verein Dataset" is composed of brand-related social media images as follows:

(a) (b) (c)

Figure 3.7: Brand Related Social Media Pictures of "GfK Verein Dataset". Figure 3.10a is an example of a picture with overall negative sentiment, Figure 3.10c represents an image with overall neutral sentiment, and Figure 3.10d is a picture with overall positive sentiment

- 1400 images with positive sentiment;
- 1400 images with neutral sentiment;
- 1400 images with negative sentiment.

To obtain the ground truth of the collected pictures, the true sentiment has been manually estimated by human annotators, thus providing a more precise and less noisy dataset compared to automatically generated labels from image captions or hashtags. All pictures are annotated with respect to their visual, textual and overall sentiment.

Figure 3.10 shows three examples of brand-related social media pictures of "GfK Verein Dataset". As can be seen, the overall sentiment towards a brand or product does not only depend on the visual content of a picture but also on its textual content.

Since sentiment estimation is a subjective task where different persons may assign different sentiments to images, it is asked two persons to judge the sentiment of the images and measured their agreement. The inter-annotator-agreement is a common approach to determine the reliability of a dataset and the difficulty of the classification task [251]. It is calculated Cohen's Kappa Coefficient k which measures the agreement between two annotators beyond chance [252]. The values of Kappa range from -1 to 1, with 1 indicating perfect agreement, 0 indicating agreement expected by chance, and negative values indicating systematic disagreement. The inter-annotator-agreement for the visual ($k = 0.82$), textual ($k = 0.82$) and overall ($k = 0.84$) sentiment assignment is high, assuring good quality of the dataset and feasibility of the machine learning task.

Table 3.5: Performance of the visual DCNN model, predicting visual sentiment based only on visual features

Category	Precision	Recall	F1-Score
positive	0.83	0.82	0.82
neutral	0.86	0.89	0.88
negative	0.72	0.67	0.69
MEAN()	***0.81***	***0.79***	***0.80***

3.2.3 Performance Evaluation and Results

In this section, the results of the experiments conducted on "GfK Verein Dataset" are reported. In addition to the performance of the overall sentiment classifier, it is presented the performance of the visual and textual sentiment classifiers which form the basis of the visual and textual feature extractors and are key to the overall sentiment classification.

The experiments are based only on these images of the dataset, where both annotators have agreed on the overall, visual and textual sentiment. By removing pictures with ambiguous sentiment, it is increased the quality of the dataset and ensure the validity of the experiments. The final dataset is comprised of a total amount of 3452 pictures, including 1149 pictures with overall positive sentiment, 1225 pictures with overall neutral sentiment and 1078 pictures with overall negative sentiment.

The experiments are performed by splitting the labeled dataset into a training set and a test set. Each classifier will only be trained based on the training set. Likewise, the test set is also fixed in the beginning and used for all test purposes. The dataset is split into 80% training and 20% test images, taking into account all permutations of overall, visual, and textual annotations.

In order to create the visual feature extractor it is trained a DCNN to classify the visual sentiment of a picture. The performance of the visual sentiment classification is reported in Table 3.5. As can be seen, high values of precision and recall can be achieved, especially for pictures with positive and neutral visual sentiment. The recognition of visually negative pictures is more difficult due to the smaller amount of available training data and the higher variation in motives. Consumers tend to express their overall negative sentiment towards brands by adding negative text to neutral or positive motives. As people avoid posting pictures with negative facial expressions on social media, the most frequent form of visual negative sentiment is graphics with many different motives.

For creating a textual feature extractor, it is trained a DCNN to estimate the sentiment of the text in the pictures. Table 3.6 depicts precision and recall of the textual sentiment classification. The performance of the textual sentiment classification is good, but lower than the performance of the visual

Table 3.6: Performance of the textual DCNN model, predicting textual sentiment based only on textual features

Category	Precision	Recall	F1-Score
positive	0.71	0.68	0.70
neutral	0.84	0.61	0.71
negative	0.67	0.89	0.76
MEAN()	***0.74***	***0.73***	***0.74***

Table 3.7: Performance of the overall classifier, predicting overall sentiment based on both visual and textual features

Classifier	Precision	Recall	F1-Score
NB	0.72	0.72	0.72
DT	0.72	0.72	0.72
RF	0.74	0.74	0.74
SVM	0.77	0.77	0.77
kNN	0.78	0.78	0.78
ANN	**0.79**	**0.79**	**0.79**

sentiment classification. While the judgment of visual and textual sentiment is equally difficult for humans, the classification of text in pictures is much more challenging for machines as the text has to be detected and recognized first before it can be classified, thus being more error-prone. Comparing the different classes reveals that negative and neutral texts can be recognized better than positive texts. This fact is also reflected by the characteristics of the dataset. As consumers prefer visual clues such as happy people or smileys to textual clues for showing their overall positive sentiment towards brands, positive texts are less expressive.

Based on the visual and textual features, a machine learning classifier is trained to identify the overall sentiment of a picture. It is trained several classifiers, namely SVM, DT, NB, RF, and ANN and compare their performance for different parameter settings. Table 3.7 reports the results of the best parameter setting for each classifier. As can be seen, the performance of all classifiers is good, with F1-Scores ranging from 0.72 for NB to 0.79 for ANN, thus demonstrating the effectiveness and the suitability of the proposed approach. The performance of the overall sentiment classification is much higher than the performance of the textual sentiment classification but slightly lower than the performance of the visual sentiment classification. This comparison shows that recognizing the overall sentiment is more challenging than only the visual sentiment. Estimating the overall sentiment, however, is crucial for understanding consumers' attitudes towards brands. Relying on the visual sentiment only can be misleasing in many cases since consumers often embed text in their pictures to verbalize their sentiment. Especially, overall negative sentiments are often expressed by adding negative text to neutral or positive visual motives.

3.3 Video Surveillance

Camera installations are widespread in several domains that range from small business and large retail applications, to home surveillance applications, environment monitoring, facility access, sports venues, and mass-transit. Identification cameras are widely employed in most of public places like malls, office buildings, airports, stations and museums. In these applications, it is desirable to determine if different instances or images of the same person, recorded in different moments, belong to the same subject. This kind of process is commonly known as "person re-identification" (re-id). It has a wide range of utilities and is of great commercial value.

Research in people behaviour analysis has been thoroughly focused on person re-id during the last decade, which has seen the exploitation of many paradigms and approaches of PR [142]. In challenging situations, algorithms need to be robust to deal with issues such as widely varying camera viewpoints and orientations, rapid changes in appearance of clothing, occlusions, varying poses and varied lighting conditions [253, 254].

The first studied re-id problem was related to vehicle tracking and traffic analysis, where objects move in well defined paths, have almost uniform colours and are rigid. Features as colour, speed, size, lane position are generally embedded in Bayesian frameworks. However, person re-id requires more elaborate methods in order to deal with the widely varying degrees of freedom of a person's appearance [255].

Much of the research on person re-id has been devoted to model the human appearance. In fact, descriptors of image content have been proposed in order to discriminate identities while compensating for appearance variability due to changes in illumination, pose, and camera viewpoint. Re-id is also a learning problem in which either metrics or discriminative models are actually learned [256, 254]. Labelled training data are required for metric learning approaches and new training data are needed whenever a camera setting changes [257].

Recently, person re-id is emerging as a very interesting field and future solutions could be exploited as a tool for the efforts devoted to the development of systems that can carry out detection and tracking of people, in the case of occlusion or where there is a partial camera coverage. Moreover, person re-id is an important part of modern surveillance systems. In fact, the amount of data generated by such systems is often too large to be manually analysed. Automatic and semi-automatic re-id approaches can help speed up this process by presenting the operator with likely matches of a query person. The re-id task has been used for long-term tracking: people should be tracked as long as possible, using one cameras. Thanks to the short (or null) temporal gap

between samples, geometric and positional features are usually enough and the requirements on the camera quality and resolution are loose. Video sequences are usually available as input and processed using the common surveillance chain composed by background/foreground segmentation and intracamera object tracking [142].

This aspect can be closely linked to the use of RGB-D cameras that provides affordable and additional rough depth information coupled with visual images, offering sufficient accuracy and resolution for indoor applications. This camera has already been successfully applied in the surveillance field to univocally identify people in a scene and to analyse behaviours and interactions between them.

The choice of the RGB-D camera in a top view configuration is due to its greater suitability compared with a front view configuration, usually adopted for gesture recognition or even for video gaming. The top-view configuration reduces the problem of occlusions and has the advantage of being privacy preserving, because a person's face is not recorded by the camera [258]. Top-view people counting applications are the most accurate (with accuracy up to 99%) even in very crowded scenarios; this camera point of view is also the only one that allow to measure at the same time people passing by and interactions among shoppers and products on the shelf [165].

Currently, several datasets are available for the study of person re-id and cover the many aspects of this problem, such as shape deformation, occlusions, illumination changes, very low resolution images and image blurring, etc. [259]. The most famous are VIPeR [260], iLIDS Multi Camera Tracking Scenario [261], ETHZ [262] and CAVIAR4REID [263]. Another re-id dataset is proposed in [264]. Data are gathered using RGB-D technology, but are not suitable for the listed purposes.

Due to this fact and against this background, for re-id evaluation, a new dataset has been built for person re-id that uses an RGB-D camera in a top-view configuration: the TVPR dataset. An Asus Xtion Pro Live RGB-D camera has choosen because it allows the acquisition of colour and depth information in an affordable and fast way [265]. The camera was installed on the ceiling above the area to be analysed. This dataset collects the data of 100 people, acquired across intervals of days and in different times.

3.3.1 Person Re-Identification with an RGB-D camera in Top-view configuration

A method for person re-id is introduced. It is based on a set of features extracted by the colour and depth images that is used to perform a classification process, selecting the first or initial passage under the camera for training and

using returns to the initial position as the testing set. Furthermore, a gender classification, focused on colour and length hair, is carried out with the aim to improve surveillance applications. All feature sets using kNearest Neighbors (k-NN), Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RF) classifiers are tested. The performance evaluation demonstrates the effectiveness of the proposed approach, achieving good results in term of Precision, Recall and F1-score.

The first step involves the processing of the data acquired from the RGB-D camera. The camera captures depth and colour images, both with dimensions of 640×480 pixels, at a rate up to approximately $30\,fps$ and illuminates the scene/objects with structured light based on infrared patterns.

Seven out of the nine features selected are *anthropometric features* extracted from the depth image: distance between floor and head, d_1; distance between floor and shoulders, d_2; area of head surface, d_3; head circumference, d_4; shoulders circumference, d_5; shoulders breadth, d_6; thoracic anteroposterior depth, d_7. The remaining two *colour-based features* are acquired by the colour image. It is also defined TVH, TVD and $TVDH$. TVH is the colour descriptor:

$$TVH = \{H_h^p, H_o^p\} \tag{3.1}$$

TVD is the depth descriptor:

$$TVD = \{d_1^p, d_2^p, d_3^p, d_4^p, d_5^p, d_6^p, d_7^p\} \tag{3.2}$$

Finally, $TVDH$ is the the signature of a person defined as:

$$TVDH = \{d_1^p, d_2^p, d_3^p, d_4^p, d_5^p, d_6^p, d_7^p, H_h^p, H_o^p\} \tag{3.3}$$

Colour is an important visual attribute for both computer vision and human perception. It is one of the most widely used visual features in image/video retrieval. To extract these two features it is used HSV histograms. Local histograms have proven to be largely adopted and are very effective. The signature of a person is also composed by two colour histograms computed for head/hairs and outerwear: H_h^p, H_o^p in eq. 3.1, such as in [266], with $n = 10$ bin quantization, for both H channel and S channel.

Figure 3.8 depicts the set features considered: anthropometric and colour-based.

For the experiments, person re-id classification is performed selecting the first passage under the camera for training and using a reset to the initial position as the testing set. All feature sets using k-Nearest Neighbors (kNN) classifier [5], Support Vector Machine [10, 11, 12], Decision Tree [14] and Random Forest [13] are tested and the performance in terms of precision, recall and F1-score are

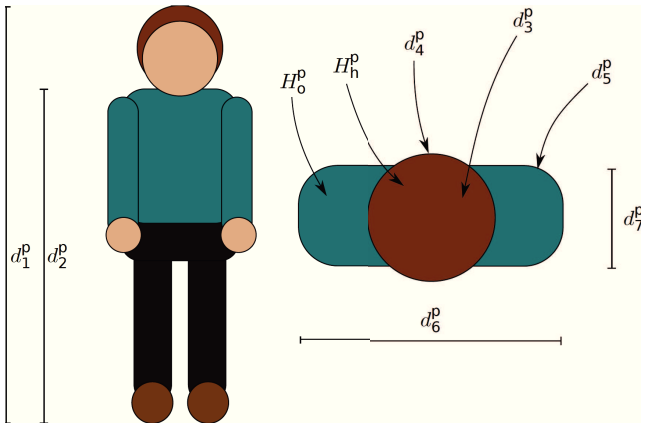


Figure 3.8: Anthropometric and colour-based features.

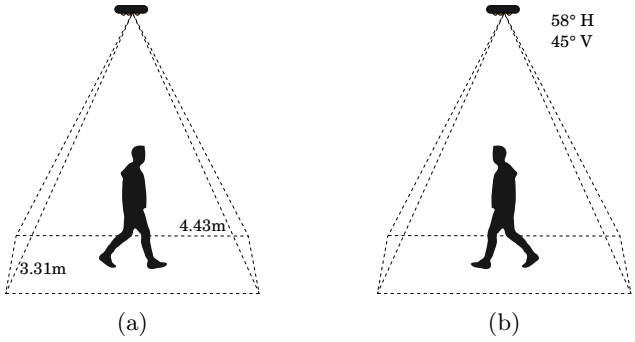


Figure 3.9: System architecture.

evaluated.

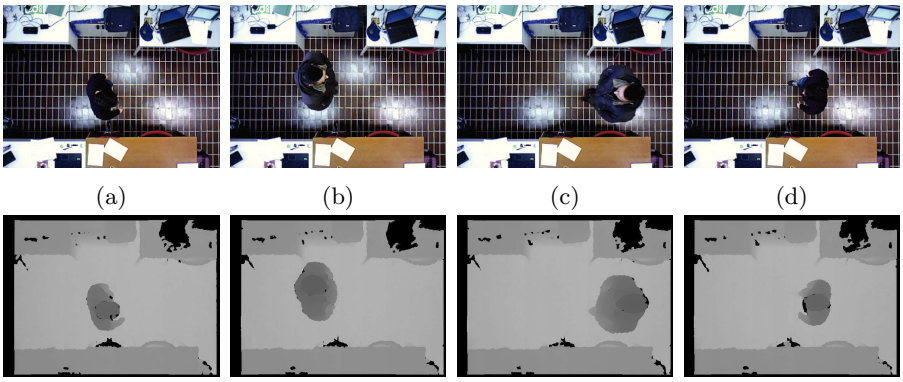
3.3.2 TVPR Dataset

The proposed representation has been experimentally validated on TVPR (Top View Person Re-identification) dataset [5] for person re-id [267].

TVPR contains videos of 100 individuals recorded in several day from an RGB-D camera installed in a top-view configuration. The camera is installed on the ceiling of a laboratory at $4\,m$ above the floor and covers an area of $14.66\,m^2$ ($4.43\,m \times 3.31\,m$). The camera is positioned above the surface which is to be analysed (Figure 3.9).

The 100 people of TVPR dataset were acquired in 23 registration session. Each of the 23 folders contains a video of one registration sessions. Acquisitions have been performed in 8 days and the total recording time is about 2000

[5] `http://vrai.dii.univpm.it/re-id-dataset`



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3.10: Snapshots of a registration session of the recorded data, in an indoor scenario, with artificial light. People passed under the camera installed on the ceiling. The sequence 3.10a-3.10e, 3.10b-3.10f corresponds to the sequence 3.10d-3.10h, 3.10c-3.10g respectively training and testing set of the classes `8-9` for the registration session `g003`.

seconds.

Registrations are made in an indoor scenario, where people pass under the camera installed on the ceiling. Another big issue is environmental illumination. In each recording session, the illumination condition is not constant, because it varies in function of the different hours of the day and it also depends on natural illumination due to weather conditions. The video acquisitions, in the scenario, are depicted in Figure 3.10, which are examples of person registration respectively with sunlight and artificial light.

Each person during a registration session walked with an average gait within the recording area in one direction subsequently turning back and repeated over the same route in the opposite direction. This methodology is used for a better split of the TVPR in training set (the first passage of the person under the camera) and testing set (when the person passes a second time under the camera).

3.3.3 Performance Evaluation and Results

To evaluate TVPR dataset, the performance results are reported in terms of recognition rate, using the CMC curves, as previously described in [267]. Figure 3.11 depicts a comparison among *TVH* and *TVD* in terms of CMC curves, to compare the ranks returned by using these different descriptors, where the horizontal axis is the rank of the matching score, the vertical axis is the probability of correct identification.

In particular, Figure 3.11a represents the CMC obtained for *TVH*. Figure 3.11b provides the CMC obtained for *TVD*. It is compared these results with the average obtained by *TVH* and *TVD*. The average CMC is displayed in Figure 3.11d.

It can be assumed that the best performance is achieved by the combination of descriptors. It is possible infer this aspect from Figure 3.11d where the combination of descriptors improve the results obtained by each of the descriptor separately. This result is due to the depth contribution that may be more informative. In fact, the depth outperforms the colour measure, giving the best performance for rank values higher than 15 (Figure 3.11b). Its better performance suggests the importance and potential of this descriptor. Considering TVPR dataset, it is depicted a comparison among *TVH* and *TVD* in terms of CMC curves, to compare the ranks returned by using these different descriptors.

The *CMC* curve represents the expectation of finding the correct match in the top n matches. It is equivalent of the ROC curve in detection problems. This performance metric evaluates recognition problems, through some assumptions about the distribution of appearances in a camera network. It is considered the primary measure of identification performance among biometric researchers.

As well-established in recognition and in re-id tasks, for each testing item it is ranked the training gallery elements using standard distance metrics. It is examined the effects of 3 distance measures as the matching distance metrics: the L1 City block, the Euclidean Distance and the Cosine Distance.

Figure 3.11a provides the CMC obtained for *TVH*. Figure 3.11b represents the CMC obtained for *TVD*. These results are compared with the average obtained by *TVH* and *TVD*. The average CMC is displayed in Figure 3.11d.

It is observed that the best performance is achieved by the combination of descriptors. In Figure 3.11d, it can be seen that the combination of descriptors improve the results obtained by each of the descriptor separately. This result is due to the depth contribution that may be more informative. In fact, the depth outperforms the color measure, giving the best performance for rank values higher than 15 (Figure 3.11b). Its better performance suggests the importance and potential of this descriptor.

The classification process is performed with kNN, SVM, DT and RF classifiers. Two experiments: are carried out: a classic training/testing experiment and a gender classification, both based on TVPR dataset.

The first task is solved using as a TVD descriptor an SVM with a quadratic degree of the polynomial kernel function, while the others descriptors are solved with SVM with a cubic degree of the polynomial kernel function. For the kNN classifier the “minkowski” as metric distance and “n neighbors = 5” has been chosen.

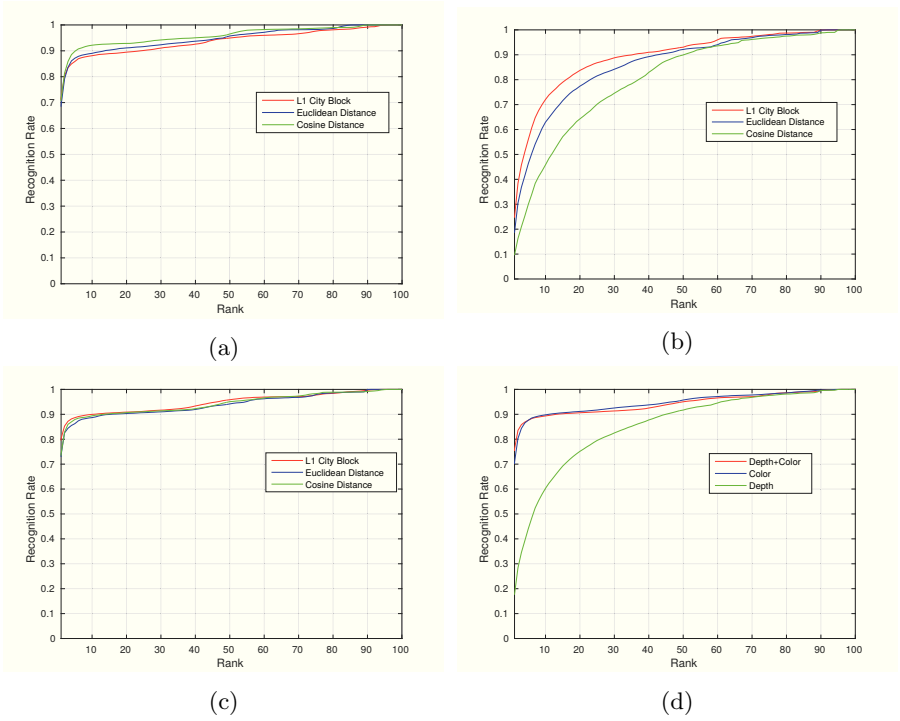


Figure 3.11: The CMC curves obtained on TVPR Dataset.

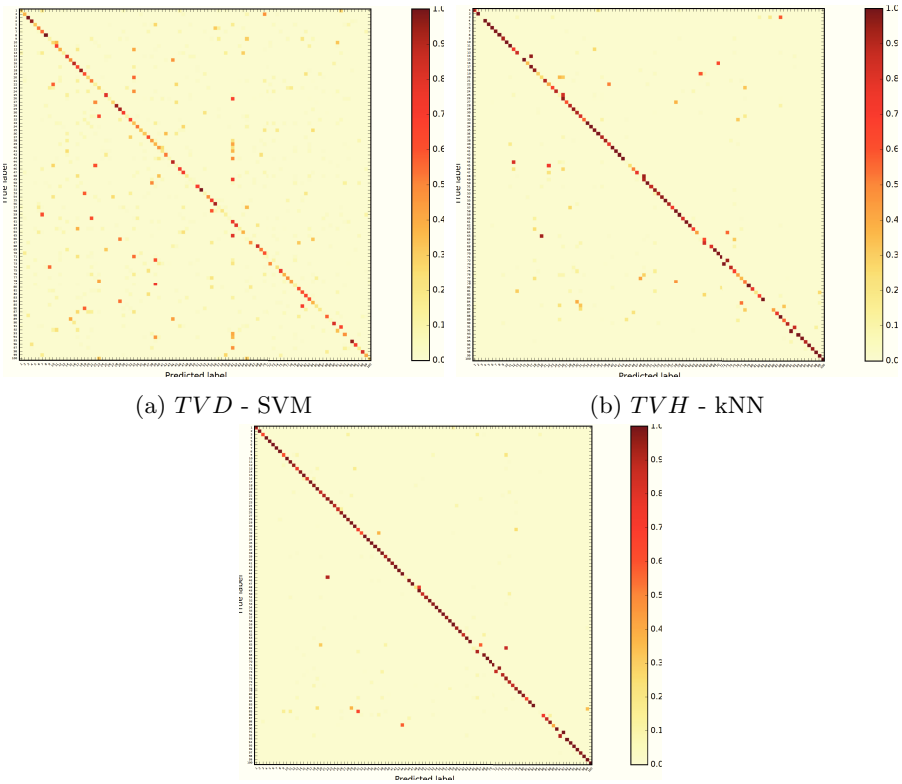
For the first case, it is considered the first passage under the camera as training set and the going back in the initial position as the testing set. The dataset is composed by 21685 instances divided in 11683 for training and 10002 for testing.

Table 3.8 reported, for each person of TVPR, the recognition results for kNN classifier with the TVDH descriptor.

The re-id classification performance of TVPR is summarized in Table 3.8 with a comparison among the descriptors TVH, TVD and TVDH. Figure 3.12 are the best confusion matrices for the three descriptors: TVD with SVM classifier (Figure 3.12a, TVH with kNN classifier (Figure 3.12b) and TVDH with kNN classifier (Figure 3.12c).

In this case, high performance for the proposed approach to re-identify people is observed. This accentuates the feasibility of utilizing colour as an effective cue in re-id scenarios. Moreover, by conducting the comparative study for the two descriptors TVD and TVH, it is possible to observe the influence of colour for the re-id top view scenario. However, TVD descriptor is important for re-id, because it improves the overall precision as Figure 3.12c shows.

In the second experiment, the aim is to classify gender considering the length



(a) TVD - SVM (b) TVH - kNN

(c) $TVDH$ - SVM

Figure 3.12: Confusion Matrices.

of hair and colour. The results is summarized in Table 3.10. Figure 3.13 depicts the confusion matrix for the kNN classifier.

Results shows the effectiveness and the suitability of the proposed approach. In fact, the class FSD "Female with dark and short hair" is confused, because females commonly have hair with considerable length. Same thing goes for class MLD "Male with dark and long hair", because generally short hair is an Italian male hairstyle. For the other class, classification overall precision is over 76%.

3.4 Intelligent Retail Environment

In recent years, a lot of efforts have been devoted to the retail space optimization and to the item assortment in that space. The reasons are manifold and easy to understand. From the retailer's viewpoint, shelf space is an important asset and the assortment variety increases inventory cost. Assortment is

Table 3.8: Classification results for each person of TVPR for kNN classifier with the TVDH descriptor.

ID	Precision	Recall	F1-S	Sup.	ID	Precision	Recall	F1-S	Sup.
1	0.90	0.85	0.87	53	51	0.84	0.20	0.33	103
2	0.70	0.74	0.72	43	52	0.58	1.00	0.73	110
3	1.00	0.91	0.95	54	53	0.99	0.87	0.93	100
4	0.90	1.00	0.95	69	54	1.00	0.94	0.97	101
5	0.93	0.98	0.95	86	55	0.99	1.00	0.99	94
6	1.00	0.95	0.98	109	56	0.92	0.97	0.94	67
7	0.85	0.98	0.91	63	57	0.99	1.00	1.00	105
8	1.00	1.00	1.00	102	58	1.00	1.00	1.00	76
9	1.00	1.00	1.00	86	59	1.00	1.00	1.00	93
10	1.00	1.00	1.00	85	60	0.96	1.00	0.98	91
11	1.00	1.00	1.00	84	61	0.94	1.00	0.97	120
12	1.00	1.00	1.00	101	62	0.96	0.94	0.95	126
13	1.00	1.00	1.00	73	63	1.00	1.00	1.00	65
14	1.00	1.00	1.00	82	64	1.00	0.88	0.94	68
15	0.96	1.00	0.98	73	65	0.93	0.99	0.96	145
16	0.75	0.62	0.68	73	66	1.00	1.00	1.00	125
17	1.00	1.00	1.00	116	67	0.00	0.00	0.00	98
18	0.88	0.99	0.93	113	68	0.03	0.04	0.03	112
19	0.95	0.96	0.95	93	69	0.00	0.00	0.00	101
20	1.00	0.98	0.99	93	70	1.00	1.00	1.00	157
21	0.90	1.00	0.95	94	71	1.00	1.00	1.00	163
22	0.99	0.84	0.90	91	72	0.98	0.98	0.98	121
23	0.99	1.00	0.99	98	73	0.00	0.00	0.00	82
24	0.79	0.97	0.87	107	74	0.00	0.00	0.00	149
25	0.73	1.00	0.85	77	75	0.96	0.91	0.93	107
26	0.71	0.88	0.79	94	76	0.48	0.96	0.64	114
27	0.98	0.91	0.94	140	77	0.76	0.91	0.83	78
28	0.23	0.97	0.37	31	78	0.99	0.88	0.93	179
29	1.00	0.98	0.99	123	79	0.71	0.94	0.81	64
30	0.97	0.86	0.92	169	80	1.00	0.97	0.98	131
31	0.86	0.97	0.91	171	81	1.00	0.68	0.81	62
32	1.00	1.00	1.00	151	82	1.00	0.99	0.99	83
33	0.91	0.97	0.94	111	83	1.00	1.00	1.00	77
34	0.74	1.00	0.85	112	84	0.00	0.00	0.00	80
35	0.94	0.99	0.96	134	85	0.12	0.01	0.02	76
36	0.50	0.75	0.60	84	86	1.00	0.73	0.85	49
37	0.95	0.61	0.74	88	87	1.00	0.88	0.93	72
38	0.99	1.00	1.00	102	88	0.91	0.96	0.94	84
39	1.00	1.00	1.00	97	89	1.00	0.41	0.58	139
40	1.00	1.00	1.00	77	90	0.00	0.00	0.00	103
41	0.65	1.00	0.79	72	91	0.00	0.00	0.00	100
42	0.83	0.99	0.90	101	92	1.00	1.00	1.00	152
43	0.89	0.92	0.90	98	93	1.00	1.00	1.00	99
44	0.99	1.00	1.00	130	94	0.98	1.00	0.99	100
45	1.00	0.97	0.98	100	95	1.00	1.00	1.00	92
46	1.00	1.00	1.00	118	96	1.00	0.97	0.99	110
47	1.00	1.00	1.00	101	97	1.00	1.00	1.00	157
48	0.59	1.00	0.74	116	98	0.74	1.00	0.85	87
49	1.00	0.09	0.16	113	99	1.00	1.00	1.00	91
50	0.99	1.00	1.00	100	100	0.95	0.67	0.78	93
					AVG	**0.85**	**0.85**	**0.83**	10002

also critical to demand generation and shopper satisfaction. From the manufacturer's standpoint, assortment selection and category space allocation are equally critical. Manufacturers cannot generate an adequate profit if their products are not on store shelves.

Understanding shopper behaviour is one of the keys to success for retailers. In particular, for an appropriate retail strategy development, it is necessary that managers know which retail attributes are important to which shoppers. The important question, how do people shop, is not well answered by conventional methods so retailers rely on shopper insight to justify making changes to the store.

In the last years, the concept of shopping experience is changed. Nowadays, the stores are not only the place in which consumers go for searching and buying products [268], but they become the place where shoppers spend their time,

Table 3.9: Training/Testing Classification results for TVD, TVH and TVDH descriptors.

	Classifier	Precision	Recall	F1-Score
TVD	KNN	0.35	0.32	0.31
	SVM	**0.48**	**0.43**	**0.42**
	Decision Tree	0.37	0.34	0.33
	Random Forest	0.46	0.43	0.42
TVH	KNN	**0.75**	**0.73**	**0.71**
	SVM	0.70	0.67	0.64
	Decision Tree	0.49	0.46	0.45
	Random Forest	0.71	0.70	0.68
TVDH	KNN	0.81	0.80	0.79
	SVM	**0.85**	**0.85**	**0.83**
	Decision Tree	0.52	0.50	0.48
	Random Forest	0.74	0.71	0.69

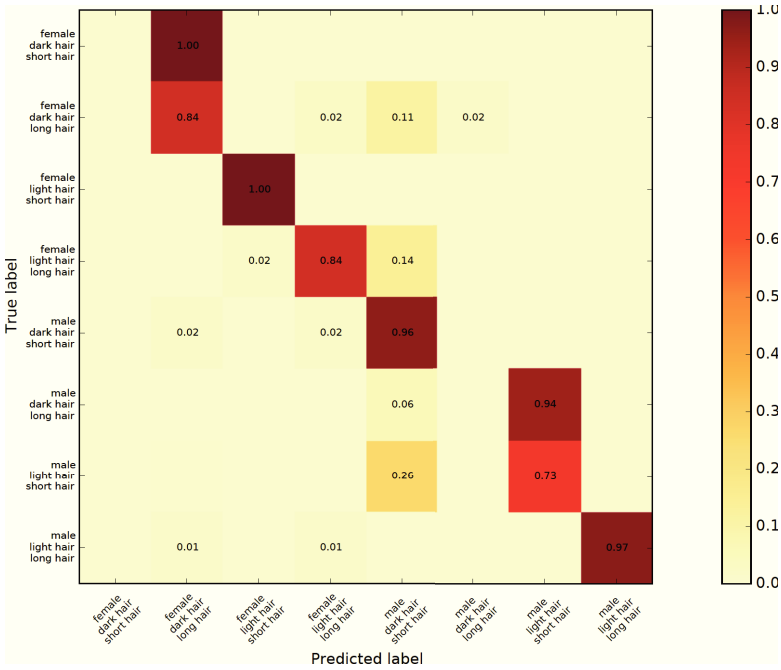


Figure 3.13: Gender Classification Confusion Matrix with kNN classifier.

test products in real time and look for information about the last trends. In a shopping centre, there are many possible activities such as product trying, seeing, touching, professional consulting and ambience shopping, so the consumers are attracted not only by the prices or offers [269].

Currently, artificial intelligence is pervading retail environments and technologies able to supply humans with supplementary knowledge for making bet-

Table 3.10: Gender Classification results with kNN classifier.

Class	Gender	Hair Type	Precision	Recall	F1-S	Sup.
FSD	Female	Short Dark	0.00	0.00	0.00	101
FLD	Female	Long Dark	0.93	0.84	0.88	3036
FSL	Female	Short Light	0.92	1.00	0.96	157
FLL	Female	Long Light	0.76	0.84	0.80	708
MSD	Male	Short Dark	0.89	0.96	0.92	5222
MLD	Male	Long Dark	0.00	0.00	0.00	98
MSL	Male	Short Light	0.82	0.73	0.77	612
MLL	Male	Long Light	1.00	0.97	0.99	68
			0.87	**0.88**	**0.88**	10002

ter decisions have been conceived [160], [161]. The shop is becoming a digital ambient with the artificial intelligence that allows environments to be sensitive and adaptive to the human presence [162], [163].

Generally, computer vision and image processing demonstrated to hold great potential for retail practice and research [164]. Integrated systems able to monitor shoppers in intelligent retail environments have been developed with the aim of learning shopper skills [165], [166]. Data gathered from sensors installed in the shop are used with the aim of evaluating the *attraction* (the level of attraction that the store is creating on consumers), the *attention* (the time consumers spend in front of brand display) and the *action* (the number of consumers that enter in the store and interact with merchandise).

The interest is in improving and personalising shopping experience across digital devices installed in the shopping carts and baskets. For this purpose, sCREEN (Consumer REtail ExperieNce), an intelligent mechatronic system for indoor navigation assistance in retail environments is proposed. It does not require the use of metrics maps but only topological ones. The tracking system is based on Ultra-WideBand (UWB) technology. The system provides the use of several UWB antennas properly positioned inside a predetermined area and powered battery tags free to move inside the area3.14.

It is also able to monitor the movement of consumers in stores and send tracking data to a cloud server [270]. Starting from the analysis of the project, the configuration and the structure of the store up to the position of products on the shelves, by recording the trajectories of consumers in a retail store and by predicting the probability of an analysed subject attraction in front of a shelf, sCREEN helps consumers in navigating retail environments. Experiments are performed in a real retail environment that is a German supermarket, during business hours.

The work contributes the design of an intelligent mechatronic system with the use of Hidden Markov Models (HMMs) to the representation and recognition of shopper attraction and special retail scenarios (shelf out-of-stock or new store layout). Observations are viewed as a perceived intelligent system

Figure 3.14: Example of tag installed in shopping carts.

performance. By forecasting consumers' next shelf/category attraction, the system can present the item location information to the consumer, including a walking route map to a location of the product in the retail store, and/or the number of an aisle in which the product is located. Effective and efficient design processes for mechatronic systems are a prerequisite for competitiveness in an intelligent retail environment.

Traditional automated systems are rigid and are not capable of responding rapidly to changes in demand and supply. Instead, an intelligent mechatronic system is capable of achieving given goals under conditions of uncertainty. Intelligence can be designed into these systems using traditional Artificial Intelligence methods such as expert systems, fuzzy logic or neural networks, but the most cost-effective and powerful implementation is through the use of distributed artificial intelligence, where a community of intelligent agents decides on the optimal or near-optimal action through a process of negotiation [271].

The proposed processing algorithm is inspired at a high conceptual level of abstraction level in other understanding tasks, such as speech and music

recognition or human activity sequence recognition. This work is novel in retail field within a HMM based approach. However, HMMs have been used in other recognition tasks, speech recognition, gesture recognition, human daily activities sequence recognition and robot skill learning [272], [273].

A dataset of consumers' trajectories, with timestamps and the corresponding ground truth (i.e., the attraction shelf) for training as well as for evaluating the HMM has been built and made publicly available.

Furthermore, an approach for clustering the trajectories acquired by shopping carts and baskets is proposed. In fact, the trajectories of shopping carts and baskets in a supermarket are an information-rich feature that reveals the structure of the retail environment, provides clues to the events taking place, and allows inference about the interactions between objects. Firsty, it is assumed that clustering trajectories as a whole could not detect similar trajectories portions, because a trajectory may have a long and complicated path. In fact, even if some trajectories portions show a common behaviour, the whole trajectories might not [274]. Discovering common sub-trajectories could be very useful in this case, where there are regions of special interest for analysis. This framework, in which the solution is to partition a trajectory into a set of line segments and then group similar line segments, is called a "partition-and-group framework". Then, the subtrajectories are stored in clusters and two different algorithm are applied and their performances are compared: Agglomerative [275] and Spectral [276, 277].

Several contributions are made by this work. First of all, in retail environments, such mechatronic systems are capable of achieving considerably better results in terms of performance/cost ratio and reliability than conventional systems and devices. The major elements of the proposed system are battery based intelligent mechatronic devices, which are software objects capable of communicating with each other and localize carts and baskets moving inside a store with high accuracy. Furthermore, this system can model and forecast the shelf/category attraction and it assists customers in store navigation without the use of metrics maps but only topological maps. Traditional automated systems are rigid and are not capable of responding rapidly to changes in demand and supply. Also, this approach enables retailers to compare the impact of different store layouts (maps) or shelf layouts (planograms) on issues such as ease of selection, trading up and the overall shopping experience. Finally, the approach is validated using real data gathered from a real store which helps to make the results more confident and the experiments repeatable; used data are publicly available in an open dataset that is the first dataset based on extensive real data in this field. The innovative aspects are in proposing an adequate HMM structure together with the use of trajectories and time of consumer attraction in front of a shelf/category to forecast that a certain at-

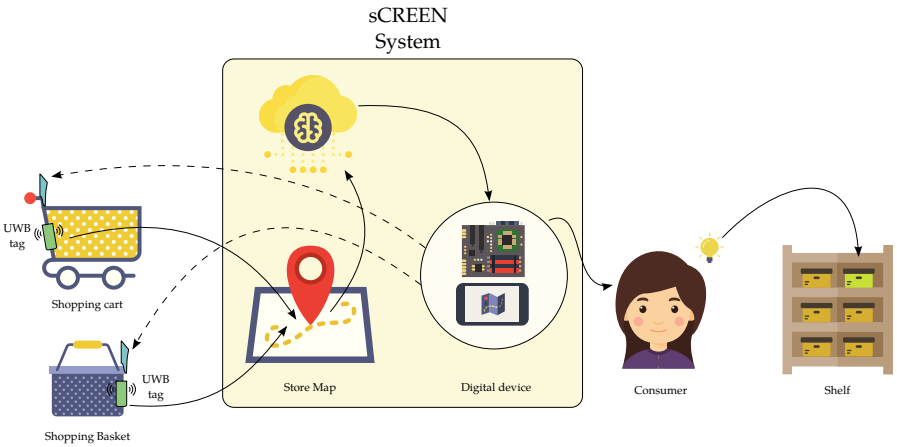


Figure 3.15: The main idea of the sCREEN intelligent retail system.

traction will be performed. The continuous modelling approach and the novel HMM structure is also able to model changes in planogram and store layouts or usual retail events such as products experiencing shelf out-of-stock. Moreover, it is proposed a method for trajectory clustering useful for synthetic data visualization of store flows. A method able to automatically cluster different shopper behaviors useful to identify better store design, secondary placement locations, way-findings, etc. is tested.

3.4.1 Modelling and Forecasting customers navigation in Intelligent Retail Environment.

The sCREEN intelligent mechatronic system provides accurate navigation in retail environments without access to updated metric maps but requires only topological maps. The system leverages the structured movement trajectories of consumers in store to enhance accuracy and minimize active user tagging. sCREEN intelligent system and its forecasting process is depicted in Figure 3.15.

The design and analysis of a mechatronic device have to be easily integrated into the shopping carts and baskets. Reliability of the mechatronic system can generally be expressed as the ability of a system to perform its intended functions under stated conditions for a specified period of time. To cope with the challenges of intelligent mechatronic system design and complexity handling, key elements of the mechatronic system are:

- Mechanic components that are shopping carts and baskets equipped with electronic devices. These allow to track and assist the customers during their shopping experience.

- Embedded devices mounted on the shopping carts and basket to give the information to the shopper. These devices have been developed using a Raspberry Pi 3 Model B, connected to an LCD display.

- HMM-based software that it is able to forecast customer next shelf attraction.

- Automatic component that learn the type of customer. The device retrieves the data predicted by the model from the cloud server, based on the current position of the shopping cart/basket and provides the corresponding route information to the shopper on the display.

During system design and implementation, several issues are faced up and it is designed specific integrated information to enable the provision of retail service. The mechatronic system helps people navigate by providing customized indoor maps that can be used in digital screens. By offering the shelves in sCREEN, it is possible to increase their visibility. Selected shelves can be highlighted at all zoom levels to differentiate and create extra value. It also provides audio tours based on where the individual is in the store. Forecasting shop locations presents obvious benefits for both shoppers and retailers, saving everyone time, improving customers' experiences, streamlining sales and making easier to navigate. Customers can rely on these maps to get turn-by-turn directions inside store, finding specific products category easily.

Structure of Retail Environments

Retail environments such as supermarkets have structured movement trajectories for consumers [278]. The stores generally consist of aisles with products placed in shelves along them. In addition, products may be placed on islands along the broader walkway along the perimeter called the racetrack. The product placement limits the trajectory of costumer movement to approximately unidirectional trajectories within aisles or along the perimeter. Due to the trend in marketing, supermarkets are similar in layout and design. Supermarket is a large self-service grocery store, selling a wide variety of food, groceries, dairy products and household goods. The supermarket commonly comprises of meat, fresh products and baked goods categories, along with shelf space reserved for canned and packaged goods as well as for various non-food items such as household cleaners, pharmacy products and pet supplies. Fresh products are generally located near the entrance of the store. Milk, bread, and other essential staple items are typically situated toward the rear of the store and in other out-of-the-way places, purposely done to maximize the consumer's time spent in the store, strolling past other items and capitalizing on impulse

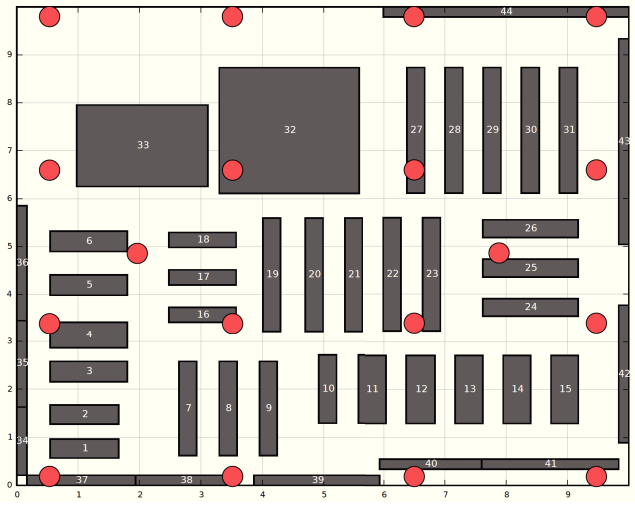


Figure 3.16: Store layout and shelves placement with red dots representing the anchors placed in the dropped ceiling of the store. The total number of anchors used is 18.

buying. The front of the store is the area where, the cash registers are usually located.

Figure 3.16 represents the store layout in which the experiments are performed. It shows the structure and shelves placement, with red dots representing the anchors (Section 3.4.1). It is a supermarket during the business hour. The store layout and product placement (combination of products locating on shelves) with the visual effect of the shopping environment and the sales productivity contribute on shoppers attraction time in front of a shelf.

The sCREEN intelligent system leverages this structure to automatically learn the graph of walkable trajectories in a store from user movement traces.

System Operation

The sCREEN system operation can be described in four steps:

- UWB technology (Section 3.4.1);
- recording of the consumer trajectories, measuring the attraction time in front of a shelf and building the sCREEN dataset (Section 3.4.2);
- problem formulation for the HMM development (Section 3.4.1);
- designing HMM for predicting the shelf attraction and personalising the shopping experience (Section 3.4.1).

This section gives an overview of the approach.

UWB Technology

The sCREEN intelligent system is based on UWB technology, which is able to monitor the consumers trajectories in stores and send the collected tracking data to a cloud server. These data, properly processed and stored in a database, are useful to obtain information about consumers' behaviour during the shopping experience. The technology has been used to deploy a RTLS (Real Time Locating System), in order to collect real time localization data from shopping carts and baskets. Figure 3.16 shows the system deployment. Tests performed in the store achieved an accuracy in the position measurement of 20 cm in terms of indoor localization and owing to a smart power management provides a high autonomy for the battery-powered tags. This value is not achievable with conventional wireless applications (such as RFID, WLAN and others), and so it is very useful for applications that require a high levels of precision in real time for 2-D and 3-D localization. The UWB radio module DWM1000 advent (IEEE802.15.4-2011 UWB compliant) by decaWAVE, at a low price (10$ per 1000 unit), has given a great number of tracking solutions, developed by European companies. The costumers' trajectories collected, with other information, such as average walking and attraction time in front of a shelf/category are useful information for the retailers with the aim of improving the shopping experience of their customers. It is possible to measure:

- Time of arrival;
- Direction of arrival;
- Signal strength.

These data have a great importance, because they represent how the consumer interacts with the environment, which areas are the most visited, the difficulties in finding a product and what are the most effective strategies in terms of marketing, communication and space design [270]. UWB signals are transmitted with a shorter duration, consuming less power and can operate in a wide area of the radio spectrum. UWB and RFID can operate in the same area without interference thanks to the differences in signal types and radio spectrum. Moreover, UWB signal is able to pass through walls, devices and clothes with no interferences.

The tracking system adopted is composed by:

- **Anchors**: static devices (antennas) with a known position, placed in the dropped ceiling of the store, to form a homogeneous grid in order to cover the entire store. The anchors gather signals from the tags and forward this data (timestamps of received signals and tags related information such as ID and battery level) to the RTLS server. The anchors are connected and

powered via Ethernet through a PoE switch to the RTLS server. Since all the anchors have to share the same time, one of them is chosen as the master anchor for synchronizing them.

- **Tags**: mobile devices to be tracked. They send data to the anchors at a specified transmission rate and they also send a broadcast message that is received by the anchors in a communication system (this makes their number fully scalable as they do not have to directly negotiate a transmission). The tag is also characterised by an accelerometer for movement detection in order to prolong battery lifetime. In this way, new data are sent only if the tags exceed a certain and adjustable threshold.
- **RTLS Server**: collect data from the anchors, estimate the 3D position of tags and send these to the cloud server. In order to estimate a tag position through multilateration, a TDOA (Time Difference of Arrival) algorithm that takes into account only timestamps coming from at least three anchors with the same time is used.

The RTLS Server sends localization data on a TCP/IP socket. A software is developed to collect from the stream the following information:

- master ID;
- tag ID;
- position coordinates (x, y and z);
- battery level of the tag and timing information (such as tracking system and RTLS server timestamp).

Design of HMM structure: problem formulation

Let:

$$X = \{x_1, x_2, \ldots, x_n\}$$

be a discrete finite shelf attraction space and

$$O = \{o_1, o_2, \ldots, o_m\}$$

the observation space of a HMM [279]. Let T be the transition matrix of this HMM, with $T_{x,y}$ representing the probability of transitioning from attraction in shelf $x \in X$ to attraction in shelf $y \in X$, and $p_x(o)$ be the emission probability of observation $o \in O$ in attraction in shelf $x \in X$.

The probability that HMM trajectory follows the attraction sequence s given the sequence of n observations, is denoted as:

$$P(X_{1:n} \in seq_n(s)|o_{1:n})$$

where $seq_n(s)$ is a set of all length n trajectories whose duration free sequence equals to s.

Finding the most probable attraction sequence can be seen as a search problem that requires evaluation of probabilities of attraction sequences. The Viterbi algorithm [280] based on dynamic programming can be used to efficiently find the most probable trajectory. In fact, it makes use of the Markov property of an HMM (that the next state transition and symbol emission depend only upon the current state) to determine, in linear time with respect to the length of the emission sequence, the most likely path through the states of a model which might have generated a given sequence.

Forecasting customers attraction with HMMs

In this section, a trajectory model is derived by reformulating the standard HMM whose state output vector includes both static and dynamic feature vectors. To evaluate the approach usefulness for attraction prediction, a new dataset is built. The sCREEN dataset contains consumers trajectories. The data were collected over a period of 30 days. By imposing the explicit relationships between the static and dynamic features, the standard HMM is naturally translated into a trajectory model, referred to as trajectory HMM in the present paper. The trajectory HMM can overcome the above two limitations of the standard HMM without any additional parameters. A Viterbi-type training algorithm based on the maximum likelihood criterion is also derived. Information provided by shopping carts and baskets detection algorithms can be used as input for a set of HMMs. Each of these recognise different trajectories.

After training the model, it is considered a trajectory

$$s = \{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$$

and calculate its probability λ for the observation sequence $P(s|\lambda)$. Then, the trajectory is classified as the one which has the largest posterior probability. To each observation the ground truth (state) is associated that is the shelf from which the consumer was attracted. The observation are the grid cells in which the store map is divided.

Figure 3.17 depicts the general scheme of the attraction prediction process. In particular, different HMMs are used, which have as observations trajectory of:

- the shopping baskets (HMM^S_{SB});
- both shopping baskets and carts together (HMM^S_{SBSC});

- shopping carts and baskets during the slot 06.00 a.m. - 11.00 a.m ($HMM^S_{06am-11am}$);
- shopping carts and baskets during the slot 11.00 a.m. - 04.00 p.m ($HMM^S_{11am-04pm}$);
- shopping carts and baskets during the slot 04.00 p.m. - 09.00 p.m ($HMM^S_{04pm-09pm}$).

The different observations clusters are described to test different scenario and different targets and prove the effectiveness of the proposed approach. In particular shoppers using carts or basket are usually following very different flows inside the store visiting less than 5 category (shoppers with baskets) or more than 10 (shoppers with carts). The same different behaviours can be found for shoppers visiting the store at different day time.

Table 3.11 indicates the number of vertical and horizontal layers used in the quantization step for each HMM and the total number of observations.

Finally, the classification module provides the shelf x_j^S or the category x_j^c that maximizes $P^S_{HMM_i}$ or $P^C_{HMM_i}$ respectively. It is the HMM trajectory probability that follows the trajectory s given the sequence of n observations, i.e.:

$$x_j^\alpha = \arg\max_i P_{HMM_i^\alpha}(X_{1:n} \in seq_n(s)|o_{1:n}) \quad \alpha \in \{S, C\} \tag{3.4}$$

In the retail field, the right space allocation for products and categories plays a critical role. For this reason, Equation 3.4 can be redefined as follows:

$$x_j^\alpha = \arg\max_i P_{HMM_i^\alpha}(X_{1:n} \in seq_n(s)|o_{1:n}) \cdot k_i \cdot f_i(t) \quad \alpha \in \{S, C\} \tag{3.5}$$

where $k_i \in [0, 1]$ is the retailer coefficient that considers the out-of-stock of some products on the shelves. Low values of k_i indicate an out-of-stock or a market penalty of the shelf. Instead, $f_i(t) \in [0, 1]$ is a function (figure 3.18a) that models design changes of store over time. It is defined as:

$$f_i(t) = \frac{\tanh(\alpha t - 1 - t_0 - t^*) + \tanh(t_f - t^* + 1 - \alpha t)}{2} \qquad \in [0, 1]$$

where t_0 and t_f are respectively first and last time of store particular arrangement; t^* is the time required by the model to learn the consumers' behaviour; $\alpha \in [0, 1]$ takes into account the learning speed of model. If the product is not on the shelf, it is not necessary that the customer approaches the shelf. This is a very important value for promoting brands with remarkable promo products. When the retailer wants to increase a promotional brand or product he/she has to descrease this value. In this way, the store can create personal profiles and thus improve the success of offers and promotions tailored to individual con-

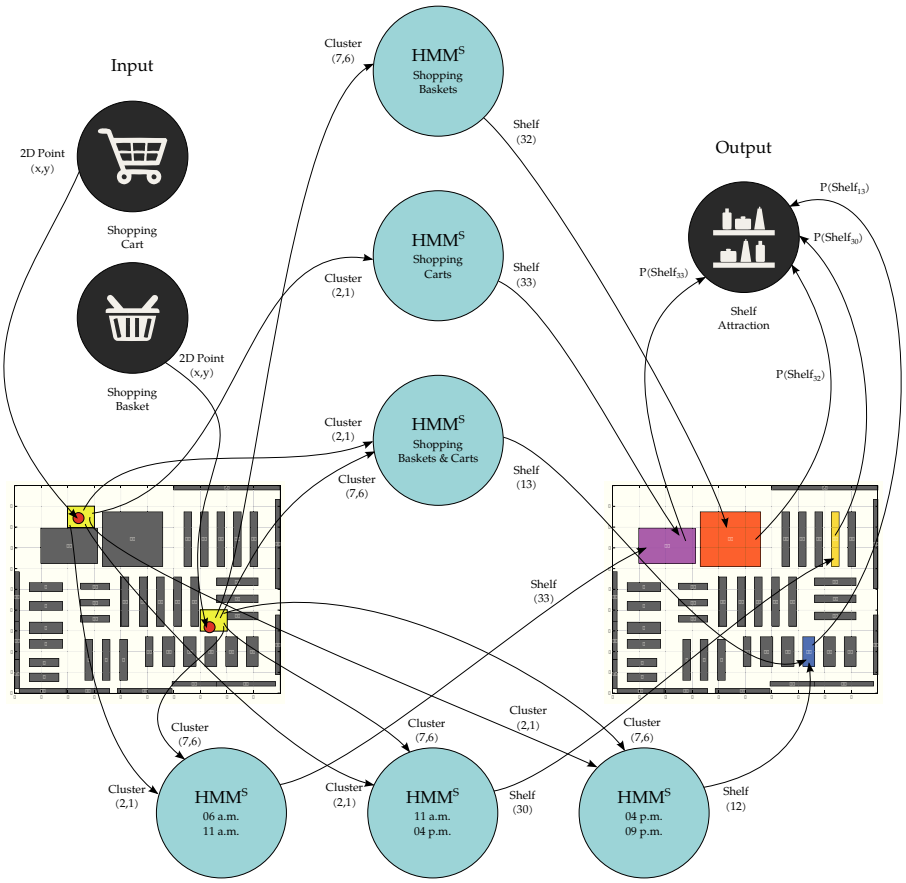
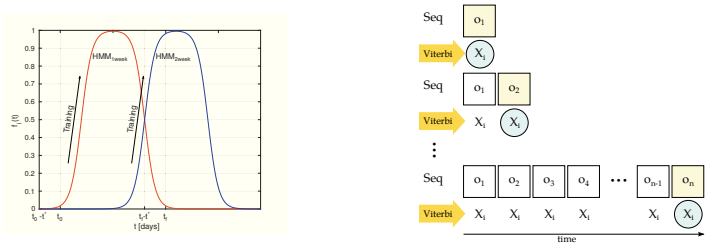


Figure 3.17: General schema of architecture. For each HMMs are shown the inputs and the outputs.

sumers, and respond to individual profiles of persons and objects with whom it has no previous relationship.

The overlying HMM is valid every time as long as the store arrangement remains unchanged.

HMMs process for attraction forecasting is illustrated in Figure 3.18b. The system automatically learns customers target. It builds up a prediction system consisting of HMMs that fit and model the customer behaviour in store. When customer is attracted by other products the system automatically updates model over time. In fact, this model adapts rapidly to any changes to customer attraction.



(a) (b)

Figure 3.18: Store arrangement function and forecasting with HMMs process representation.

Table 3.11: Number of observations for each HMMs (v: vertical layer, h: horizontal layer).

Trajectories	**# layers**	**# observ.**	**# states**	**type**
Shopping Baskets	$v:10,\ h:10$	100	44	shelf
Shopping Carts	$v:10,\ h:10$	100	44	shelf
Shopping Baskets & Shopping Carts	$v:10,\ h:10$; $v:10,\ h:10$	100	44	shelf
Slot 06.00 am - 11.00 am	$v:10,\ h:10$; $v:10,\ h:10$	100	44	shelf
Slot 11.00 am - 04.00 pm	$v:10,\ h:10$; $v:10,\ h:10$	100	44	shelf
Slot 04.00 pm - 09.00 pm	$v:10,\ h:10$; $v:10,\ h:10$	100	44	shelf
Shopping Baskets & Shopping Carts	$v:10,\ h:10$; $v:10,\ h:10$	100	10	category

Trajectory Clustering

In this section, a design overview is presented. It is assumed that each carts/basket trajectory in a supermarket begins in the same area, where they will be placed after the shopping. This area, called "origin area" or "origin", is the spatial delimiter of the trajectories. Some other "control areas" (or, simply, "controls") are also defined. Figure 3.34 depicts the store layout and the shelves placement. The highlighted areas (c1, c3, c7) are the "origin area" and the "control areas". "Controls" have been placed where most of the trajectories pass through these areas, therefore considered strategic areas, because there are generally high densities of carts/baskets.

Section 3.4.1 formally presents the problem statement. Section 3.4.1 describes the filtering operations applied. Section 3.4.1 and Section 3.4.1 presents the two clustering methods that are used for the comparison. In the following exposition, d_{ij} is the distance between the ith and jth input trajectories (used by the agglomerative method), and k_{ij} represents their corresponding similarity after appropriate scale selection (used in the spectral method). Finally,

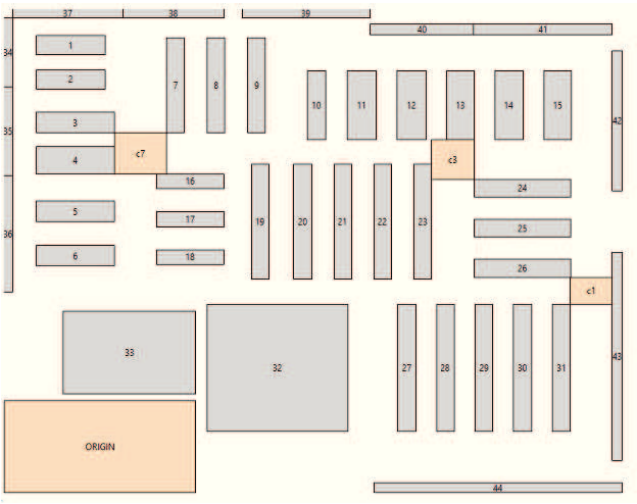


Figure 3.19: Store layout and shelves placement. The highlighted areas (c1, c3, c7) are the "origin area" and the "control areas"

Section 3.4.1 describes the innovative approach to cluster trajectory data.

Problem Statement

A clustering algorithm based on the partition-and-group framework [274] is developed. Given a set of trajectories $I = \{T_1, T_2, \ldots, T_n\}$; the algorithm generates a set of clusters $O = \{C_1, C_2, \ldots, C_n\}$ as well as a representative trajectory for each cluster C_i, where the trajectory, cluster, and representative trajectory are defined as follows. A trajectory is a sequence of multi-dimensional points. It is denoted as $T_i = \{p_1, p_2, \ldots, p_j, \ldots, p_k\}$, $(1 \leq i \leq n)$. Here p_j, $(1 \leq j \leq k)$ is a d-dimensional point. The length k of a trajectory can be different from those of other trajectories. A trajectory $p_{d1}, p_{d2}, \ldots, p_{dt}$, $(1 \leq d_1 < d_2 < d_t \leq k)$ is called a sub-trajectory of T_i.

A cluster is a set of trajectory partitions, i.e., a line segment $p_i p_j$ (i ¡ j), where p_i and p_j are the points chosen from the same trajectory. For the distance measure, line segments belonging to the same cluster are close to each other. It is important to highlight that a trajectory can belong to multiple clusters. In fact, a trajectory is partitioned into multiple line segments, and clustering is performed over these line segments. A representative trajectory (that indicates a common sub-trajectory) is a sequence of points such as an ordinary trajectory. It is an imaginary trajectory that represents the behaviour of the trajectory partitions (i.e., line segments) that belong to the cluster.

Filtering

Before clustering, three filtering operations are applied: one during the trajectory identifying process ("positions out-of-bounds") and two after the identification ("densities of points" and "Kalman filter").

- *Positions out-of-bounds*: some cart positions may be placed outside the bounds of the map (positions with negative x and/or y coordinates). Carts having positions out-of-bounds are skipped in the trajectory identifying process.

- *Densities of points*: some trajectories may be characterized by segments with an high density of sparse points within a small area. This fact is probably due to the inaccuracy of the RTLS or to the chaotic behaviour of the customers. These segments are smoothed removing the unnecessary points densities.

- *Kalman filter*: a standard Kalman filter [281] is then applied to each trajectory to smooth them, by removing the noise and the RTLS inaccuracies.

Agglomerative Clustering

In the agglomerative clustering, each trajectory is initially assigned to a single cluster, as done in [275]. The two closest clusters are iteratively merged until all clusters are separated by a distance greater than a prespecified threshold μ. If P and Q denote the sets of trajectories indices of two disjoint clusters, the distance between the clusters is

$$d_{avg} = \frac{1}{|P||Q|} \sum_{i \in P} \sum_{j \in Q} \tag{3.6}$$

In the experiments, as the clusters number is known from the manual ground-truth data, the desired clusters number has been specified rather than choosing a value of μ. The range of values for μ that produce the same number of clusters is indicative of the sensitivity of the underlying distance measure.

Spectral Clustering

As mentioned above, it is used the clustering algorithm proposed in [282]. The clusters number is choosen automatically using the distortion metric to select the optimal globalscale [276]. The input to the spectral clustering method is the affinity matrix K with elements k_{ij} for $1 \leq i,\ j \leq n$. At first, the process requires to normalize the rows and columns of K, which yields a normalized

affinity L, as

$$L = W^{-1/2} K W^{-1/2} \tag{3.7}$$

where the ith element of the diagonal matrix W is defined as

$$w_{ij} = \sum_{1 \leq j \leq n} k_{ij} \tag{3.8}$$

The L matrix is a block diagonal and it has g (current clusters number in the data) eigenvalues equal to 1 and $n - g$ eigenvalues equal to zero. Counting the eigenvalues number with value 1 is not useful for understanding the clusters number [277]. For this reason, a range for the clusters number is found using the L spectrum. Assuming that $1 \geq \lambda_1 \geq \lambda_2 \ldots \geq \lambda_n \geq 0$ are the L eigenvalues, g_{min} (minimum clusters number in the data) can be evaluate counting the eigenvalues grater that 0.99, instead g_{max} can be estimate counting the eigenvalues greater than 0.8. These thresholds were experimentally chosen. For some experiments, the possible range is very narrow, whereas it was large enough for others. After choosing the search range for the clusters number, the L eigenvectors are computed that correspond to the top g_{max} eigenvalues. Assuming v_i represents these vectors for $1 \leq i \leq g_{max}$. In case repeated eigenvalue, it is required that the corresponding eigenvectors are mutually orthogonal. The clustering algorithm proceeds using the following steps, which are repeated for each integer value of g, inclusively ranging from g_{min} to g_{max}.

- For all integer g values between g_{min} and g_{max}, the $n \times g$ $V = [v_1, \ldots, v_q]$ is formed.

- Each row of V is normalized to have a unit length $R = SV$, where S is diagonal with elements $s_i = (\sum_{j=1}^{c} V_{ij})^{-1/2}$

- The rows of R have to be considered as g-dimensional data points and clustered using k-means. Let $\mu_1, \mu_2, \ldots, \mu_q$ represent the g cluster centers (as row vectors), and let $c(i)$ represent the cluster that corresponds to the ith row $r_{(i)}$.

- The within-class scatter is $W = \sum_{i=1}^{n} ||r_{(i)} - \mu_{c_{(i)}}||_2^2$ and the total scatter is $T = \sum_{i=1}^{n} \sum_{j=1}^{k} ||r_{(i)} - \mu_j||_2^2$. The distortion score $p_g = \frac{W}{(T-W)}$ is computed as the ratio of the within-class to between-class scatter. The g value that produces the least distortion p_g is the number of clusters automatically determined, and $c(i)$, which is obtained with the clusters number, is the class indicator function for the input trajectories.

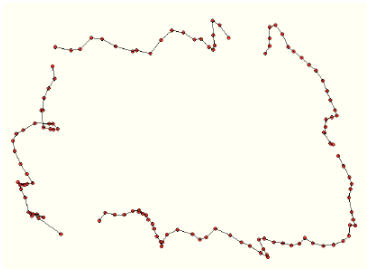


Figure 3.20: A trajectory splitted in four sub-trajectories.

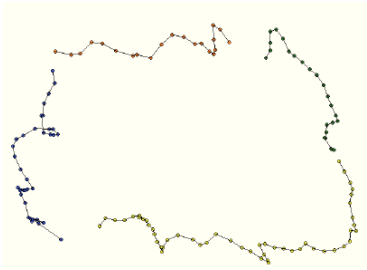


Figure 3.21: After the clustering, each sub-trajectory is merged to form a track. The ordered set of clusters (yellow→ green → orange → blue) is the macro-cluster to which the track belongs.

Macro Clustering

The innovative approach mainly relies on the concepts of macro-cluster. Since a whole trajectory can be conceived as an ordered set of split sub-trajectories (Figure 3.20), after the individual (agglomerative or spectral) clustering process each sub-trajectory is assigned to a certain cluster, identified by a color (Figure 3.21). The single sub-trajectories are then merged to form a track, which is approximately equal to the original whole trajectory. The ordered set of clusters, to which they belong, can be seen as a new bigger cluster, called "macro-cluster".

As a sub-trajectory belongs to a cluster, a track belongs to a macro-cluster. The macro-clustering process is implemented to find the most common paths in the store target by obtaining the most frequent patterns of tracks. Furthermore, as a cluster can be identified by all the subtrajectories which belong to it, a macro-cluster gathers all the tracks characterized by the same ordered set of clusters (Figure 3.22 and Figure 3.23).

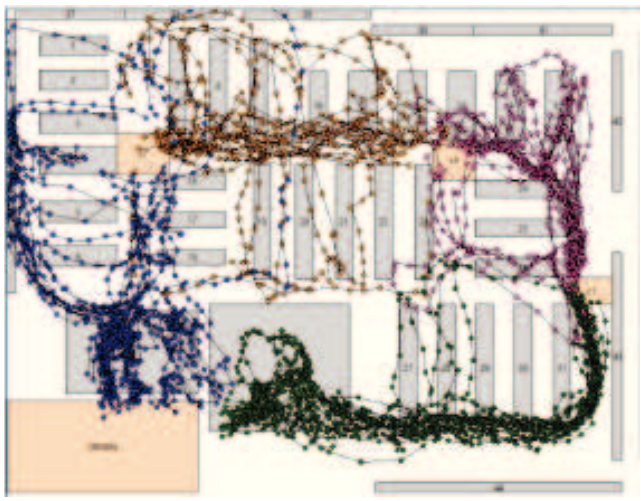


Figure 3.22: Example of all the tracks belonging to the macro-cluster (yellow→ green → orange → blue).

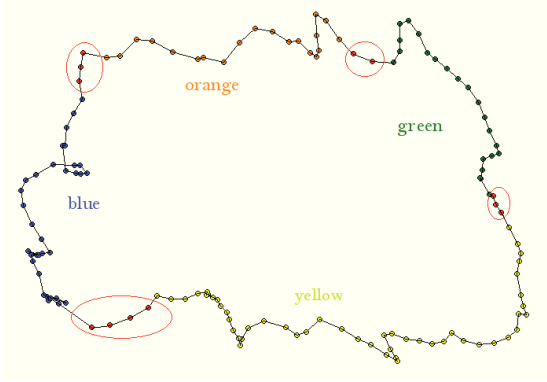


Figure 3.23: Example of track: Composed by four sub-trajectories (yellow → green → orange → blue), it belongs to the macro-cluster yellow-green-orange-blue (red dots inside red ellipses has been inserted to simulate and highlight the complete trajectory).

3.4.2 sCREEN Dataset

The store target in which sCREEN dataset[6] was collected is a German supermarket, during the business hours. This dataset is composed by the information gathered through the sensors installed in the shopping carts and baskets in the analysed supermarket. Location data generated from sensor equipped moving shopping carts and baskets are typically collected as streams of spatio-temporal (x, y, t) points that when put together form corresponding

[6] `http://vrai.dii.univpm.it/screen-dataset`

trajectories. As a prior step, the points are filtered with an attraction time in front of a shelf/category less than 5 seconds and the trajectories made in less than 2 minutes because too short and not suitable to this analysis. Trajectory construction is evidently necessary for mobility data processing and understanding, including tasks like trajectory data cleaning, compression, and segmentation so as to identify trajectory episodes like stops (e.g. while observing a shelf) and moves. However, for each consumer trajectory construction it is considered that when the cart or basket is stopped for 5 or more minutes it is taken by another consumer so new trajectory is obtained. The data were collected over a period of 30 days.

3.4.3 Performance Evaluation and Results

HMMs Results

An architecture to implement HMMs shopper attraction prediction is proposed. The architecture uses the shopping carts and baskets trajectories to learn the probability of shelf/category attraction. The algorithm for HMM training is the forward-backward, or Baum-Welch algorithm [283]. This is an algorithm that uses an iterative expectation/maximization process to find an HMM which is a local maximum in its likelihood to have generated a set of training' observation sequences. This step is needed because the state paths are hidden, and the equations cannot be solved analytically.

In this study, the Baum–Welch algorithm was employed to estimate a transition probability matrix and an observation emission matrix so that the model best fits the training dataset. Since the discrete observation density is used in implementing HMMs, clustering step is required to map the continuous observation in order to convert continuous data to discrete data.

To evaluate the performance of the algorithms the following metrics are used [284]:

- *Precision*: is a function of true positives and examples misclassified as positives (false positives).
- *Recall*: is a function of correctly classified examples (true positives) and misclassified examples (false negatives).
- *F1-score*: is the usual measure that combines precision and recall through the harmonic mean as a single measure of a test's accuracy.

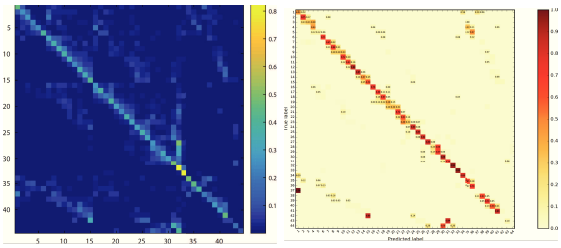
The information about actual and predicted classifications done by the system is depicted by confusion matrix [236] reported for every test case using both transition matrix and a normalized confusion matrix.

Below, the results were obtained performing different experiments with different shoppers clusters to prove the ability of the proposed approach to model with high accuracy different behaviours. The main goal is to gradually forecast the next shelf or category attraction. In both cases, it is randomly splitted the dataset into training and test sets. In particular, it is selected 66% of the instances for training and use the remaining 34% as a test dataset.

At first, the consumers behaviour analysing their trajectory is predicted when they use in the target store the cart or the basket. Three different HMMs are modelled: for the shopping baskets only (HMM^S_{SB}), then the carts (HMM^S_{SC}) and finally, the combination of both (HMM^S_{SBSC}).

The results for each HMM are shown in Figures 3.24, 3.25, 3.26. As It can be see in the confusion matrices most of the trajectories are detected with high accuracy, proving the effectiveness and suitability of this model. Final overall accuracies are then evaluated and compared.

Figure 3.24 represents the HMMs results for shopping baskets (HMM^S_{SB}). In particular, the first one is the transition matrix for HMM^S_{SB} and the matrix on the right is the normalised confusion matrix for HMM^S_{SB}. Transition matrix allows to determine the shelves with higher initial probability, that are: $\{33, 32, 34, 31, 41, 15\}$. Their position in the store is represented in Figure 3.16.



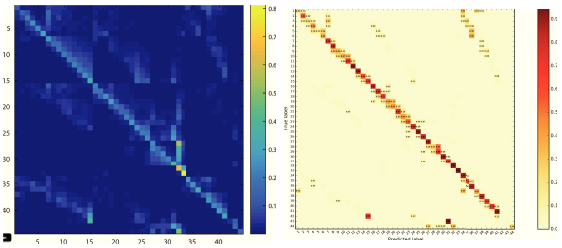
(a) Transition matrix. (b) Confusion matrix.

Figure 3.24: HMMs Results for shopping baskets (HMM^S_{SB}) showing high precision in the confusion matrix. Results are comparable with other target groups presented in the following figures. The test case accuracy has an overall precision of 0.79.

Figure 3.25 depicts the HMMs results for shopping carts (HMM^S_{SC}), with the figure on the right that is the normalised confusion matrix for HMM^S_{SC}. In this case, the shelves with higher initial probability are: $\{33, 32, 31, 15, 36, 41\}$.

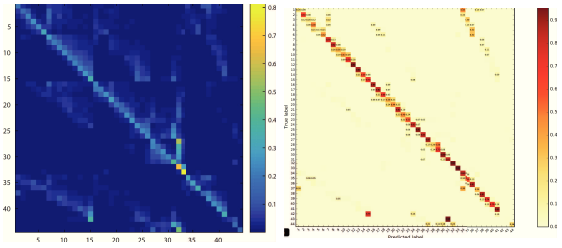
In Figure 3.26 are reported the HMMs results for both shopping baskets and carts (HMM^S_{SBSC}) that are respectively the transition matrix and the normalised confusion matrix for HMM^S_{SBSC}. The shelves with higher initial probability are: $\{33, 32, 31, 15, 41, 36\}$.

It has been tried to forecast consumers' behaviour according to the time slot in which they make purchases in the store. For this reason, three different slot



(a) Transition matrix. (b) Confusion matrix.

Figure 3.25: HMMs Results for shopping carts (HMM^{S}_{SC}). Results are comparable with other target groups presented in the previous and following figures. The test case accuracy has an overall precision of 0.80.



(a) Transition matrix. (b) Confusion matrix.

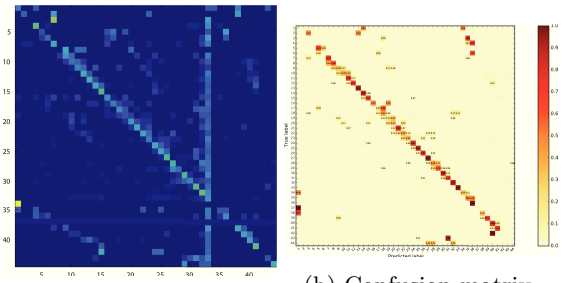
Figure 3.26: HMMs Results for both shopping baskets and carts together (HMM^{S}_{SBSC}). Results are comparable with other target groups presented in the previous and following figures. The test case accuracy has an overall precision of 0.80.

during the day are analysed: morning (06.00 a.m. - 11.00 a.m), middle of the day (11.00 a.m. - 04.00 p.m.) and evening (04.00 p.m. - 09.00 p.m.).

The attraction probability in front of a shelf change during the three analysed daily slot. This fact is highlited by the results obtained. In particular, in the morning slot, 06.00 -11.00 a.m, the shelves with a higher attraction probability are: $\{32, 41, 4, 31, 15, 29\}$. Instead, in the middle of the day slot, 11.00 a.m - 04.00 p.m, the shelves that are more attractive are: $\{33, 32, 31, 15, 41, 40\}$. As well as, the last slot taken into account (04.00 - 09.00 p.m) has the following shelves that capture the customers attention: $\{33, 32, 31, 36, 10, 6\}$.

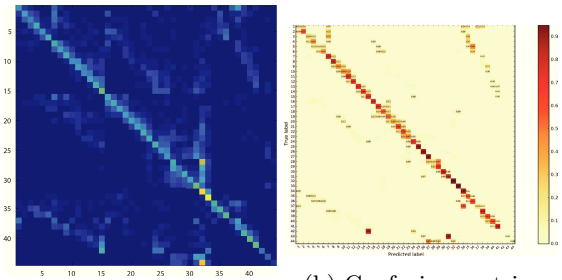
Figure 3.27 depicts the HMMs results for shopping carts and baskets during the slot 06.00 -11.00 a.m ($HMM^{S}_{06am-11am}$), with the transition matrix for $HMM^{S}_{06am-11am}$ and the normalised confusion matrix for $HMM^{S}_{06am-11am}$.

In Figure 3.28 are represented the HMMs results for shopping carts and baskets during the slot 11.00 a.m - 04.00 p.m ($HMM^{S}_{11am-04pm}$). There is the transition matrix for $HMM_{11am-04.00pm}$ and on the right there is the



(a) Transition matrix. (b) Confusion matrix.

Figure 3.27: HMMs results during the slot 06.00 -11.00 a.m ($HMM^{S}_{06am-11am}$). Results are comparable with other target groups presented in the previous and following figures. The test case accuracy has an overall precision of 0.76.



(a) Transition matrix. (b) Confusion matrix.

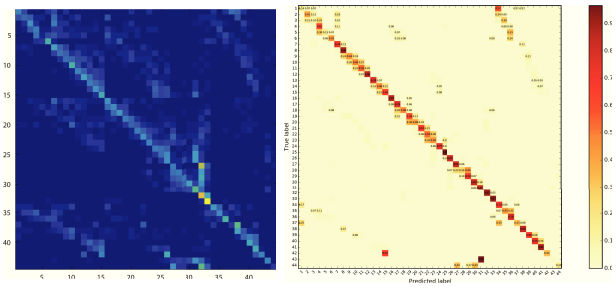
Figure 3.28: HMMs results during the slot 06.00 -11.00 a.m ($HMM^{S}_{06am-11am}$). Results are comparable with other target groups presented in the previous and following figures. The test case accuracy has an overall precision of 0.76.

normalised confusion matrix for $HMM^{S}_{11am-04pm}$. For the middle day slot, the shelves with higher initial probability are: $\{32, 41, 4, 31, 15, 29\}$.

Finally, in Figure 3.29 are reported the HMMs results for shopping carts and baskets during the slot 04.00 - 09.00 p.m ($HMM^{S}_{04pm-09pm}$), with respectively the transition matrix for $HMM^{S}_{04pm-09pm}$ and the normalised confusion matrix for the evening slot time $HMM^{S}_{04pm-09pm}$. For this slot, the shelves with higher initial probability are: $\{33, 32, 31, 15, 41, 40\}$, $\{33, 32, 31, 36, 10, 6\}$.

Table 3.12 summarises the results in terms of Precision, Recall and F1-score demonstrating the effectiveness and suitability of the approach, with a precision value of about 80%, proving that the proposed approach can model very different behaviours with a high accuracy.

As it can be inferred from the results in Table 3.12, the shopping carts and baskets together (HMM^{S}_{SBSC}) achieve greater performance with a precision



(a) Transition matrix. (b) Confusion matrix.

Figure 3.29: HMMs results during the slot 04.00 - 09.00 p.m ($HMM^{S}_{04pm-09pm}$). Results are comparable with other target groups presented in the previous and following figures. The test case accuracy has an overall precision of 0.78.

Table 3.12: Classification Results

	Precision	**Recall**	**F1-score**
HMM^{S}_{SB}	0.79	0.78	0.78
HMM^{S}_{SC}	0.80	0.80	0.79
HMM^{S}_{SBSC}	0.80	0.81	0.80
$HMM^{S}_{06am-11am}$	0.76	0.76	0.75
$HMM^{S}_{11am-04pm}$	0.77	0.77	0.77
$HMM^{S}_{04pm-09pm}$	0.78	0.78	0.78

value of 80%. It can be also assumed that the shopping baskets (HMM^{S}_{SB}) outperform the shopping carts (HMM^{S}_{SC}). This is due to the fact that the consumer chooses a shopping cart or basket that fits the kind of shopping to do. In fact, observing the consumers shopping habits, it is possible realized that the shopping carts could provide better services and sell more groceries if only consumers need more merchandise. The shopping baskets are commonly chosen by single people that spend less time in a supermarket and have few items to take. Instead, a cart must be available to hold a worth week of groceries or a list of items. For this reason, the carts are generally used by consumers that have much items to take and spent much time inside the store and in front of a shelf (they are usually used by family and elderly). It could be noted that for the time slots the best value of precision (78%) is for the slot 04.00 p.m. - 09.00 p.m. Probably, it is due to the fact that the customers mostly preferred purchasing time slot was evening, as majority were salaried people shopping after their office hours. The shopping is concentrated in the evening and in middle day. Instead, the precision value of $HMM^{S}_{06am-11am}$ lower than the other slots is probably due to a limited number of shoppers in the morning that are commonly elderly.

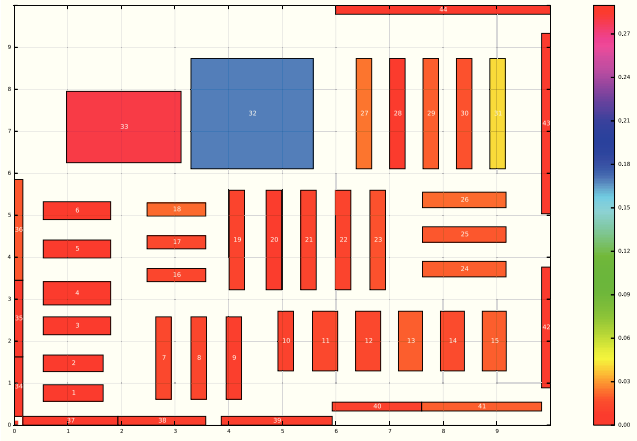


Figure 3.30: Heatmap of shelves attractiveness in the target store using test set and HMM^{S}_{SBSC}.

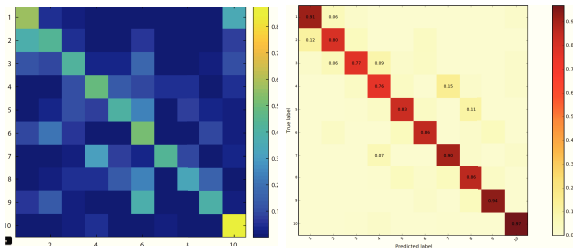
Figure 3.30 is the heatmap probability of shelves attractiveness in the target store using test set and HMM^{S}_{SBSC}, defined as,

$$P(S_i) = \frac{n_{a_i}}{n_i} \tag{3.9}$$

where n_i is the total number of observations and n_{a_i} is the number of correct predictions. In this Figure, it is possible to notice the forecast of customers attraction time in front of the shelves of the store analysed. The heatmap enables retailers to compare the impact of different shelf layouts (planograms) on issues such as ease of selection, trading up and the overall shopping experience. It could be a tool for planogram optimization solutions that help retailers make the right decisions on floor space allocation and merchandising. By observing the shelves more attractive, retailers can customize their store assortments to consumer demand and composition. The heatmap can be highly effective for positioning a retailer in the market and attracting and retaining customers. It also allows to drive sales for the product category more attractiveness, and it is also useful to evaluate the less attractiveness for the other shelves.

The last experiment is performed with the aim to forecast the next category attraction. The store tagert is divided in ten categories and Table 3.13 gives information about the division of shelves into these categories with their attraction probability. Even in this case, the dataset is randomly splitted into training and test sets. In particular, it is selected 66% of the instances for training and use the remaining 34% as a test dataset.

Table 3.14 summarises the results in terms of Precision, Recall and F1-score for each category analysed. It is possible to infer from these results, that high



(a) Transition matrix. (b) Confusion matrix.

Figure 3.31: HMMs Results for both shopping baskets and carts together (HMM^{C}_{SBSC}). The test case accuracy has an overall precision of 0.91.

Table 3.13: Categories division with their attraction probability.

Category	**Shelves**	**Probability (P)**
1	27, 28, 29, 30, 31, 44	0.12
2	26, 25, 24, 42, 43	0.08
3	19, 20, 21, 22, 23	0.06
4	18, 17, 16, 6, 5, 4, 3, 2, 1	0.07
5	7, 8, 9	0.04
6	10, 11, 12, 13, 14, 15	0.11
7	36, 35, 34	0.04
8	37, 38, 39	0.02
9	40, 41	0.04
10	33, 32	0.42

values of precision and recall can be achieved, especially for categories 6 and 10. The recognition of categories 7 and 8 is more difficult due to the smaller amount of available training data and the higher variation in motives. This fact is also reflected by the characteristics of the dataset. However, the average values of the classification performance are very high, thus demonstrating the effectiveness and the suitability of the proposed approach.

Figure 3.32 is the heatmap probability of categories attractiveness in the target store using test set and HMM^{C}_{SBSC}, previously defined in equation 3.9. In this Figure, it is possible to notice the forecast of customers attraction time in front of the category of the store analysed.

Trajectory clustering Results

In addition to the performance of the agglomerative and the spectral clustering, the performance of the macro-clustering, which is the innovative aspect of this work, are presented. The experiments are based only on a partition of the dataset. 390 trajectories have been computed and clustered following the steps previously mentioned.

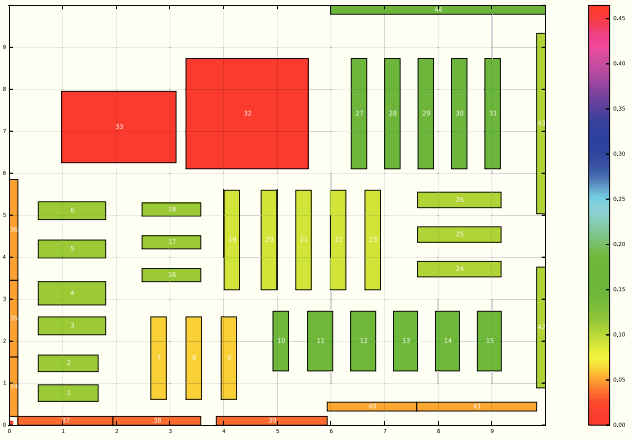


Figure 3.32: Heatmap of categories attractiveness in the target store using test set and HMM^{C}_{SBSC}.

Table 3.14: Classification Results for each category of the target store

Category	Precision	Recall	F1-Score	Support
1	0.91	0.91	0.91	2442
2	0.80	0.80	0.80	1459
3	0.84	0.77	0.80	1357
4	0.76	0.76	0.76	1567
5	0.78	0.83	0.80	809
6	0.93	0.86	0.89	2187
7	0.55	0.90	0.69	427
8	0.65	0.86	0.74	370
9	0.83	0.94	0.88	627
10	0.99	0.97	0.98	10223
avg / total	0.91	0.90	0.90	21468

For the agglomerative clustering the trajectories are grouped into 8 different clusters, while for the spectral clustering the number is automatically computed at run-time.

However, experiments demonstrate that this number is generally lower than using agglomerative clustering. This means that many anomalous trajectories, which should be excluded after the agglomerative (yellow, gray, purple and cyan trajectories in Table 3.15), are put inside other main clusters by the spectral, making it more difficult to exclude the unnecessary ones. Finally, tracks are reconstructed starting from clustered trajectories, and macro-clusters are computed. Table 3.17 and Table 3.18 respectively shows results about the computation of agglomerative and spectral macro-clusters. Analysing the results obtained from the tests performed, it is possible to infer that both agglomerative and spectral clustering are suitable for this problem, thus demonstrating

Table 3.15: Performance of the agglomerative clustering.

Number of Trajectories	Type	Accuracy
99	blue	25.38%
103	orange	26.41%
72	pink	18.46%
95	green	24.36%
21	others	5.39 %

Table 3.16: Performance of the spectral clustering.

Number of Trajectories	Colour	Accuracy
79	blue	20.26%
127	orange	32.56%
66	pink	16.92%
46	green	11.79%
34	yellow	8.72%
38	gray	9.74%

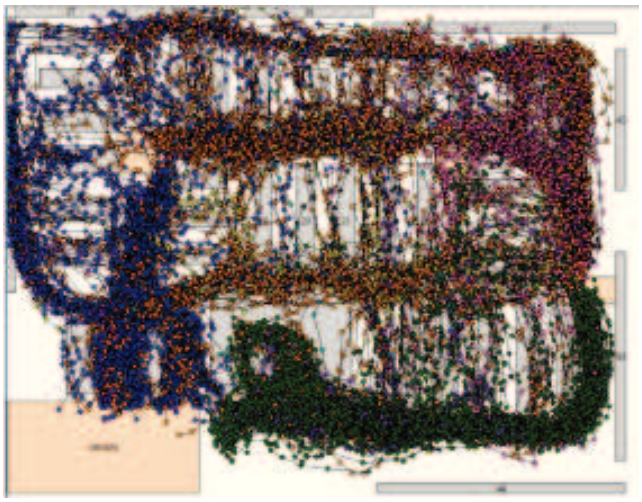


Figure 3.33: Agglomerative clustering.

the effectiveness of the proposed approach. Furthermore, agglomerative clustering produces more macro-clusters than the spectral one (21, 20, 14 compared to 13, 12, 12). However, the first approach shows more abnormal or singular trajectories (1 cyan, 6 purple, 6 yellow, 8 gray).

Table 3.19 and Table 3.20 provide the analysis of the persistence of the clusters obtained with the agglomerative and with the spectral algorithms respectively, as the days wore on. In particular, one working week (from monday to Saturady) of the store target is taken into exam.

From Table 3.19, it can be evaluated the following aspects. The blue is present in every day of the week. It is a normal occurrence and it is in line with

Table 3.17: List of agglomerative macro-clusters.

Code	Number
blue → orange → pink → green	21
blue → orange → green	20
orange → green	14
blue → blue → orange → pink → green	6
blue → orange → pink → pink → green	6
blue → pink → green	6
orange	6
others	22

Table 3.18: List of spectral macro-clusters.

Code	Number
orange → blue	13
green → orange → blue	12
orange	12
orange → pink → blue	11
gray → yellow → pink → blue	8
green → yellow → pink → blue	7
others	32

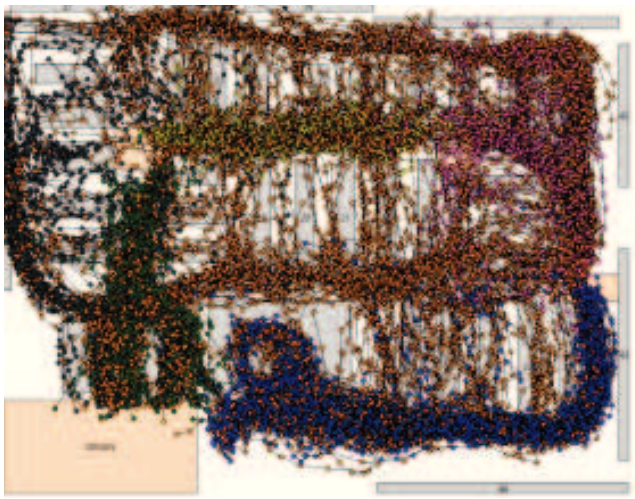


Figure 3.34: Spectral clustering.

the customers trajectories, because the blue cluster is located in the checkout aisles, thus near the exit of the supermarket. The same inference can be done for the cluster green, which is come along every day of the week being placed near the entrance in the fruit and vegetables section. This cluster is consistent with a huge number of sub-trajectories.

Instead, from the results obtained with the spectral algorithm it can be inferred that the pink cluster is consisted of a large number of sub-trajectories. It is located near the entrance in the fruit and vegetables section, for this reason

Table 3.19: Performance of the agglomerative clustering.

Cluster	Day1	Day2	day3	day4	day5	day6
pink	49	37	35	43	72	95
green	0	0	4	0	0	0
yellow	0	26	24	19	0	72
grey	34	38	0	4	37	105
cyan	0	0	0	0	0	0
orange	47	36	35	36	92	99
others	80	10	49	50	86	19
Total	210	147	147	152	287	390

Table 3.20: Performance of the spectral clustering.

Cluster	Day1	Day2	day3	day4	day5	day6
pink	31	32	23	33	73	123
green	11	14	13	0	0	32
yellow	10	10	0	17	14	28
grey	10	16	21	14	27	27
cyan	34	18	16	20	43	46
orange	0	16	18	16	18	36
others	114	41	57	52	112	98
Total	210	147	148	152	287	390

it is noticed a marked permanence of it during the days. The green cluster comes along every day of the week and it is placed in the refrigerated area and in the personal care zone. It has a smaller number of trajectory than the pink one. However, it is well defined and present every day. The same evaluation can be done for the yellow (oil and salt shelves) and the grey (butcher area). In this case, for the friday and saturday (the days prior to the closing day of the supermarket) it is possible to notice an increase of the sub-trajectories.

3.5 Digital Cultural Heritage

Nowadays, museum visits are perceived as an opportunity for individual to explore and make up their own minds, to test their own interpretations instead of the expert's once and they became a tool of entertainment such as theaters or cinemas. Over time, museums and art galleries have preserved the important Cultural Heritage (CH) and served as important sources of education and learning. Moreover, visitors are increasingly taking an active role within museums. The visitor experience is not adequately described by understanding the content, the design of exhibitions, or even by understanding visit frequency or the social arrangements in which people enter the museum. To get a more complete answer to the questions of why people do or do not visit museums, what they do there, and what learning/meaning they derive from the experience, researchers' efforts have been aimed at better describing and understanding the

museum visitor experience.

In addition, nowadays museums have a unique perspective on technology's impact. A modern approach to the fruition of art is actually based on a wide and targeted use of technologies [285, 286] and a growing number of museums is adopting digital tools as integral part of the exhibition, providing users new instruments to study art deeply [287, 288]. The increasing technical capabilities of Augmented Reality (AR) technology have raised audience expectations, advancing the use of mobile AR in CH settings [289]. At the same time, the attention regarding the use of AR has shifted from purely attracting and entertaining audiences, to finding proper ways of providing contextually relevant information that can enhance visitor experiences. The visualization of digital contents through a display is allowed with the same point of view of the user, by superimposing virtual objects on the real scene. As well, AR permits the visualization of virtual objects (e.g. 3D models, audio, text, images) avoiding the use of artefacts (i.e. QR code) to retrieve contents, besides permitting an automatic and interactive visualization of Points of Interest (POIs) [290]. This might positively influence visitors' experience, hence museum appeal.

3.5.1 An HMM-based approach to eye-tracking data for Augmented Reality Applications.

For this reason, in a previous work [217] a protocol to understand the visual behaviour of subjects looking at paintings is defined, using eye-tracking technology, in order to optimize an existing AR application [289, 291].

In fact, eye tracking is a methodology whereby the position of the eye is used to determine gaze direction of a person at a given time and also the sequence in which they have been moved [292]. Eye-movement data consist of eye fixations and saccades. The first one are brief moments that the eye is still and information is extracted from the stimulus (about two to four times per second). The saccades, instead, are rapid jumps of the eye between fixations to redirect the line of sight to a new location [293].

Eye gaze data have been used in many fields such as psychology, neurology, ophthalmology and related areas to study oculomotor characteristics and abnormalities, and their relation to mental states and cognition, because of their relation with internal brain processes(thinking, cognition, etc.). In many research fields eye-tracking devices are used for analysing user behaviour [294], such as in market research [295], human computer interaction [292] and visualization research [296]. Due to the wide fields of eye tracking applications and the types of research questions, different approaches have been developed to analyze eye-tracking data such as statistical algorithms (either descriptive or inferential) [297, 298], string editing algorithms [299], visualization-related

techniques, and visual analytics techniques [300]. Eye-tracking can be used to determine what objects in a visual scene a person is interested in, and thus might like to have annotated in their augmented reality view.

In this paper, the main goal is to provide a truly predictive model of the museum visitor visual behaviour. Eye-tracking could provide quantifiable learning outcomes and rich contextual customized learning environment as well as contents for each single individual. An Hidden Markov Model (HMM) approach to predict users' attention is presented, in front of a painting, measured by an eye tracker. Therefore, areas of interest (AOI) most visited are used and it is tried to predict the next transitions between AOIs. Furthermore, this work provides a users behaviour comparison between adults and children in front of a painting. The application of this approach yields good results in terms of precision, recall and F1-score and demonstrates its effectiveness.

Several contributions are made by this work. First of all, the model is generic, so it can be applied to any sequential datasets or sensor types. Second, this model deals with the problem of scalability. Finally, the approach is validated using real data gathered from eye-tracking acquisitions which helps to make results more confident and experiments repeatable. The innovative aspects of this paper lies in proposing an adequate HMM structure and also the use of eye trajectories to estimate the probability that a certain AOI transition will be performed. This model could be a representation of the attention scheme that can be incorporated in the AR applications to have a transition probability or to guide the user on a novel AR interaction scheme. In fact, the test are performed on two class of users adults and children, thus proving a reliable approach to eye trajectories.

Figure 3.35 summarizes and compares the approach with a classical approach. After the definition of AOI, the approaches could be: Expert Based or User Data Driven. In the first one, there is a manual AOI definition done by expert, then the usability of AR applications is evaluated. In the User Data driven Approach, which is the one proposed in this paper, an eye-tracking dataset is built and a HMMs are designed with the aim of estimating the AOIs transition probability. The accuracy is estimated and these models step by step help the user, with a voice guidance, in the painting vision. Finally, the Usability of this approach is compared with the usability of the first one.

3.5.2 Eye-Tracking Dataset.

The eye-tracking data are recorded by using a Tobii Eye-Tracker X2-60 (Figure 3.37) and the Imotions®Attention Tool software (vers. 5.7), as described in [217]. The eye-tracking dataset stores eye tracking data collected from 80 participants to the tests. In particular, the 40 adults taken in exam in this

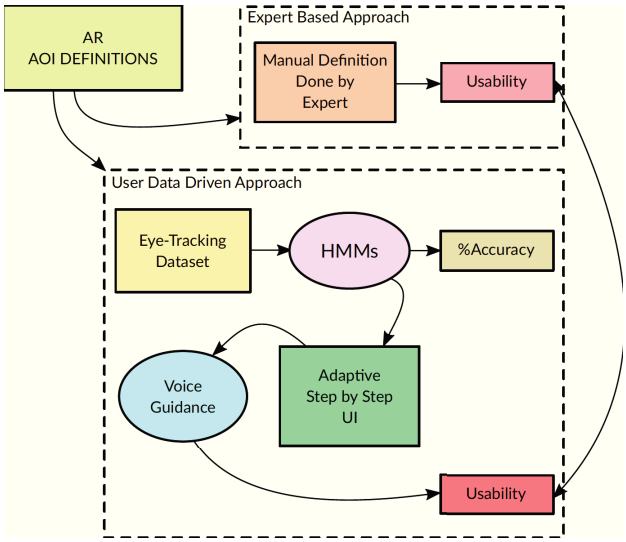


Figure 3.35: Workflow of the approach.

work are forty Italian students and employees at Universitá Politecnica delle Marche. Instead, the 40 children are students of primary school. In both cases, all the acquisitions are collected in a quiet room and under standard illumination conditions. Each participant was seated 60 cm from the eye-tracker and monitor (Figure 3.36).

The digital versions of the painting was showed in a 23" inches monitor, at a resolution of 1920×1080 pixels, preserving the original aspect ratio. The eye-movement indicators, on which the analysis is based, are fixations and saccades. Fixations are eye pauses over a particular of interest averaging about 300 ms. Saccades are rapid eye-movements between fixations. Participants were informed that their eye-movements were recorded. Each trial started with a 9-point calibration pattern, then the study starts. The focus is on the eye trajectories collected for the painting "The Ideal City". The subjects analysed have to observe a faithful reproduction of the picture "The Ideal City" as if they were at the museum. For this test, it is used the Eye-Glasses mobile eye-tracker. The average time of observation registered was 64 seconds. It is performed a pre-test, useful for comparing the outcomes between using digital image and using the real-size artwork. The six-framed details, shown in Figure 3.38 were defined according to the existing AR application for "The Ideal City". They include some architectural details that experts considered relevant in this painting: the doves, the vanishing point; the capitals; the landscape in the background; the floor and the geometry that characterize the whole painting.

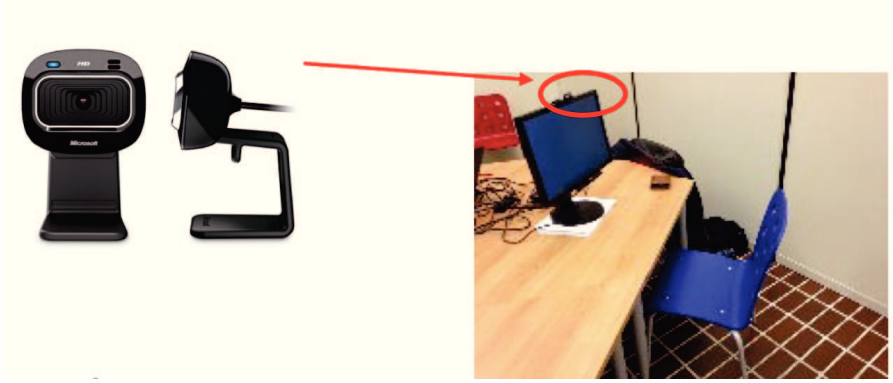


Figure 3.36: Eye-tracking device and subjects position in front of the screen for eye-tracking acquisitions.

Figure 3.37: Eye-tracking Device.

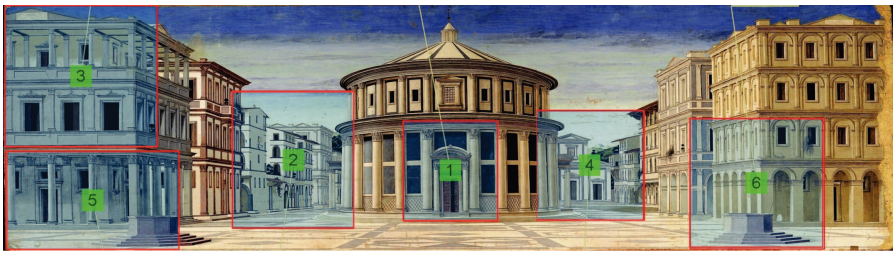


Figure 3.38: Areas of Interest of "The Ideal City".

Figure 3.39: Adults Heatmap of "The Ideal City".

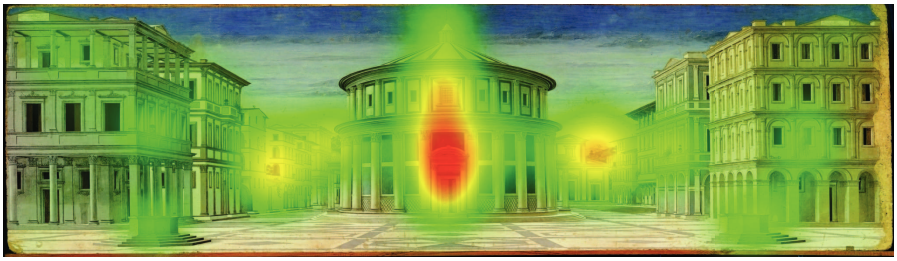


Figure 3.40: Children Heatmap of "The Ideal City".

Data collected were extracted using the IMotions®Attention Tool software and they are analysed using STATA vers.13. IMotions®provides different metrics for each AOI such as the TTFF-F, the Time spent-F, x and y (the coordinates of fixation). The TTFF-F represents the time to first fixation or in other words, it identifies which AOI the participants saw at first sight. The Time spent-F provides the time spent in a specific AOI. In general, a low time value of TTFF-F indicates that the participant's fixation for that particular AOI started immediately as the image appeared on the screen. Instead a high time value of TTFF-F shows that the fixation has started late or not started. The TTFF-F value is equal to the entire exposure time of the image when the fixation not started. Figure 3.39 represents the heat map, for the 40 adults participants, when they were asked to observe the painting as they were at the museum. Figure 3.39 represents the heat map, for the 40 children participants, when they were asked to observe the painting as they were at the museum.

Design of HMM Structure

Let:

$$X = \{x_1, x_2, \ldots, x_n\}$$

be a discrete finite AOI attraction space and

$$O = \{o_1, o_2, \ldots, o_m\}$$

the observation space of a HMM [279]. Let T be the transition matrix of this HMM, with $T_{x,y}$ representing the probability of transitioning from attraction in AOI $x \in X$ to attraction in AOI $y \in X$, and $p_x(o)$ be the emission probability of observation $o \in O$ in attraction in AOI $x \in X$.

The probability that HMM trajectory follows the attraction sequence s given the sequence of n observations, is denoted as:

$$P(X_{1:n} \in seq_n(s)|o_{1:n})$$

where $seq_n(s)$ is a set of all length n trajectories whose duration free sequence equals to s.

Finding the most probable attraction sequence can be seen as a search problem that requires evaluation of probabilities of attraction sequences. The Viterbi algorithm [280] based on dynamic programming can be used to efficiently find the most probable trajectory. In fact, it makes use of the Markov property of an HMM (that the next state transition and symbol emission depend only upon the current state) to determine, in linear time with respect to the length of the emission sequence, the most likely path through the states of a model which might have generated a given sequence. A Viterbi-type training algorithm based on the maximum likelihood criterion is also derived.

After training the model, a trajectory

$$s = \{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$$

is consedered and its probability λ is calculated for the observation sequence $P(s|\lambda)$. Then, the trajectory is classified as the one which has the largest posterior probability. To each observation it is associated the ground truth (state) that is the AOI from which the user was attracted. The observation are the grid cells in which the painting is divided. The number of vertical layers is 10 and it is the same of the horizontal layers used in the quantization step for each HMM .

3.5.3 Performance Evaluation and Results

An architecture to implement HMMs eye-tracking trajectories is proposed. The architecture uses the eye-tracking data to classify different AOIs.

The standard algorithm for HMM training is the forward-backward, or Baum-Welch algorithm [283]. Baum-Welch is an iterative algorithm that uses an iterative expectation/maximization process to find an HMM which is a local maximum in its likelihood to have generated a set of training' observation sequences. This step is required because the state paths are hidden, and the equations cannot be solved analytically.

Table 3.21: Adults Classification Results Cross Validation HMM

AOIs	precision	recall	f1-score
AOI_1	0.97	0.88	0.92
AOI_2	1.00	0.94	0.97
AOI_3	0.53	0.82	0.64
AOI_4	0.83	0.89	0.86
AOI_5	0.74	0.95	0.83
AOI_6	0.18	0.99	0.31
avg / total	0.92	0.90	0.90

In this work, the Baum–Welch algorithm has been employed to estimate a transition probability matrix and an observation emission matrix so that the model best fits the training dataset.

Since the discrete observation density is adopted in HMMs implementation, a Vector Quantization and clustering step is required to map the continuous observation in order to convert continuous data to discrete data. The results were obtained using the cross validation technique.

Below, the results are given for each AOI in which the examinated painting is divided. The main goal is to gradually estimate the AOIs transition probability. A k-fold cross-validation approach (with $k = 5$) to test the HMM is applied. The resulting confusion matrix for adults is shown in Figure 3.41. As it is possible to see in the confusion matrix the AOIs eye transitions are detected with high accuracy.

Table 3.21 summarises the results demonstrating the effectiveness and suitability of the approach.

From the classification results, it is possible to infer that the AOIs 1, 2, 4 reach high values of precision, recall and F1-score because are the AOIs in which the attention is most focused, as shown by the heatmap (Figure 3.39).

In Figure 3.42, that represents the modelling of the transition matrix of the total recorded period, it could be seen from the matrix representation, the probability of transition from one state to another, whose values range from zero to one. The fact that the probability values along the matrix diagonal are very high means that the probability that in the next instant (K + 1) eyes remains in the same AOI of the previous instant is very high. In other words, it is very likely that once eyes observe certain AOIs, they continues observing those same AOIs also in the next instant. Furthermore, it is possible to see other elements, not along the diagonal, with significant probability values such as the elements (1, 5), (1,6), (5,3) because the user has a high probability to focus his/her attention passing from the center of the painting to the AOI in the corners. Table 3.22 reports the results for children eye-tracking.

Figure 3.44 depicts the modelling of the transition matrix of the total recorded

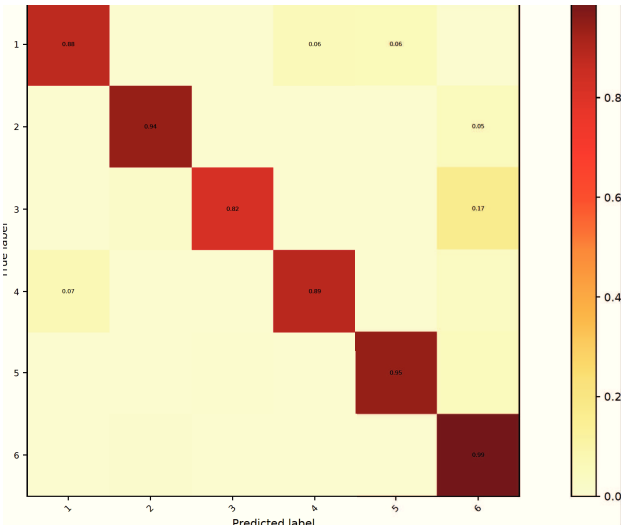


Figure 3.41: Adults Confusion Matrix.

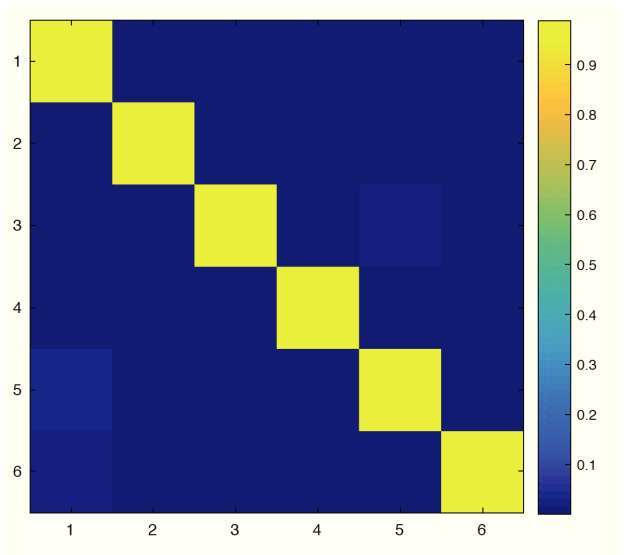


Figure 3.42: Adults Transition Matrix.

Table 3.22: Children Classification Results Cross Validation HMM

AOIs	precision	recall	f1-score
AOI_1	0.97	0.76	0.84
AOI_2	0.75	0.91	0.83
AOI_3	0.53	1.00	0.69
AOI_4	1.00	0.75	0.85
AOI_5	0.58	0.70	0.64
AOI_6	0.78	0.31	0.44
avg / total	0.83	0.77	0.77

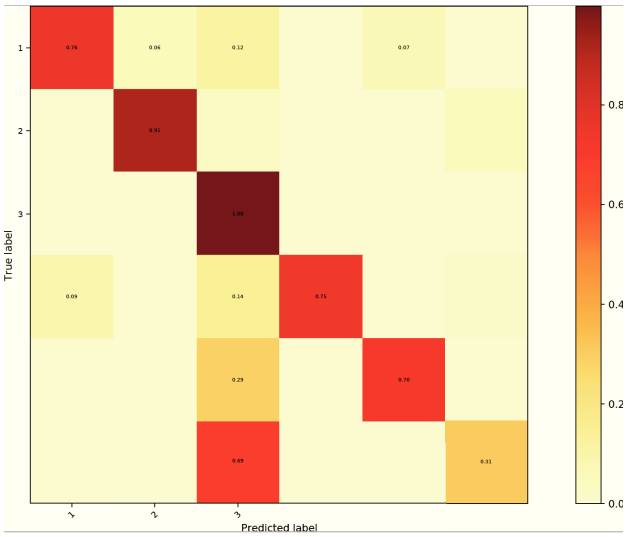


Figure 3.43: Children Confusion Matrix.

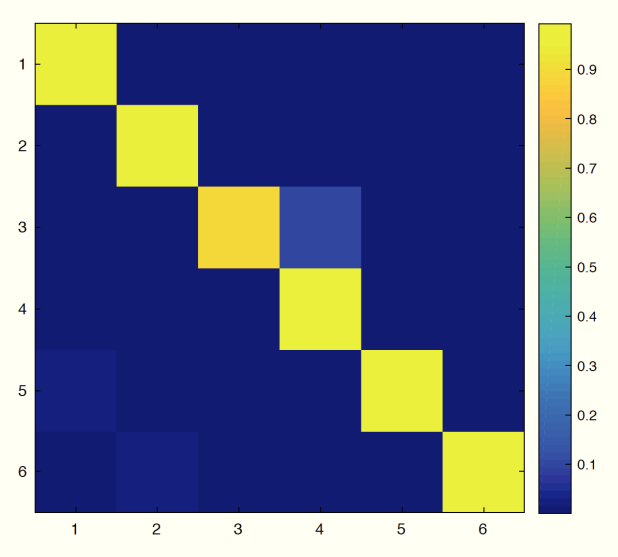


Figure 3.44: Children Transition Matrix.

period for the children. Even in this case, the fact that the probability values along the matrix diagonal are very high means that the probability that in the next instant (K + 1) eyes remains in the same AOI of the previous instant is very high. It is entirely possible that children eyes observe certain AOIs, they continues observing those same AOIs also in the next instant. Moreover, other elements, not along the diagonal, with significant probability values such as the elements are (1, 5), (2,1), (4,3) because the user has a high probability to focus his/her attention passing from the center of the painting to the AOI in the corners.

Experimental results demonstrate that the eye-tracking system provides useful input information to design a personalised user experience in AR applications. In fact, this kind of model will be used as a seed to automatically complete all the information needed to build the AR application.

Chapter 4

Discussion: Limitations, challenges and lesson learnt

The examples illustrated and the use cases described so far, have proved that the followed pipeline, raised in the Introduction (Chapter 1), is suitable for the development of the challenging applications of computer vision in which PR is the main core. Even if each scenario has different features and needs, it has been possible to outline a common path in every ambit. The different PR approaches experimented, have demonstrated that it is possible to cope with each need. The overriding goal was not only to present interesting PR solutions, but also to introduce challenging computer vision problems in the increasingly important five domains chosen, accompanied with benchmark datasets and suitable performance evaluation methods. For a given problem, information can be obtained from multiple sources at different abstraction levels. Combining information from multiple sources can further boost the performance. For example, in SMI field combining images and text to obtain joint features representation. Furthermore, despite a method thriving in a specific computer vision problem, it has not been nearly as impactful in another research area. Collecting and comparing contributions that identify such challenges, and propose methods to overcome current limits is the thesis milestone.

Faced with many types of data, such as images, biological data and trajectories, a key difficulty was to find relevant vectorial representations. While this problem had been often handled in an ad-hoc way by domain experts, it has proved useful to learn these representations directly from data, and Machine Learning algorithms, HMM and DCNN have been particularly successful. The representations are then based on compositions of simple parameterized processing units, the depth coming from the large number of such compositions. It was desirable to develop new, efficient data representation or feature learning/indexing techniques, which can achieve promising performance in the related tasks.

In tables 4.1 and 4.2 a comprehensive comparison of the issues related to the creation of PR applications, needs and limitations are described.

4.1 Thesis Contributions

The main contribution of this thesis can be summarized in the following aspects: the definition and development of more efficient and effective frameworks that learn and recognize patterns in different applications. The computer vision applications in which PR is the key core in their design, starting from general methods, that can be exploited in more fields, and then passing to methods and techniques addressing specific problems. The applications of PR techniques are devoted to real-world problems, and on interdisciplinary research, experimental and/or theoretical studies yielding new insights that advance PR methods. New research lines, are to spur especially within interdisciplinary research scenarios.

To make PR techniques tailored for the aforementioned challenging applications, considerations such as computational complexity reduction, hardware implementation, software optimization, and strategies for parallelizing solutions must be observed. The overarching goal of the work consisted of presenting a pipeline to select the model that best explains the given observations; nevertheless, it does not prioritize in memory and time complexity when matching models to observations.

Extensive efforts are devoted to collecting training and testing data and five newly challenging datasets are specifically designed for the described task. The design of benchmark involved several issues that range from the objective collection in order not to give any method unfair advantages to the consideration of the specificity of the concrete situation (the methods design that is tuned to a specific problem do not work properly on other problems).

In Figure 4.1 are highlighted in red the main contributions of this thesis for the analysed fields. In particular, from this figure, it is possible to deduce that five newly challenging datasets are specifically designed for each task and the learning methods are evaluated on the proposed datasets: Biological Reference Dataset (BRD), sCREEN, Top View Person Re-identification (TVPR), GfK Verein dataset and Eye-Tracking dataset. In Biology and SMI domains, the *Features Extraction* is an innovative aspect, instead in retail the *Tailored Learning Approach* is the step in which the new contribution of this work is relevant. In Biology, SMI, Surveillance, Retail and DCH, in the step *Comparative Analysis and Assessment* an important contribution is made, since various details have investigated such as domain dependence and prior information, computational cost and feasibility, discriminative features, similar values for similar patterns, different values for different patterns, invariant features with respect to translation, rotation and scale, robust features with respect to occlusion, distortion, deformation, and variations in environment. Moreover, performances with training sample were estimated, performances with future data were predicted and problems of overfitting and generalization were evaluated.

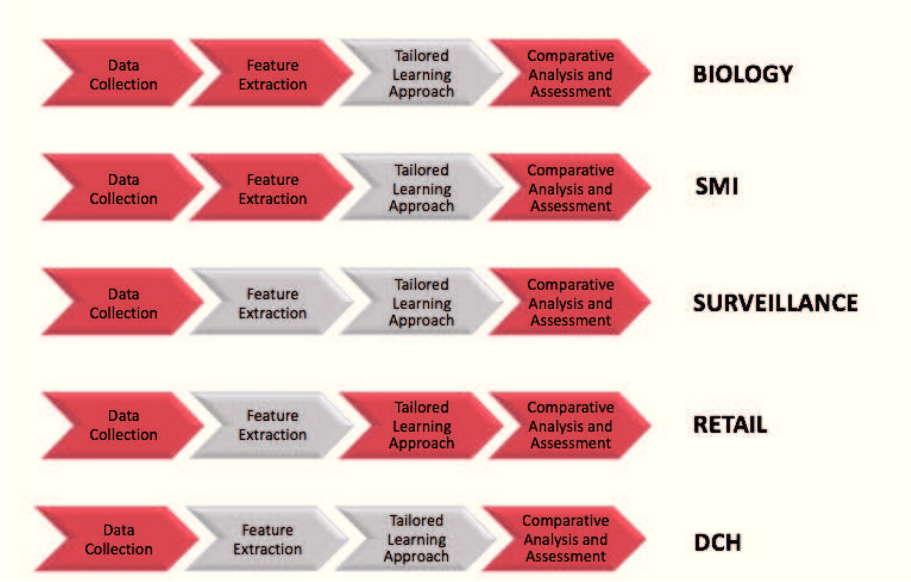


Figure 4.1: Pipelines in which are highlighted the main contributions of this thesis for each domain taken into exam.

4.2 Limitations

The common thread has been outlined and depicted in Figure 4.1. It was also given an exhaustive description of the new possibilities offered by different techniques and approaches to provide and collect data. However, during the studies for this thesis, several limitations emerged that still exist and that are preventing the effectiveness PR applications.

In case of *Biology*, emerged the necessity to have more "A" data; this to make the BRD more consistent and the prediction more precise. This aspect is not simple to tackle. In ART, the percentage of positive outcome is extremely low due to the psychological problem of women related to infertility. Thus, collecting data of class "A" is extremely complex as well as making prediction in this domain. In fact, the lack of success in implantation depends not only on the right prediction but also to other unpredictable factors such as woman's predisposition towards the embryo or the hormonal status. A possible solution involves virtualization, standardization of technologies and special regulations for the protection of personal data, which will define the basis of a distributed system cloud-centered. All the data could be stored in a cloud that collects oocytes data from all centres, healthcare institutions/hospitals and assisted reproductive clinics. The cloud technology allows updating the system constantly. In this way, the classifiers accuracy will be increased.

An issue arises in *SMI* domain was related to the sentiment evaluation so

the Dataset annotation. In fact, it is difficult to judge the sentiment in a standard way because it is a subjective Task. For this reason, the Cohen's Kappa Coefficient k is calculated for measuring the agreement between the two annotators beyond chance. Another aspect was the collection of negative visual brand-related social media pictures. Consumers tend to express their overall negative sentiment towards brands by adding negative text to neutral or positive motives. As people avoid posting pictures with negative facial expressions on social media, the most frequent form of visual negative sentiment is graphics with many different motives. For the same reason, the low availability of negative visual brand-related social media pictures was a limit, since consumers prefer visual clues such as happy people or smileys to textual clues for showing their overall positive sentiment towards brands, positive texts are less expressive.

For *Surveillance*, the task is difficult since recognizing a person with a camera in a top-view configuration is a complex. However, as stated in Chapter 3, the accuracy is good even if people that have similar features are confused. For example, two guys in TVPR dataset that have same T-shirt, height and colour of hair are confused by the classifiers.

In *Retail*, the limitations of sCREEN is due to the RTLS. In particular, the whole store must be covered by anchors and each shopping cart/basket must be equipped by tag. Another problem is the anchor installation made difficult to the store layout.

Finally, the *DCH* application was limited to the HMMs structure that does not take into account the sequence of states leading into any given state and for their Markovian nature, the time spent in a given state is not captured explicitly.

4.3 Challenges

The proposed applications, described in Chapter 3 open up a wealth of novel and important opportunities for the PR community. The dataset collected as well as the complex areas taken into exam, make the research challenging. Intensive attention has been drawn to the exploration of tailored learning models and algorithms, and their extension to more application areas. The combination of the new deep learning and traditional methods in PR and artificial intelligence has also demonstrated benefits. However, most existing methods directly borrow the models for multimedia tasks without considering the distinctiveness of multimedia data and multimedia tasks. As a result, these methods hardly fit the requirements of these multimedia tasks. The tailored methods, adopted for the development of the proposed applications, have shown to be capable of extracting complex statistical features and efficiently learning

their representations, allowing it to generalize well across a wide variety of computer vision tasks, including image classification, text recognition and so on. In order to cope with these new multimedia tasks, current models, including their architectures, training and inference methods, must be adapted or even re-designed. A number of fundamental issues had to be solved for emerging multimedia data, multimedia computing and applications. For example, how to train the DNN for brand-related social media pictures; how to design novel architectures for emerging multimedia tasks; how to conduct multimedia caption at multi-granularity from global correspondence to local correspondence; how to enhance the models to support simultaneous multimedia knowledge extraction and reasoning, to name a few.

For the *Biology* domain, in ART implantation there is not a criteria for the oocyte choice. Its selection is based on the morphological features of the cytoplasm, polar body and cumulus cells. However, all these criteria for grading and screening oocytes are subjective and controversial, and seem to not be related to the intrinsic competence of the oocyte. As a consequence of this limitation, there is a strong demand for a complete automation of the procedure, that would result in increased test repeatability and reliability, easier and faster result reporting. It is the first automatic approach for learning and classifying oocytes quality.

In *SMI*, with the increasing popularity of social networks and image sharing platforms more and more opinions are expressed by pictures. Several solutions have been proposed for the sentiment analysis of visual content. However, a multitude of consumers' pictures does not only include visual elements, but also textual elements. For example, people take pictures of advertisement posters or insert text into photos with the aid of photo editing software. For the companies, it is of great importance estimate the overall sentiment of a picture in order to have a direct feedback on a product. Until now sentiment analysis has been performed on either only textual content or only visual content. This is the first approach to consider visual and textual information in pictures at the same time.

The *Surveillance* is a rather broad concept. Apart from safety and security, video surveillance has a wide variety of applications in numerous other aspects of life. Digital video pays a pivotal role in a myriad of surveillance applications. Automated means of video surveillance for public safety enhancement have existed for quite some time but have gained significant popularity only in recent years. Enhanced awareness for public safety is leading to innovative research by making use of multimedia information and telecommunications for disaster and crime prevention and management of secure environments. All of this, in turn, is leading to new applications of surveillance in homeland security and crime prevention through indoor and outdoor monitoring and monitoring of

critical infrastructures, highways, parking garages, and shopping malls. The challenge, in this case, is to collect data of 100 people with an RGB-D camera installed in top-view configuration, acquired across intervals of days and at different times. This choice is due to its greater suitability compared with a front view configuration, usually adopted for gesture recognition or even for video gaming. The top- view configuration reduces the problem of occlusions and has the advantage of being privacy preserving, because the face is not recorded by the camera.

The innovative aspects of *Retail* application are in proposing an adequate HMM structure together with the use of trajectories and time of consumer attraction in front of a shelf/category to forecast that a certain attraction will be performed. The continuous modelling approach and the novel HMM structure is also able to model changes in planogram and store layouts or usual retail events such as products experiencing shelf out-of-stock. Furthermore, a framework for clustering customers trajectories in stores is proposed. This framework enables retailers to discover common shoppers subtrajectories, whereas previous frameworks do not. Also, this approach enables retailers to compare the impact of different store layouts (maps) or shelf layouts (planograms) on issues such as ease of selection, trading up and the overall shopping experience. Finally, the work is validated using real data gathered from a real store which helps to make the results more confident and our experiments repeatable; used data are publicly available in an open dataset that is the first dataset based on extensive real data in this field.

The challenge faced in *DCH* lies in proposing an adequate HMM structure and also the use of eye trajectories to estimate the probability that a certain AOI transition will be performed. This model could be a representation of the attention scheme that can be incorporated in the AR applications to have a transition probability or to guide the user on a novel AR interaction scheme. In fact, the tests are performed on two class of users adults and children, thus proving a reliable approach to eye trajectories.

4.4 Lesson Learnt

The PR applications described in this thesis deserve some comments. PR techniques are delivering a promising solution to develop systems and to make innovation at a rapid pace. The combination of ICT technologies offers a framework for building large scale applications relying on data gathered from a complex infrastructure of sensors and smart devices. Numerous challenges exist in implementing such a framework, one of them is to meet the data and services requirements on informatics based applications in terms of energy efficiency, sensing data quality, network resource consumption, and latency. In

fact, the convergence of PR techniques (machine learning, statistical methods and Deep Learning) with reference to quality of data and services for real-world applications has three main components: intelligent devices, intelligent system of systems, and end-to-end analytics. Further, PR approaches had addressed various challenges such as anomaly detection, multivariate analysis, streaming and visualization of data.

As outlined in Chapter 2, recent literature has addressed the inherent power of fusion between different PR approaches. It can provide effective solutions for machine understanding of data (structured/semi structured), optimization problems, specifically, dealing with incomplete or inconsistent information. It is concerned with constructing systems that can improve the experiences in this work. The PR techniques proposed to meet the challenge of massive data processing, of which semi-supervised learning is a hot topic and should be one of the most important techniques. However, the cost of labelling the data is large because of expert experience or experiments, so only part of the data is labelled. Semi-supervised learning can utilize the unlabelled data. There are different methods to utilize the unlabelled data, of which clustering is a state-of-the-art method. But it does not work when it meets huge data.

Many improvements, from different perspectives, should be considered in the technology, so that the challenging nature of the requirements for the current and future computing environments can be accommodated.

Table 4.1: Summary of challenges, limitation and issues related to the creation of PR applications in the different domain described.

Application domain	**Challenge**	**Pattern**	**Limitations**	**Main Contributions**
Biology	Automatic classification of human oocytes in ART.	Human GCs.	No consistent Dataset due to the low percentage of positive outcome; The positive outcome depends not only to the right prediction but also to other unpredictable factors.	First Machine Learning approach for learning and classifying oocytes quality; Proposal of a significant set of bio-features able to achieve relevant classification performances; The first comparison between different classifiers to evaluate a family of machine learning algorithms able to obtain significant results in this field.
SMI	Visual and Textual Sentiment Analysis of brand-related social media pictures	Social Media images brand-related	Subjective Task; Low availability of negative visual brand-related social media pictures; Low availability of positive text in brand-related social media pictures.	The first study on sentiment analysis of brand-related pictures on Instagram; Sentiment analysis for both visual and textual information, with text included in a picture, which is more challenging since the text has to be detected and recognized first, before its sentiment can be identified; Visual and textual features extracted from two specially trained DCNNs.
Surveillance	Person Re-identification with RGB-D camera in Top-View configuration	Video of people passing under the camera	People that have similar features are confused.	The top-view configuration reduces the problem of occlusions and has the advantage of being privacy preserving, because a person's face is not recorded by the camera.

Table 4.2: Summary of challenges, limitation and issues related to the creation of PR applications in the different domain described.

Application domain	Challenge	Pattern	Limitations	Main Contributions
Retail	Modelling and Forecasting customers navigation in store	Shopping carts and baskets Trajectories	The whole store must be covered by anchors and each shopping cart/basket must be equipped by tag; Problem related to anchor installation due to the store layout.	This framework enables retailers to discover common shoppers subtrajectories, whereas previous frameworks do not; approach that enables retailers to compare the impact of different store layouts (maps) or shelf layouts (planograms) on issues such as ease of selection, trading up and the overall shopping experience; work validated using real data gathered from a real store which helps to make our results more confident and our experiments repeatable.
DCH	An HMM-based approach to eye-tracking data for AR Applications	Eye trajectories	The HMMs do not take into account the sequence of states leading into any given state; due to their Markovian nature, the time spent in a given state is not captured explicitly.	An adequate HMM structure and the use of eye trajectories to estimate the probability that a certain AOI transition will be performed; This model could be a representation of the attention scheme that can be incorporated in the AR applications to have a transition probability or to guide the user on a novel AR interaction scheme.

Chapter 5

Conclusions and future works

In this thesis, the development of challenging computer vision applications for real applications such as Biology, Retail, Surveillance, Social Media Intelligence and Digital Cultural Heritage, in which PR is the main core, is described. Even if each scenario presents different features and needs, it has been possible to outline a common path in every ambit. The extensive use of PR in many areas has motivated and led to comprehend if such methods are able to exploit multimedia data.

Experiencing solutions in different domains, the studies were oriented towards understanding potentials and weak points, from a multidisciplinary perspective. The main scope has been to bring together PR methods for computer vision applications in order to give a landscape of techniques that can be successfully applied and also to show how such techniques should be adapted to each particular domain. The privileged focus was on up-to-date applications of PR techniques to real-world problems, and on interdisciplinary research, experimental and/or theoretical studies yielding new insights that advance PR methods.

The thesis has presented reviews, perspectives, new methods and applications in PR. The described contributions reflect a significant advancement both in the state-of-the-art as well as for the scenarios.

Chapter 1 portrays the concept of *PR*, starting from the importance of data, regardless the type of data dealing with.

With the second Chapter a review the state of art about methods and applications has been provided with a specific focus on the state of the art in the five chosen domain. In particular, for each research field was analysed the methods and techniques, also main paths that most approaches follow are summarized and their contributions are pointed out. The reviewed approaches were categorized and compared from multiple perspectives, including methodology, function and analyse the pros and cons of each category.

The use cases faced during the research activity are collected in Chapter 3. It was mainly devoted to the exploitation of cutting edge scientific methodologies for the solution of problems of relevant interest, as the summa of different

experiences, activities and best practices which can help to contribute, compared with the state of art, to better solving various real-world problems of the computer vision applications in different research areas. The purpose of each scenario was to demonstrate the possibility to apply the same pipeline in several domains; it was also the opportunity to outline points of contact among disciplines apparently different, but that can be joined thank to this common thread. In fact, as stated in Chapter 4, given the large number of case studies experiences, it was possible to outline needs, bottlenecks and weakness points for each domain.

To conclude, it has been done the first steps in introducing these fascinating applications for computer vision. Furthermore, the introduction and release of open benchmark datasets attract more colleagues from the computer vision community in quest toward advancing the state of the art. It has been interesting to investigate the adversarial actions performed on PR techniques, to understand how their analysis can be biased and perturbed by means of the injection of fake data or adversarial examples and how the trustworthiness of the produced knowledge is diminished in relation with the kind and intensity of the performed manipulation (e.g. forged images and videos).

5.1 Future Works

The work described in this document paves the way for future research. Future research directions include the improvement of the algorithms to use other comprehensive features, thereby achieving better performance. Other aspect concerns the increase of the five datasets in order the make the prediction more accurate and precise.

The further efforts in biology involve cybersecurity and cryptography with virtualization, standardization of technologies and special regulations for the protection of personal data, which will define the basis of a distributed system cloud-centered.

In SMI the investigation will be devoted to improve the approach by extracting additional informative features such as peoples' emotions as well as positive and negative symbols.

In Surveillance, an interesting approach is setting up a full neural network for the real time processing of video images as well as the evaluation of the necessary resources for the design of CNN layers. The long term goal could include the integration of this re-id system with an audio framework and the use of other types of RGB-D camera, such as time of flight (TOF) ones. The system can additionally be integrated as a source of high semantic level information in a networked ambient intelligence scenario, to provide cues for different problems, such as detecting abnormal speed and dimension outliers, that can alert one to

a possible uncontrolled circumstance. It would also be interesting to evaluate both colour and depth images in a way that does not decrease the performance of the system when the colour image is being affected by changes in pose and/or illumination.

Analysis based on other indicators will be conducted in retail domain. In this way, an optimization technology to support decision-making on point of sale and shelf locations that will reduce customer traffic congestion, automate some of the sales processes, expand automated services to improve customer service and maximize profit for retailers will be developed. Moreover, the evaluation will be extended by combining the analysis with legacy system information such as customer purchase history for operational efficiency and expanding sales. More analysis models need to be developed for more detailed analysis of the shopping behaviour of the several types of customers, along with the development of various measurement indexes to analyse the store environment. A mobile robot will also be developed that searches for the best product location and improve the retail operation sells. A future vision, following results here presented, will be the integration of the proposed system on intelligent assistive robotics systems. Research on spatial navigation for blind or elder shoppers suggests that a device aimed at helping them to shop independently should provide the shopper with effective interfaces to the locomotor and haptic spaces of the supermarket.

An important research investigation in DCH will improve users' attitude toward mobile advertising. By sensing the user awareness to the system, for instance allowing the application to know what area the user is interested in at the moment with the aid of gaze tracking devices, contents will be provided in a more reliable and proper way. In fact, the model learns from real experiences and updates the probabilities automatically when the function is applied in practice. Similarly, AR users can use their devices to identify which application best fits their profile. The application will search for the most suitable experience by matching user personal profile with the user's content plus context information.

Bibliography

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[2] P. E. Ross, "Flash of genius," in *Forbes*, 1998, pp. 98–104.

[3] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[4] S. Watanabe, *Pattern recognition: human and mechanical*. John Wiley & Sons, Inc., 1985.

[5] R. O. Duda, P. E. Hart *et al.*, *Pattern classification and scene analysis*. Wiley New York, 1973, vol. 3.

[6] T. Pavlidis, "Shape description by region analysis," in *Structural Pattern Recognition*. Springer, 1977, pp. 216–257.

[7] R. C. Gonzalez and M. G. Thomason, "Syntactic pattern recognition: An introduction," 1978.

[8] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic press, 2013.

[9] R. J. Schalkoff, *Pattern recognition*. Wiley Online Library, 1992.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[11] V. N. Vladimir and V. Vapnik, "The nature of statistical learning theory," 1995.

[12] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992, pp. 144–152.

[13] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[14] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[15] H. B. Demuth, M. H. Beale, O. De Jess, and M. T. Hagan, *Neural network design*. Martin Hagan, 2014.

[16] T. H. Bø, B. Dysvik, and I. Jonassen, "Lsimpute: accurate estimation of missing values in microarray data with least squares methods," *Nucleic acids research*, vol. 32, no. 3, pp. e34–e34, 2004.

[17] I. Rish, "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22. IBM New York, 2001, pp. 41–46.

[18] R. Lippmann, "An introduction to computing with neural nets," *IEEE Assp magazine*, vol. 4, no. 2, pp. 4–22, 1987.

[19] S. Kotsiantis and P. Pintelas, "Recent advances in clustering: A brief survey," *WSEAS Transactions on Information Science and Applications*, vol. 1, no. 1, pp. 73–81, 2004.

[20] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.

[21] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 13, no. 8, pp. 841–847, 1991.

[22] M. Sato, Y. Sato, L. C. Jain, and J. Kacprzyk, *Fuzzy clustering models and applications*. Physica-Verlag, 1997.

[23] G. McKachlan and T. Krishnan, "The em algorithm and extensions. john eiley & sons," 1997.

[24] F. Corpet, "Multiple sequence alignment with hierarchical clustering," *Nucleic acids research*, vol. 16, no. 22, pp. 10 881–10 890, 1988.

[25] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.

[26] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview, ii," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 7, no. 6, 2017.

[27] A.-A. Liu, Y.-T. Su, W.-Z. Nie, and M. Kankanhalli, "Hierarchical clustering multi-task learning for joint human action grouping and recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 1, pp. 102–114, 2017.

[28] T. Zhang, R. Ramakrishnan, and M. Livny, "Birch: an efficient data clustering method for very large databases," in *ACM Sigmod Record*, vol. 25, no. 2. ACM, 1996, pp. 103–114.

[29] S. Guha, R. Rastogi, and K. Shim, "Cure: an efficient clustering algorithm for large databases," *Information Systems*, vol. 26, no. 1, pp. 35–58, 2001.

[30] G. Karypis, E.-H. Han, and V. Kumar, "Chameleon: Hierarchical clustering using dynamic modeling," *Computer*, vol. 32, no. 8, pp. 68–75, 1999.

[31] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[32] W. Wang, J. Yang, R. Muntz *et al.*, "Sting: A statistical information grid approach to spatial data mining," in *VLDB*, vol. 97, 1997, pp. 186–195.

[33] G. Sheikholeslami, S. Chatterjee, and A. Zhang, "Wavecluster: A multi-resolution clustering approach for very large spatial databases," in *VLDB*, vol. 98, 1998, pp. 428–439.

[34] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, *Automatic subspace clustering of high dimensional data for data mining applications*. ACM, 1998, vol. 27, no. 2.

[35] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, *Advances in knowledge discovery and data mining*. AAAI press Menlo Park, 1996, vol. 21.

[36] J. M. Peña, J. A. Lozano, and P. Larrañaga, "Learning recursive bayesian multinets for data clustering by means of constructive induction," *Machine Learning*, vol. 47, no. 1, pp. 63–89, 2002.

[37] V. Jensen, "Introduction to bayesian networks springer-verlag new york," *Inc. Secaucus, NJ, USA*, 1996.

[38] T. Kohonen, "Selfrorganizing maps. sec—ond ed," 1997.

[39] A. Strehl and J. Ghosh, "Cluster ensembles-a knowledge reuse framework for combining partitionings," in *Aaai/iaai*, 2002, pp. 93–99.

[40] K. Das and R. N. Behera, "A survey on machine learning: Concept, algorithms and applications," 2017.

[41] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[42] A. J. Izenman, "Linear discriminant analysis," in *Modern multivariate statistical techniques*. Springer, 2013, pp. 237–280.

[43] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial intelligence*, vol. 97, no. 1, pp. 245–271, 1997.

[44] W. M. Van der Aalst, "Data mining," in *Process Mining*. Springer, 2011, pp. 59–91.

[45] B. Settles, "Active learning literature survey," *University of Wisconsin, Madison*, vol. 52, no. 55-66, p. 11, 2010.

[46] L. B. Holder, M. M. Haque, and M. K. Skinner, "Machine learning for epigenetics and future medical applications," *Epigenetics*, no. just-accepted, pp. 00–00, 2017.

[47] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden markov models for complex action recognition," in *Computer vision and pattern recognition, 1997. proceedings., 1997 ieee computer society conference on*. IEEE, 1997, pp. 994–999.

[48] A. Duric and F. Song, "Feature selection for sentiment analysis based on content and syntax models," *Decision Support Systems*, vol. 53, no. 4, pp. 704–711, 2012.

[49] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[50] K. Noda, Y. Yamaguchi, K. Nakadai, H. G. Okuno, and T. Ogata, "Audio-visual speech recognition using deep learning," *Applied Intelligence*, vol. 42, no. 4, pp. 722–737, 2015.

[51] G. Wu, W. Lu, G. Gao, C. Zhao, and J. Liu, "Regional deep learning model for visual tracking," *Neurocomputing*, vol. 175, pp. 310–323, 2016.

[52] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[53] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[54] N. Hou, H. Dong, Z. Wang, W. Ren, and F. E. Alsaadi, "Non-fragile state estimation for discrete markovian jumping neural networks," *Neurocomputing*, vol. 179, pp. 238–245, 2016.

[55] F. Yang, H. Dong, Z. Wang, W. Ren, and F. E. Alsaadi, "A new approach to non-fragile state estimation for continuous neural networks with time-delays," *Neurocomputing*, vol. 197, pp. 205–211, 2016.

[56] Y. Yu, H. Dong, Z. Wang, W. Ren, and F. E. Alsaadi, "Design of non-fragile state estimators for discrete time-delayed neural networks with parameter uncertainties," *Neurocomputing*, vol. 182, pp. 18–24, 2016.

[57] Y. Yuan and F. Sun, "Delay-dependent stability criteria for time-varying delay neural networks in the delta domain," *Neurocomputing*, vol. 125, pp. 17–21, 2014.

[58] J. Zhang, L. Ma, and Y. Liu, "Passivity analysis for discrete-time neural networks with mixed time-delays and randomly occurring quantization effects," *Neurocomputing*, vol. 216, pp. 657–665, 2016.

[59] G. P. Zhang, "Neural networks for classification: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 30, no. 4, pp. 451–462, 2000.

[60] C. Neocleous and C. Schizas, "Artificial neural network learning: A comparative review," *Methods and Applications of Artificial Intelligence*, pp. 750–750, 2002.

[61] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2014.

[62] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[63] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[64] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 1017–1024.

[65] Y. Bengio, N. Boulanger-Lewandowski, and R. Pascanu, "Advances in optimizing recurrent networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on.* IEEE, 2013, pp. 8624–8628.

[66] I. Sutskever, "Training recurrent neural networks," *University of Toronto, Toronto, Ont., Canada*, 2013.

[67] J. Martens, "Deep learning via hessian-free optimization," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 735–742.

[68] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843–2851.

[69] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, Q. V. Le *et al.*, "Large scale distributed deep networks," in *Advances in neural information processing systems*, 2012, pp. 1223–1231.

[70] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[71] Q. V. Le, "Building high-level features using large scale unsupervised learning," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on.* IEEE, 2013, pp. 8595–8598.

[72] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, and G. Penn, "Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on.* IEEE, 2012, pp. 4277–4280.

[73] O. Abdel-Hamid, L. Deng, and D. Yu, "Exploring convolutional neural network structures and optimization techniques for speech recognition." in *Interspeech*, 2013, pp. 3366–3370.

[74] O. Abdel-Hamid, L. Deng, D. Yu, and H. Jiang, "Deep segmental neural networks for speech recognition." in *INTERSPEECH*, vol. 36, 2013, p. 70.

[75] T. N. Sainath, A.-r. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for lvcsr," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on.* IEEE, 2013, pp. 8614–8618.

[76] L. Deng, O. Abdel-Hamid, and D. Yu, "A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on.* IEEE, 2013, pp. 6669–6673.

[77] A.-r. Mohamed, D. Yu, and L. Deng, "Investigation of full-sequence training of deep belief networks for speech recognition," in *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[78] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Transactions on audio, speech, and language processing*, vol. 20, no. 1, pp. 30–42, 2012.

[79] T. N. Sainath, B. Kingsbury, and B. Ramabhadran, "Improving training time of deep belief networks through hybrid pre-training and larger batch sizes," in *Proc. NIPS Workshop on Log-linear Models*, 2012.

[80] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, and M. A. Musen, "Bioportal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications," *Nucleic acids research*, vol. 39, no. suppl 2, pp. W541–W545, 2011.

[81] E. E. Schadt, M. D. Linderman, J. Sorenson, L. Lee, and G. P. Nolan, "Computational solutions to large-scale data management and analysis," *Nature Reviews Genetics*, vol. 11, no. 9, pp. 647–657, 2010.

[82] A. Rosenthal, P. Mork, M. H. Li, J. Stanford, D. Koester, and P. Reynolds, "Cloud computing: a new business paradigm for biomedical information sharing," *Journal of biomedical informatics*, vol. 43, no. 2, pp. 342–353, 2010.

[83] N. H. Shah and J. D. Tenenbaum, "The coming age of data-driven medicine: translational bioinformatics' next frontier," *Journal of the American Medical Informatics Association*, vol. 19, no. e1, pp. e2–e4, 2012.

[84] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for dna microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.

[85] D. J. Stekhoven and P. Bühlmann, "Missforest—non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2012.

[86] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE transactions on knowledge and data engineering*, vol. 26, no. 1, pp. 97–107, 2014.

[87] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of dna-and rna-binding proteins by deep learning," *Nature biotechnology*, vol. 33, no. 8, pp. 831–838, 2015.

[88] J. Lanchantin, R. Singh, Z. Lin, and Y. Qi, "Deep motif: Visualizing genomic sequence classifications," *arXiv preprint arXiv:1605.01133*, 2016.

[89] S. A. Danziger, R. Baronio, L. Ho, L. Hall, K. Salmon, G. W. Hatfield, P. Kaiser, and R. H. Lathrop, "Predicting positive p53 cancer rescue regions using most informative positive (mip) active learning," *PLoS computational biology*, vol. 5, no. 9, p. e1000498, 2009.

[90] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of online learning and an application to boosting," in *European conference on computational learning theory*. Springer, 1995, pp. 23–37.

[91] A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: challenges and opportunities," 2016.

[92] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings in bioinformatics*, p. bbw068, 2016.

[93] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[94] P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov, "Applications of deep learning in biomedicine," *Molecular pharmaceutics*, vol. 13, no. 5, pp. 1445–1454, 2016.

[95] Z. Gillani, M. S. H. Akash, M. M. Rahaman, and M. Chen, "Comparesvm: supervised, support vector machine (svm) inference of gene regularity networks," *BMC bioinformatics*, vol. 15, no. 1, p. 395, 2014.

[96] B. F. Huang and P. C. Boutros, "The parameter sensitivity of random forests," *BMC bioinformatics*, vol. 17, no. 1, p. 331, 2016.

[97] H. Wu, H. Li, M. Jiang, C. Chen, Q. Lv, and C. Wu, "Identify high-quality protein structural models by enhanced-means," *BioMed Research International*, vol. 2017, 2017.

[98] P. D. Reeb, S. J. Bramardi, and J. P. Steibel, "Assessing dissimilarity measures for sample-based hierarchical clustering of rna sequencing data using plasmode datasets," *PloS one*, vol. 10, no. 7, p. e0132310, 2015.

[99] N. Patel and J. T. Wang, "Semi-supervised prediction of gene regulatory networks using machine learning algorithms," *Journal of biosciences*, vol. 40, no. 4, pp. 731–740, 2015.

[100] S. Ciucci, Y. Ge, C. Durán, A. Palladini, V. Jiménez-Jiménez, L. M. Martínez-Sánchez, Y. Wang, S. Sales, A. Shevchenko, S. W. Poser *et al.*, "Enlightening discriminative network functional modules behind principal component analysis separation in differential-omic science studies," *Scientific Reports*, vol. 7, 2017.

[101] K. Severson, B. Monian, J. Christopher Love, and R. D. Braatz, "A method for learning a sparse classifier in the presence of missing data for high-dimensional biological datasets," *Bioinformatics*, 2017.

[102] H. Franken, R. Lehmann, H.-U. Häring, A. Fritsche, N. Stefan, and A. Zell, "Wrapper-and ensemble-based feature subset selection methods for biomarker discovery in targeted metabolomics," in *IAPR International Conference on Pattern Recognition in Bioinformatics*. Springer, 2011, pp. 121–132.

[103] R. K. Padmanabhan, V. H. Somasundar, S. D. Griffith, J. Zhu, D. Samoyedny, K. S. Tan, J. Hu, X. Liao, L. Carin, S. S. Yoon *et al.*, "An active learning approach for rapid characterization of endothelial cells in human tumors," *PloS one*, vol. 9, no. 3, p. e90495, 2014.

[104] H. Cho, B. Berger, and J. Peng, "Reconstructing causal biological networks through active learning," *PloS one*, vol. 11, no. 3, p. e0150611, 2016.

[105] W. Lin and D. Xu, "Imbalanced multi-label learning for identifying antimicrobial peptides and their functional types," *Bioinformatics*, vol. 32, no. 24, pp. 3745–3752, 2016.

[106] P. Kelchtermans, W. Bittremieux, K. Grave, S. Degroeve, J. Ramon, K. Laukens, D. Valkenborg, H. Barsnes, and L. Martens, "Machine learning applications in proteomics research: How the past can boost the future," *Proteomics*, vol. 14, no. 4-5, pp. 353–366, 2014.

[107] M. Lan, C. L. Tan, and J. Su, "Feature generation and representations for protein–protein interaction classification," *Journal of biomedical informatics*, vol. 42, no. 5, pp. 866–872, 2009.

[108] X. Wang, B. Zheng, S. Li, J. J. Mulvihill, M. C. Wood, and H. Liu, "Automated classification of metaphase chromosomes: optimization of an adaptive computerized scheme," *Journal of biomedical informatics*, vol. 42, no. 1, pp. 22–31, 2009.

[109] G. Díaz, F. A. González, and E. Romero, "A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images," *Journal of Biomedical Informatics*, vol. 42, no. 2, pp. 296–307, 2009.

[110] H.-Q. Wang, H.-S. Wong, H. Zhu, and T. T. Yip, "A neural network-based biomarker association information extraction approach for cancer classification," *Journal of Biomedical Informatics*, vol. 42, no. 4, pp. 654–666, 2009.

[111] M. A. Mazurowski, J. Y. Lo, B. P. Harrawood, and G. D. Tourassi, "Mutual information-based template matching scheme for detection of breast masses: From mammography to digital breast tomosynthesis," *Journal of biomedical informatics*, vol. 44, no. 5, pp. 815–823, 2011.

[112] S. A. Kostopoulos, P. A. Asvestas, I. K. Kalatzis, G. C. Sakellaropoulos, T. H. Sakkis, D. A. Cavouras, and D. T. Glotsos, "Adaptable pattern recognition system for discriminating melanocytic nevi from malignant melanomas using plain photography images from different image databases." *International Journal of Medical Informatics*, 2017.

[113] R. Llobet, J. C. Pérez-Cortés, A. H. Toselli, and A. Juan, "Computer-aided detection of prostate cancer," *International Journal of Medical Informatics*, vol. 76, no. 7, pp. 547–556, 2007.

[114] J. Mallik, A. Samal, and S. L. Gardner, "A content based image retrieval system for a biological specimen collection," *Computer Vision and Image Understanding*, vol. 114, no. 7, pp. 745–757, 2010.

[115] P. P. de San Roman, J. Benois-Pineau, J.-P. Domenger, F. Paclet, D. Cataert, and A. de Rugy, "Saliency driven object recognition in egocentric videos with deep cnn: toward application in assistance to neuroprostheses," *Computer Vision and Image Understanding*, 2017.

[116] C. Blaiotta, M. J. Cardoso, and J. Ashburner, "Variational inference for medical image segmentation," *Computer Vision and Image Understanding*, vol. 151, pp. 14–28, 2016.

[117] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.

[118] C. Strapparava, A. Valitutti *et al.*, "Wordnet affect: an affective extension of wordnet." in *LREC*, vol. 4. Citeseer, 2004, pp. 1083–1086.

[119] A. Esuli, "a.(2006). sentiwordnet: A publicly available lexical resource for opinion mining," *Proceedings of Language Resources And Evaluation (LREC). Genoa, Italy*, pp. 24–26.

[120] E. Cambria and B. White, "Jumping nlp curves: a review of natural language processing research [review article]," *IEEE Computational Intelligence Magazine*, vol. 9, no. 2, pp. 48–57, 2014.

[121] Y.-M. Li and T.-Y. Li, "Deriving market intelligence from microblogs," *Decision Support Systems*, vol. 55, no. 1, pp. 206–217, 2013.

[122] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[123] G. Mesnil, T. Mikolov, M. Ranzato, and Y. Bengio, "Ensemble of generative and discriminative techniques for sentiment analysis of movie reviews," *arXiv preprint arXiv:1412.5335*, 2014.

[124] T. Chen, D. Borth, T. Darrell, and S.-F. Chang, "Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks," *arXiv preprint arXiv:1410.8586*, 2014.

[125] C. Xu, S. Cetintas, K.-C. Lee, and L.-J. Li, "Visual sentiment prediction with deep convolutional neural networks," *arXiv preprint arXiv:1411.5731*, 2014.

[126] E. Cambria, S. Poria, F. Bisio, R. Bajpai, and I. Chaturvedi, "The clsa model: A novel framework for concept-level sentiment analysis," in *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, 2015, pp. 3–22.

[127] Q. You, J. Luo, H. Jin, and J. Yang, "Joint visual-textual sentiment analysis with deep neural networks," in *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 2015, pp. 1071–1074.

[128] Y. Yu, H. Lin, J. Meng, and Z. Zhao, "Visual and textual sentiment analysis of a microblog using deep convolutional neural networks," *Algorithms*, vol. 9, no. 2, p. 41, 2016.

[129] H. Kang, S. J. Yoo, and D. Han, "Senti-lexicon and improved naïve bayes algorithms for sentiment analysis of restaurant reviews," *Expert Systems with Applications*, vol. 39, no. 5, pp. 6000–6010, 2012.

[130] F. Xianghua, L. Guo, G. Yanyan, and W. Zhiqiang, "Multi-aspect sentiment analysis for chinese online social reviews based on topic modeling and hownet lexicon," *Knowledge-Based Systems*, vol. 37, pp. 186–195, 2013.

[131] K. Choi, K.-A. Toh, and H. Byun, "Incremental face recognition for large-scale social network services," *Pattern Recognition*, vol. 45, no. 8, pp. 2868–2883, 2012.

[132] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 513–520.

[133] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1631–1642.

[134] S. Poria, E. Cambria, and A. Gelbukh, "Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 2539–2544.

[135] E. Boiy and M.-F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," *Information retrieval*, vol. 12, no. 5, pp. 526–558, 2009.

[136] J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification," in *Proceedings of the ACL student research workshop*. Association for Computational Linguistics, 2005, pp. 43–48.

[137] H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010, pp. 291–300.

[138] M. Salathé and S. Khandelwal, "Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control," *PLoS computational biology*, vol. 7, no. 10, p. e1002199, 2011.

[139] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman *et al.*, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PloS one*, vol. 8, no. 9, p. e73791, 2013.

[140] Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," *Expert systems with applications*, vol. 36, no. 3, pp. 6527–6535, 2009.

[141] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern recognition letters*, vol. 34, no. 1, pp. 3–19, 2013.

[142] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, p. 29, 2013.

[143] K. McKenzie, M. A. Campbell, D. A. Scott, T. R. Discoll, J. E. Harrison, and R. J. McClure, "Identifying work related injuries: comparison of methods for interrogating text fields," *BMC medical informatics and decision making*, vol. 10, no. 1, p. 19, 2010.

[144] J. Albusac, J. J. Castro-Schez, L. M. López-López, D. Vallejo, and L. Jimenez-Linares, "A supervised learning approach to automate the acquisition of knowledge in surveillance systems," *Signal Processing*, vol. 89, no. 12, pp. 2400–2414, 2009.

[145] R. R. Sillito and R. B. Fisher, "Semi-supervised learning for anomalous trajectory detection." in *BMVC*, vol. 27, 2008, pp. 1025–1044.

[146] D. Makris and T. Ellis, "Learning semantic scene models from observing activity in visual surveillance," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 35, no. 3, pp. 397–408, 2005.

[147] B. Tamersoy and J. K. Aggarwal, "Robust vehicle detection for tracking in highway surveillance videos using unsupervised learning," in *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on.* Ieee, 2009, pp. 529–534.

[148] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3908–3916.

[149] X. Wu, G. Liang, K. K. Lee, and Y. Xu, "Crowd density estimation using texture analysis and learning," in *Robotics and Biomimetics, 2006. ROBIO'06. IEEE International Conference on.* IEEE, 2006, pp. 214–219.

[150] T. Ahmed, B. Oreshkin, and M. Coates, "Machine learning approaches to network anomaly detection," in *Proceedings of the 2nd USENIX workshop on Tackling computer systems problems with machine learning techniques*. USENIX Association, 2007, pp. 1–6.

[151] N. M. Oliver, B. Rosario, and A. P. Pentland, "A bayesian computer vision system for modeling human interactions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 831–843, 2000.

[152] G. Wu, Y. Wu, L. Jiao, Y.-F. Wang, and E. Y. Chang, "Multi-camera spatio-temporal fusion and biased sequence-data learning for security surveillance," in *Proceedings of the eleventh ACM international conference on Multimedia*. ACM, 2003, pp. 528–538.

[153] Z. Lin and L. S. Davis, "Learning pairwise dissimilarity profiles for appearance recognition in visual surveillance," in *International symposium on visual computing*. Springer, 2008, pp. 23–34.

[154] W. Ouyang and X. Wang, "Joint deep learning for pedestrian detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2056–2063.

[155] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential deep learning for human action recognition," in *International Workshop on Human Behavior Understanding*. Springer, 2011, pp. 29–39.

[156] W. Zajdel, Z. Zivkovic, and B. Krose, "Keeping track of humans: Have i seen this person before?" in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*. IEEE, 2005, pp. 2081–2086.

[157] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *European conference on computer vision*. Springer, 2014, pp. 1–16.

[158] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 34–39.

[159] R. Zhang, L. Lin, R. Zhang, W. Zuo, and L. Zhang, "Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4766–4779, 2015.

[160] C. Ramos, J. C. Augusto, and D. Shapiro, "Ambient intelligence—the next step for artificial intelligence," *IEEE Intelligent Systems*, vol. 23, no. 2, pp. 15–18, 2008.

[161] D. Liciotti, E. Frontoni, A. Mancini, and P. Zingaretti, "Pervasive system for consumer behaviour analysis in retail environments," in *International Workshop on Face and Facial Expression Recognition from Real World Videos*. Springer, 2016, pp. 12–23.

[162] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J.-C. Burgelman, *Scenarios for ambient intelligence in 2010*. Office for official publications of the European Communities Luxembourg, 2001.

[163] E. Frontoni, F. Marinelli, R. Rosetti, and P. Zingaretti, "Shelf space re-allocation for out of stock reduction," *Computers & Industrial Engineering*, vol. 106, pp. 32–40, 2017.

[164] D. Liciotti, P. Zingaretti, and V. Placidi, "An automatic analysis of shoppers behaviour using a distributed rgb-d cameras system," in *Mechatronic and Embedded Systems and Applications (MESA), 2014 IEEE/ASME 10th International Conference on*. IEEE, 2014, pp. 1–6.

[165] D. Liciotti, M. Contigiani, E. Frontoni, A. Mancini, P. Zingaretti, and V. Placidi, "Shopper analytics: A customer activity recognition system using a distributed rgb-d camera network," in *International Workshop on Video Analytics for Audience Measurement in Retail and Digital Signage*. Springer, 2014, pp. 146–157.

[166] E. Frontoni, A. Mancini, P. Zingaretti, and V. Placidi, "Information management for intelligent retail environment: the shelf detector system," *Information*, vol. 5, no. 2, pp. 255–271, 2014.

[167] S. R. Ahmed, "Applications of data mining in retail business," in *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on*, vol. 2. IEEE, 2004, pp. 455–459.

[168] A. Berson and S. J. Smith, *Building data mining applications for CRM*. McGraw-Hill, Inc., 2002.

[169] C. Giraud-Carrier and O. Povel, "Characterising data mining software," *Intelligent Data Analysis*, vol. 7, no. 3, pp. 181–192, 2003.

[170] S. Mitra, S. K. Pal, and P. Mitra, "Data mining in soft computing framework: a survey," *IEEE transactions on neural networks*, vol. 13, no. 1, pp. 3–14, 2002.

[171] R. S. Swift, *Accelerating customer relationships: Using CRM and relationship technologies*. Prentice Hall Professional, 2001.

[172] A. H. Kracklauer, D. Q. Mills, and D. Seifert, "Customer management as the origin of collaborative customer relationship management," in *Collaborative Customer Relationship Management*. Springer, 2004, pp. 3–6.

[173] J. Y. Woo, S. M. Bae, and S. C. Park, "Visualization method for customer targeting using customer map," *Expert Systems with Applications*, vol. 28, no. 4, pp. 763–772, 2005.

[174] K.-W. Cheung, J. T. Kwok, M. H. Law, and K.-C. Tsui, "Mining customer product ratings for personalized marketing," *Decision Support Systems*, vol. 35, no. 2, pp. 231–243, 2003.

[175] M.-C. Chen, A.-L. Chiu, and H.-H. Chang, "Mining changes in customer behavior in retail marketing," *Expert Systems with Applications*, vol. 28, no. 4, pp. 773–781, 2005.

[176] J. H. Drew, D. Mani, A. L. Betz, and P. Datta, "Targeting customers with statistical and data-mining techniques," *Journal of Service Research*, vol. 3, no. 3, pp. 205–219, 2001.

[177] A. Prinzie and D. Van den Poel, "Investigating purchasing-sequence patterns for financial services using markov, mtd and mtdg models," *European Journal of Operational Research*, vol. 170, no. 3, pp. 710–734, 2006.

[178] C. C. Aggarwal, C. Procopiuc, and P. S. Yu, "Finding localized associations in market basket data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 1, pp. 51–62, 2002.

[179] W.-H. Au and K. C. Chan, "Mining fuzzy association rules in a bank-account database," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 2, pp. 238–248, 2003.

[180] E. W. Ngai, L. Xiu, and D. C. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert systems with applications*, vol. 36, no. 2, pp. 2592–2602, 2009.

[181] W. A. Mackaness and O. Z. Chaudhry, "Automatic classification of retail spaces from a large scale topographic database," *Transactions in GIS*, vol. 15, no. 3, pp. 291–307, 2011.

[182] T. Jambulingam, R. Kathuria, and W. R. Doucette, "Entrepreneurial orientation as a basis for classification within a service industry: the case

of retail pharmacy industry," *Journal of operations management*, vol. 23, no. 1, pp. 23–42, 2005.

[183] A. Kumar, G. Kabra, E. K. Mussada, M. K. Dash, and P. S. Rana, "Combined artificial bee colony algorithm and machine learning techniques for prediction of online consumer repurchase intention," *Neural Computing and Applications*, pp. 1–14, 2017.

[184] J. Lismont, S. Ram, B. Baesens, W. Lemahieu, and J. Vanthienen, "The (pseudo-) social behavior of products in offline retail stores: predicting increase in product interpurchase time," 2018.

[185] Y. Song, Y. Xue, C. Li, X. Zhao, S. Liu, X. Zhuo, K. Zhang, B. Yan, X. Ning, Y. Wang *et al.*, "Online cost efficient customer recognition system for retail analytics," in *Applications of Computer Vision Workshops (WACVW), 2017 IEEE Winter*. IEEE, 2017, pp. 9–16.

[186] T. Qu, J. Zhang, F. T. Chan, R. Srivastava, M. Tiwari, and W.-Y. Park, "Demand prediction and price optimization for semi-luxury supermarket segment," *Computers & Industrial Engineering*, 2017.

[187] V. S. Tallapragada, N. A. Rao, and S. Kanapala, "Emometric: An iot integrated big data analytic system for real time retail customer's emotion tracking and analysis," *International Journal of Computational Intelligence Research*, vol. 13, no. 5, pp. 673–695, 2017.

[188] E. Abrams, G. Gui, and A. Hortacsu, "Finding exogenous variation in data," *arXiv preprint arXiv:1704.07787*, 2017.

[189] S. Ma and R. Fildes, "A retail store sku promotions optimization model for category multi-period profit maximization," *European Journal of Operational Research*, vol. 260, no. 2, pp. 680–692, 2017.

[190] N. Kalaiselvi, K. Aravind, S. Balaguru, and V. Vijayaragul, "Retail price analytics using backpropogation neural network and sentimental analysis," in *Signal Processing, Communication and Networking (ICSCN), 2017 Fourth International Conference on*. IEEE, 2017, pp. 1–6.

[191] A. Yasser, K. Clawson, C. Bowerman, M. Lévêque *et al.*, "Saving cultural heritage with digital make-believe: Machine learning and digital techniques to the rescue," 2017.

[192] J. Llamas, P. M. Lerones, E. Zalama, and J. Gómez-García-Bermejo, "Applying deep learning techniques to cultural heritage images within the inception project," in *Euro-Mediterranean Conference*. Springer, 2016, pp. 25–32.

[193] T. E. Lombardi, *Classification of Style in Fine-art Painting*. Pace University, 2005.

[194] R. S. Arora and A. Elgammal, "Towards automated classification of fine-art painting style: A comparative study," in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 3541–3544.

[195] K. Barnard, P. Duygulu, and D. Forsyth, "Clustering art," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 2. IEEE, 2001, pp. II–II.

[196] H. Qi, A. Taeb, and S. M. Hughes, "Visual stylometry using background selection and wavelet-hmt-based fisher information distances for attribution and dating of impressionist paintings," *Signal Processing*, vol. 93, no. 3, pp. 541–553, 2013.

[197] A. Doulamis and T. Varvarigou, "Image analysis for artistic style identification: A powerful tool for preserving cultural heritage," *Emerging Technologies in Non-Destructive Testing V*, p. 71, 2012.

[198] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller, "Recognizing image style," *arXiv preprint arXiv:1311.3715*, 2013.

[199] R. M. Anwer, F. S. Khan, J. van de Weijer, and J. Laaksonen, "Combining holistic and part-based deep representations for computational painting categorization," in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*. ACM, 2016, pp. 339–342.

[200] C. Grana, D. Borghesani, and R. Cucchiara, "Automatic segmentation of digitalized historical manuscripts," *Multimedia Tools and Applications*, vol. 55, no. 3, pp. 483–506, 2011.

[201] Y. Bar, N. Levy, and L. Wolf, "Classification of artistic styles using binarized features derived from a deep neural network." in *ECCV Workshops (1)*, 2014, pp. 71–84.

[202] A. Salvador, M. Zeppelzauer, D. Manchon-Vizuete, A. Calafell, and X. Giro-i Nieto, "Cultural event recognition with visual convnets and temporal models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 36–44.

[203] N. Van Noord, E. Hendriks, and E. Postma, "Toward discovery of the artist's style: Learning to recognize artists by their artworks," *IEEE Signal Processing Magazine*, vol. 32, no. 4, pp. 46–54, 2015.

[204] S. Jafarpour, G. Polatkan, E. Brevdo, S. Hughes, A. Brasoveanu, and I. Daubechies, "Stylistic analysis of paintings usingwavelets and machine learning," in *Signal Processing Conference, 2009 17th European*. IEEE, 2009, pp. 1220–1224.

[205] M. Tissenbaum, M. Berland, and V. Kumar, "Modeling visitor behavior in a game-based engineering museum exhibit with hidden markov models." in *EDM*, 2016, pp. 517–522.

[206] G. Polatkan, S. Jafarpour, A. Brasoveanu, S. Hughes, and I. Daubechies, "Detection of forgery in paintings using supervised learning," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2009, pp. 2921–2924.

[207] J. Zujovic, L. Gandy, S. Friedman, B. Pardo, and T. N. Pappas, "Classifying paintings by artistic genre: An analysis of features & classifiers," in *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on*. IEEE, 2009, pp. 1–5.

[208] A. Blessing and K. Wen, "Using machine learning for identification of art paintings," *Technical report*, 2010.

[209] J. Li and J. Z. Wang, "Studying digital imagery of ancient paintings by mixtures of stochastic models," *IEEE Transactions on Image Processing*, vol. 13, no. 3, pp. 340–353, 2004.

[210] M. Paolanti, D. Liciotti, E. Frontoni, P. Zingaretti, and O. Carnevali, "Automatic classification of human oocytes for assisted reproductive technology," *To be submitted Patent Pending*, 2017.

[211] M. Paolanti, E. Frontoni, P. Zingaretti, and O. Carnevali, "Decision support system for assisted reproductive," *To be submitted Patent Pending*, 2017.

[212] M. Paolanti, C. Kaiser, R. Schallner, E. Frontoni, and P. Zingaretti, "Visual and textual sentiment analysis of brand-related social media pictures using deep convolutional neural networks," in *International Conference on Image Analysis and Processing*. Springer, 2017, pp. 402–413.

[213] D. Liciotti, M. Paolanti, E. Frontoni, A. Mancini, and P. Zingaretti, "Person re-identification dataset with rgb-d camera in a top-view configuration," in *International Workshop on Face and Facial Expression Recognition from Real World Videos*. Springer, 2016, pp. 1–11.

[214] D. Paolanti, MarinaLiciotti, A. Cenci, E. Frontoni, and P. Zingaretti, "Person re-identification with rgb-d camera in a top-view configuration," in *Submitted to Pattern Recognition Letters*.

[215] M. Paolanti, D. Liciotti, R. Pietrini, A. Mancini, and E. Frontoni, "Modelling and forecasting customer navigation in intelligent retail environments," *Journal of Intelligent & Robotic Systems*, 2017.

[216] D. Liciotti, M. Paolanti, E. Frontoni, and P. Zingaretti, "Trajectory clustering in intelligent retail environment," *Submitted to Pattern Recognition Letters*, 2017.

[217] S. Naspetti, R. Pierdicca, S. Mandolesi, M. Paolanti, E. Frontoni, and R. Zanoli, "Automatic analysis of eye-tracking data for augmented reality applications: A prospective outlook," in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 2016, pp. 217–230.

[218] M. Paolanti, R. Pierdicca, D. Liciotti, S. Naspetti, R. Zanoli, S. Mandolesi, E. Frontoni, and P. Zingaretti, "An hmm-based approach to eye-tracking data for augmented reality applications," in *Submitted to Journal on Computing and Cultural Heritage*.

[219] Y. Peng, Z. Wu, and J. Jiang, "A novel feature selection approach for biomedical data classification," *Journal of Biomedical Informatics*, vol. 43, no. 1, pp. 15–23, 2010.

[220] M. Hasan, A. Kotov, A. I. Carcone, M. Dong, S. Naar, and K. B. Hartlieb, "A study of the effectiveness of machine learning methods for classification of clinical interview fragments into a large number of categories," *Journal of biomedical informatics*, vol. 62, pp. 21–31, 2016.

[221] Z. Mao, W. Cai, and X. Shao, "Selecting significant genes by randomization test for cancer classification using gene expression data," *Journal of biomedical informatics*, vol. 46, no. 4, pp. 594–601, 2013.

[222] J. Cao, L. Zhang, B. Wang, F. Li, and J. Yang, "A fast gene selection method for multi-cancer classification using multiple support vector data description," *Journal of biomedical informatics*, vol. 53, pp. 381–389, 2015.

[223] E. Giorgini, G. Gioacchini, S. Sabbatini, C. Conti, L. Vaccari, A. Borini, O. Carnevali, and G. Tosi, "Vibrational characterization of female gametes: a comparative study," *Analyst*, vol. 139, no. 20, pp. 5049–5060, 2014.

[224] M. M. Matzuk, K. H. Burns, M. M. Viveiros, and J. J. Eppig, "Intercellular communication in the mammalian ovary: oocytes carry the conversation," *Science*, vol. 296, no. 5576, pp. 2178–2180, 2002.

[225] K. J. Hutt and D. F. Albertini, "An oocentric view of folliculogenesis and embryogenesis," *Reproductive biomedicine online*, vol. 14, no. 6, pp. 758–764, 2007.

[226] F. Guerif, M. Lemseffer, J. Leger, R. Bidault, V. Cadoret, C. Chavez, O. Gasnier, M. Saussereau, and D. Royère, "Does early morphology provide additional selection power to blastocyst selection for transfer?" *Reproductive biomedicine online*, vol. 21, no. 4, pp. 510–519, 2010.

[227] B. Balaban and B. Urman, "Effect of oocyte morphology on embryo development and implantation," *Reproductive biomedicine online*, vol. 12, no. 5, pp. 608–615, 2006.

[228] K. Nygren and A. Nyboe Andersen, "Assisted reproductive technology in europe, 2001. results generated european register by eshre," *Hum. Reprod*, vol. 20, pp. 1158–1176, 2002.

[229] E. Giorgini, C. Conti, R. Rocchetti, C. Rubini, S. Sabbatini, V. Librando, and G. Tosi, "Study of oral cavity lesions by infrared spectroscopy." *Journal of biological regulators and homeostatic agents*, vol. 30, no. 1, pp. 309–314, 2015.

[230] E. Giorgini, C. Conti, P. Ferraris, S. Sabbatini, G. Tosi, C. Rubini, L. Vaccari, G. Gioacchini, and O. Carnevali, "Effects of lactobacillus rhamnosus on zebrafish oocyte maturation: an ftir imaging and biochemical analysis," *Analytical and bioanalytical chemistry*, vol. 398, no. 7-8, pp. 3063–3072, 2010.

[231] M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys *et al.*, "Using fourier transform ir spectroscopy to analyze biological materials," *Nature protocols*, vol. 9, no. 8, pp. 1771–1791, 2014.

[232] E. Frontoni, M. Baldi, P. Zingaretti, V. Landro, and P. Misericordia, "Security issues for data sharing and service interoperability in ehealth systems: the nu. sa. test bed," in *2014 International Carnahan Conference on Security Technology (ICCST)*. IEEE, 2014, pp. 1–6.

[233] L. Li, C. R. Weinberg, T. A. Darden, and L. G. Pedersen, "Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the ga/knn method," *Bioinformatics*, vol. 17, no. 12, pp. 1131–1142, 2001.

[234] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[235] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, no. 2, 1995, pp. 1137–1145.

[236] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, no. 2-3, pp. 271–274, 1998.

[237] R. W. Carolin Kaiser, "Gaining marketing-relevant knowledge from social media photos - a picture is worth a thousand words." in *Proceedings of the 2016 ESOMAR Congress, New Orleans*, 2016.

[238] Y. Yang, J. Jia, S. Zhang, B. Wu, Q. Chen, J. Li, C. Xing, and J. Tang, "How do your friends on social media disclose your emotions?" in *AAAI*, vol. 14, 2014, pp. 1–7.

[239] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks," *arXiv preprint arXiv:1509.06041*, 2015.

[240] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques," in *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on.* IEEE, 2003, pp. 427–434.

[241] S. Mukherjee and P. Bhattacharyya, "Feature specific sentiment analysis for product reviews," *Computational Linguistics and Intelligent Text Processing*, pp. 475–487, 2012.

[242] B. Pang, L. Lee *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[243] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2544–2558, 2010.

[244] J. Yuan, S. Mcdonough, Q. You, and J. Luo, "Sentribute: image sentiment analysis from a mid-level perspective," in *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining.* ACM, 2013, p. 10.

[245] Y. Chang, L. Tang, Y. Inagaki, and Y. Liu, "What is tumblr: A statistical overview and comparison," *ACM SIGKDD Explorations Newsletter*, vol. 16, no. 1, pp. 21–29, 2014.

[246] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[247] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Advances in neural information processing systems*, 2015, pp. 649–657.

[248] M. Paolanti, E. Frontoni, A. Mancini, R. Pierdicca, and P. Zingaretti, "Automatic classification for anti mixup events in advanced manufacturing system," in *ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, 2015, pp. V009T07A061–V009T07A061.

[249] M. Liao, B. Shi, X. Bai, X. Wang, and W. Liu, "Textboxes: A fast text detector with a single deep neural network," *arXiv preprint arXiv:1611.06779*, 2016.

[250] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading text in the wild with convolutional neural networks," *International Journal of Computer Vision*, vol. 116, no. 1, pp. 1–20, 2016.

[251] P. K. Bhowmick, P. Mitra, and A. Basu, "An agreement measure for determining inter-annotator reliability of human judgements on affective text," in *Proceedings of the Workshop on Human Judgements in Computational Linguistics*. Association for Computational Linguistics, 2008, pp. 58–65.

[252] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[253] C. Chahla, H. Snoussi, F. Abdallah, and F. Dornaika, "Discriminant quaternion local binary pattern embedding for person re-identification through prototype formation and color categorization," *Engineering Applications of Artificial Intelligence*, vol. 58, pp. 27–33, 2017.

[254] W. Hariri, H. Tabia, N. Farah, A. Benouareth, and D. Declercq, "3d facial expression recognition using kernel methods on riemannian manifold," *Engineering Applications of Artificial Intelligence*, vol. 64, pp. 25–32, 2017.

[255] D. Baltieri, R. Vezzani, and R. Cucchiara, "3d body model construction and matching for real time people re-identification." in *Eurographics Italian Chapter Conference*, 2010, pp. 65–71.

[256] B. Farou, M. N. Kouahla, H. Seridi, and H. Akdag, "Efficient local monitoring approach for the task of background subtraction," *Engineering Applications of Artificial Intelligence*, vol. 64, pp. 1–12, 2017.

[257] G. Lisanti, I. Masi, A. D. Bagdanov, and A. Del Bimbo, "Person re-identification by iterative re-weighted sparse ranking," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 8, pp. 1629–1642, 2015.

[258] D. Liciotti, G. Massi, E. Frontoni, A. Mancini, and P. Zingaretti, "Human activity analysis for in-home fall risk assessment," in *Communication Workshop (ICCW), 2015 IEEE International Conference on.* IEEE, 2015, pp. 284–289.

[259] S. Gong, M. Cristani, S. Yan, and C. C. Loy, *Person re-identification.* Springer, 2014, vol. 1.

[260] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, vol. 3, no. 5. Citeseer, 2007.

[261] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," in *European Conference on Computer Vision.* Springer, 2014, pp. 688–703.

[262] A. Ess, B. Leibe, and L. V. Gool, "Depth and appearance for mobile scene analysis," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on.* IEEE, 2007, pp. 1–8.

[263] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification." in *BMVC*, vol. 1, no. 2, 2011, p. 6.

[264] I. B. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, and V. Murino, "Re-identification with rgb-d sensors," in *Computer Vision–ECCV 2012. Workshops and Demonstrations.* Springer, 2012, pp. 433–442.

[265] M. Sturari, D. Liciotti, R. Pierdicca, E. Frontoni, A. Mancini, M. Contigiani, and P. Zingaretti, "Robust and affordable retail customer profiling by vision and radio beacon sensor fusion," *Pattern Recognition Letters*, vol. 81, pp. 30–40, 2016.

[266] D. Baltieri, R. Vezzani, and R. Cucchiara, "Learning articulated body models for people re-identification," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 557–560.

[267] D. Liciotti, M. Paolanti, E. Frontoni, A. Mancini, and P. Zingaretti, "Person re-identification dataset with rgb-d camera in a top-view configuration," in *Video Analytics for Face, Face Expression Recognition, and Audience Measurement*. Springer, 2017.

[268] A. J. Newman and G. R. Foxall, "In-store customer behaviour in the fashion sector: some emerging methodological and theoretical directions," *International Journal of Retail & Distribution Management*, vol. 31, no. 11, pp. 591–600, 2003.

[269] A. Marin-Hernandez, G. de Jesús Hoyos-Rivera, M. Garcia-Arroyo, and L. F. Marin-Urias, "Conception and implementation of a supermarket shopping assistant system," in *Artificial Intelligence (MICAI), 2012 11th Mexican International Conference on*. IEEE, 2012, pp. 26–31.

[270] M. Contigiani, R. Pietrini, A. Mancini, and P. Zingaretti, "Implementation of a tracking system based on uwb technology in a retail environment," in *Mechatronic and Embedded Systems and Applications (MESA), 2016 12th IEEE/ASME International Conference on*. IEEE, 2016, pp. 1–6.

[271] G. Rzevski, "On conceptual design of intelligent mechatronic systems," *Mechatronics*, vol. 13, no. 10, pp. 1029–1044, 2003.

[272] C. Coppola, T. Krajník, T. Duckett, N. Bellotto *et al.*, "Learning temporal context for activity recognition." in *ECAI*, 2016, pp. 107–115.

[273] D. Liciotti, E. Frontoni, P. Zingaretti, N. Bellotto, and T. Duckett, "Hmm-based activity recognition with a ceiling rgb-d camera." in *ICPRAM*, 2017, pp. 567–574.

[274] J.-G. Lee, J. Han, and K.-Y. Whang, "Trajectory clustering: a partition-and-group framework," in *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*. ACM, 2007, pp. 593–604.

[275] D. Buzan, S. Sclaroff, and G. Kollios, "Extraction and clustering of motion trajectories in video," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2. IEEE, 2004, pp. 521–524.

[276] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in neural information processing systems*, 2002, pp. 849–856.

[277] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Advances in neural information processing systems*, 2005, pp. 1601–1608.

[278] J. S. Larson, E. T. Bradlow, and P. S. Fader, "An exploratory look at supermarket shopping paths," *International Journal of research in Marketing*, vol. 22, no. 4, pp. 395–414, 2005.

[279] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[280] G. D. Forney, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

[281] S. J. Julier and J. K. Uhlmann, "A new extension of the kalman filter to nonlinear systems," in *Int. symp. aerospace/defense sensing, simul. and controls*, vol. 3, no. 26. Orlando, FL, 1997, pp. 182–193.

[282] S. Atev, G. Miller, and N. P. Papanikolopoulos, "Clustering of vehicle trajectories," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 647–657, 2010.

[283] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *The annals of mathematical statistics*, vol. 41, no. 1, pp. 164–171, 1970.

[284] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation," in *Australasian Joint Conference on Artificial Intelligence*. Springer, 2006, pp. 1015–1021.

[285] F. Cameron and S. Kenderdine, "Theorizing digital cultural heritage: A critical discourse," 2007.

[286] S. Alletto, R. Cucchiara, G. Del Fiore, L. Mainetti, V. Mighali, L. Patrono, and G. Serra, "An indoor location-aware system for an iot-based smart museum," *IEEE Internet of Things Journal*, vol. 3, no. 2, pp. 244–253, 2016.

[287] K. Eghbal-Azar, M. Merkt, J. Bahnmueller, and S. Schwan, "Use of digital guides in museum galleries: Determinants of information selection," *Computers in Human Behavior*, vol. 57, pp. 133–142, 2016.

[288] S. Pescarin, A. Pagano, M. Wallergård, W. Hupperetz, and C. Ray, "Archeovirtual 2011: An evaluation approach to virtual museums," in *Virtual Systems and Multimedia (VSMM), 2012 18th International Conference on*. IEEE, 2012, pp. 25–32.

[289] P. Clini, E. Frontoni, R. Quattrini, and R. Pierdicca, "Augmented reality experience: From high-resolution acquisition to real time augmented contents," *Advances in Multimedia*, vol. 2014, p. 18, 2014.

[290] R. Pierdicca, E. Frontoni, P. Zingaretti, E. S. Malinverni, F. Colosi, and R. Orazi, "Making visible the invisible. augmented reality visualization for 3d reconstructions of archaeological sites," in *International Conference on Augmented and Virtual Reality*. Springer, 2015, pp. 25–37.

[291] M. Sturari, P. Clini, and R. Quattrini, "Advanced interaction with paintings by augmented reality and high resolution visualization: A real case exhibition," in *Augmented and Virtual Reality: Second International Conference, AVR 2015, Lecce, Italy, August 31-September 3, 2015, Proceedings*, vol. 9254. Springer, 2015, p. 38.

[292] A. Poole and L. J. Ball, "Eye tracking in hci and usability research," *Encyclopedia of human computer interaction*, vol. 1, pp. 211–219, 2006.

[293] S. W. Shi, M. Wedel, and F. Pieters, "Information acquisition during online decision making: A model-based exploration using eye-tracking data," *Management Science*, vol. 59, no. 5, pp. 1009–1026, 2013.

[294] Q.-X. Qu, L. Zhang, W.-Y. Chao, and V. Duffy, "User experience design based on eye-tracking technology: A case study on smartphone apps," in *Advances in Applied Digital Human Modeling and Simulation*. Springer, 2017, pp. 303–315.

[295] M. Wedel and R. Pieters, "A review of eye-tracking research in marketing," in *Review of marketing research*. Emerald Group Publishing Limited, 2008, pp. 123–147.

[296] A. Gegenfurtner, E. Lehtinen, and R. Säljö, "Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains," *Educational Psychology Review*, vol. 23, no. 4, pp. 523–552, 2011.

[297] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.

[298] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 9, pp. 970–982, 2000.

[299] A. T. Duchowski, J. Driver, S. Jolaoso, W. Tan, B. N. Ramey, and A. Robbins, "Scanpath comparison revisited," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM, 2010, pp. 219–226.

[300] G. Andrienko, N. Andrienko, M. Burch, and D. Weiskopf, "Visual analytics methodology for eye movement studies," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2889–2898, 2012.