UNIVERSITÀ DEGLI STUDI DI PAVIA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Gaze-Based Biometrics: Some Case Studies

Supervisor

Prof. Virginio Cantoni

Ph.D. Dissertation by

Haochen Wang

Dottorato di Ricerca in Ingegneria Elettronica, Informatica ed Elettrica

XXX Ciclo

Abstract

Biometrics is the science of establishing the identity of a person based on physical or behavioral attributes. It is one of the key issues and one of the most widely used technologies for individual recognition tasks as a supplement to traditional methods. Eye tracking is the technology used to track and record users' gaze behavior through the detection of eye data. Biometrics based on eye tracking, i.e. "gaze-based biometrics", is an innovative approach to perform individual recognition and verification tasks through the analysis of eye features and behavior.

This thesis starts by reviewing biometric techniques, eye tracking technologies, and works on gaze-based biometrics carried out in the last decade, as well as the relevant theories and principles of Machine Learning. Four case studies are then considered to investigate the feasibility and performance of gaze-based biometrics in different scenarios: doorbell names inspection for building access, static image observation, moving target observation, and eye-driven soft PIN input.

In each of these case studies, both identification (i.e., recognizing people within a group) and authentication (i.e., verifying the claimed identities of people within a group) are analyzed.

In the case study on gaze-based biometrics for building access, the "familiarity" of individuals to the stimuli is also investigated (i.e., we try to distinguish first-time visitors from frequent visitors). The obtained classification accuracies, using different classifiers, are almost always higher than 0.7. This shows that gaze-based techniques are viable solutions in the context of "soft biometrics" (which does not need to achieve very high accuracies because it is used together with traditional methods, such as those exploiting passwords or PINs, or with extremely reliable biometric traits, such as fingerprints).

CONTENTS

Chapter 1 *Introduction*

Nowadays, with the development of Human-Computer Interaction technologies and portable intelligent devices, people are paying increasing attention to information security topics, while individual recognition is an essential issue in the information security field.

Biometrics is one of the key issues and one of the most widely used technologies for individual recognition tasks as a supplement to traditional methods like the PIN code. It is the science of establishing the identity of a person based on physical or behavioral attributes [1]. The most typical applications of biometrics are *identification* and *authentication*. Identification is to find whether the person belongs to a specific already-known group. Authentication is to judge whether the person is the individual who he or she claims to be. It can be viewed as a pattern recognition problem, where the machine learns the salient features (i.e. the "patterns") in the biometric attributes of an individual and robustly matches such patterns, efficiently and effectively. The core technology to solve the pattern recognition problems is based on Machine Learning.

Machine Learning is a subset of Artificial Intelligence that uses statistical techniques to give computers the ability to "learn" with data without being explicitly programmed [2]. It is an interdisciplinary field which involves probability theory, statistics, approximation theory, convex analysis, the theory of algorithmic complexity, etc. Learning can be divided into "Supervised Learning" and "Unsupervised Learning". Supervised Learning is obtained with output labels or "goals" by which the computer will learn a general rule that maps inputs to outputs. It mainly solves classification tasks. Unsupervised Learning is used to discover hidden patterns in data without providing labels or targets to the learning algorithms. Biometric identification and authentication problems usually belong to the category of Supervised Learning.

Traditional biometrics exploits human features such as fingerprints, face, iris, and voice. Eye tracking is the technology used to track and record users' eye behavior, through the detection of eye data such as fixations, saccades, blinks, etc. In the past decade, with the progress of eye tracking technologies, gaze-based features have emerged as a novel potential soft biometrics which can be applied to human-computer interaction tasks and individual recognition.

Some works during the past decade have demonstrated that eye tracking technology can be exploited to accomplish pattern recognition tasks such as gender recognition, age recognition, and some assistance applications in the e-learning field, such as the detection of cognitive state and level of interest. Moreover, some biometric applications of eye tracking have been proposed as well.

In my Ph.D. activity, I overviewed the theories at the basis of gaze-based biometrics, studied machine learning principles, learnt and compared some typical machine learning methods, and exploited them to support my research topic. To accomplish my thesis, I designed four case studies and conducted experiments, collecting data and carrying out their analysis. The four case studies which were carried out in this research were about gaze-based biometrics for, respectively, building access through doorbell names inspection, static image observation, moving target observation, and eye-driven PIN input. Each case study contains two analysis stages, i.e. identification and authentication. In each stage, at least six classifiers were employed.

Through the above case studies, we achieved a highest classification accuracy of 0.935, with most accuracies higher than 0.7. Therefore, gaze-based biometrics through observation activities seems a viable solution for soft biometrics (which does not require very high accuracies because it is used along with other techniques). The accuracy can be also improved through data refinement, subdividing the raw data more precisely, increasing the instances of data samples, and selecting proper features by means of feature ranking procedures.

The following is the outline of the thesis:

- **Chapter 2: Biometrics**. Introduces the theoretical concepts about biometrics, including the definition and history of some typical fields of application of biometric technologies.

- **Chapter 3: Eye Tracking**. Provides a general presentation of eye tracking, including the history of the development of eye tracking technology and devices, i.e. eye trackers.
- **Chapter 4: Gaze-Based Biometrics.** Presents an overview of representative works about gaze-based biometrics in the past decade.
- **Chapter 5: Machine Learning**. Gives an introduction to Machine Learning, including basic Machine Learning technology and the mathematical principles of some typical methods which have been applied in this thesis.
- **Chapter 6: Case Study on Gaze-based Biometrics for Building Access Security**. Presents the research carried out for building access security. Experiments were carried out with 45 participants to collect gaze data and the analysis was conducted to evaluate familiarity classification, person identification and person authentication.
- **Chapter 7: Case Study on Gaze-based Biometrics through Static Images Observation**. Presents a research on gaze-based biometrics through a case study in which 40 participants were asked to freely observe some randomly shown static images. The analysis was carried out for both identification and verification.
- **Chapter8: Case Study on Gaze-based Biometrics through Moving Target Observation**. Presents a research on gaze-based biometrics through a case study in which 42 testers were asked to observe randomly moving graphic targets. The analysis was carried out for person identification and person verification.
- **Chapter9: Case Study on Gaze-based Biometrics through eye-driven PIN Input**. Presents a research on gaze-based biometrics in which 45 volunteers were asked to input PINs by looking at the keys of a virtual on-screen numeric pad. The analysis was carried out for both identification and verification.

Chapter10: Conclusions. Concludes the thesis with a critical analysis of the results obtained in this thesis, and indicates some suggestions about the relevant future work.

Chapter 2 *Biometrics*

2.1 Introduction

In modern society, with the progress of technologies and the increasing demand for improved public security, the ability to reliably identify individuals in real-time has become a fundamental requirement for many application fields, including forensic, international border crossing, financial transactions, etc. [3]. Traditionally, the logging-in method that requires username and password is adopted to solve the individual identity recognition task. However, this method has the potential risk of being wiretapped or hacked [4]. Biometric technologies provide users with friendly and reliable methods for accessing computer systems, networks, and workplaces [5][6][7]. Moreover, compared to most traditional authorization systems such as PIN code, password, or ID card, biometric traits have the advantage that they cannot be lost, forgotten, guessed, or easily cloned [8].

The term “biometric” is derived from the Greek words “bio” (life) and “metric” (measure) [9]. Biometrics offers natural and reliable solutions to certain aspects of identity check management by utilizing fully automated or semi-automated schemes to recognize individuals based on their biological characteristics [10]. Using biometrics makes it possible to establish an identity based on "who you are" rather than "what you possess" (such as PIN code, passwords, or ID card) [11].

A biometric system is essentially a pattern recognition system that acquires biometric data from an individual, extracts a salient feature set from the data, compares this feature set against the feature set(s) stored in the database, and executes an action based on the result of the comparison [12].

Usually, a biometric system is involved with two types of tasks: *identification* and *verification* [13]. In the identification phase, the user's biometric input is compared with the templates of all the people enrolled in the database and the system outputs either the identity of the person whose template has the highest degree of similarity with the user's input or a decision indicating that the user presenting the input is not an enrolled user [14]. In the verification phase, the user claims an identity. The biometric input is compared only to the template corresponding to the claimed identity (a one-to-one match). If the user's input and the template of the claimed identity have a high degree of similarity, then the claim is accepted as "genuine". Otherwise, the claim is rejected and the user is considered an "impostor". This way, the system answers the question "Are you the one which you claim to be?" [15].

2.2 Taxonomy of biometrics

Biometrics can be classified into two types: *physiological biometrics* and *behavioral biometrics* [16].

Physiological biometrics exploits the characteristics that an individual is born with, such as iris, hand, fingerprints, face, or DNA. In these cases the identification of a subject is carried out through genetic features [17].

Figure 1 Examples of some typical physiological biometrics [18][19][20][21]

Behavioral attributes establish identity based on the analysis of the way humans “do things”, such as key-stroke, signature, voice, etc. Based on the type of information about the user being collected, behavioral biometrics can be classified into five categories as follows [22].

The first category one is made up of authorship-based biometrics. It is based on examining a piece of text or a drawing produced by a tester. Verification is carried out by observing style peculiarities typical to the author of the work being examined, such as the used vocabulary, punctuation or brush strokes.

The second category involves HCI- features [23], and can be further subdivided into two groups. The first group consists of human interaction with input devices, such as keyboards, computer mice, and haptics which can register inherent, distinctive and consistent muscle actions [24]. The second group consists of HCI-based behavioral biometrics which measure advanced human behavior such as strategy, knowledge or skill exhibited by the user during the interaction with different software.

The third category uses indirect HCI-related biometrics obtained by indirectly monitoring user's HCI behaviors through observable low-level actions [25]. These include system call traces [26], audit logs [27], program execution traces [28], registry access [29], storage activity [30], call-stack data analysis [31], and system calls [32][33].

The fourth category relies on the users' motor-skills [34]. Motor-skill is the ability of human beings to utilize muscles, which, in most cases, is not inherited but learnt.

The fifth category consists of purely behavioral biometrics which measure human behavior directly, not concentrating on measurements of body parts or intrinsic, inimitable and lasting muscle actions such as the way an individual walks, types or even grips a tool [35]. For example, human beings utilize different strategies, skills and knowledge during performance of mentally demanding tasks. Purely behavioral biometrics quantifies such behavioral traits and makes successful identity verification a possibility.

Figure 2 Examples of some typical behavioral biometrics [36][37][38][39]

2.3 Performance of biometrics

The results of the output in a biometric system include two situations: *positive recognition*, meaning that the individual is recognized as "true" by the system, and the individual can be indeed the true target or actually an incorrectly recognized imposter; and *negative recognition*, meaning that the individual is recognized as "false" by the system and is considered an imposter.

To decide whether a system is “good enough”, it needs to achieve a certain level of performance. The computation of the performance of a biometric system aims to provide numerical measures on efficiency [40]. The performance of a biometric system is measured by the following typical statistical metrics [41]:

- **TP** (True Positive): number of correct positive recognitions, i.e. the number of subjects who match their claimed identities and are admitted correctly by the system;

- **FP** (False Positive): number of wrong positive recognitions, i.e. the number of subjects who do not match their claimed identities but are incorrectly admitted by the system;
- **TN** (True Negative): number of correct negative recognitions, i.e. number of subjects who do not match their claimed identities and are correctly not admitted by the system;
- **FN** (False Negative): number of wrong negative recognitions, i.e. number of subjects who should match their claimed identities but are incorrectly not admitted by the system;
- **FRR** (False Rejection Rate) = FNR (False Negative Rate) = $\frac{FN}{FN+TP}$;
- **FAR** (False Acceptance Rate) = FPR (False Positive Rate) = $\frac{FP}{FP+TN}$;
- **Sensitivity** = TPR (True Positive Rate) = $\frac{TP}{TP+FN}$; its value indicates the ability to recognize the correct individual and is maximum (= 1) when FRR = FN = 0;
- Specificity = TNR (True Negative Rate) = $\frac{TN}{TN+FP}$; its value indicates the ability to recognize the impostor and is maximum (= 1) when FAR = FP = 0;

Accuracy (biometric system's accuracy) = $\frac{TP+TN}{TP+FP+TN+FN}$.

Chapter 3 *Eye Tracking*

3.1 Physiology and Visual Function of the Human Eye

The eye is the organ for vision. The light of external objects enters the eyes and images are focused on the retina. On the retina, the light is converted into nerve impulses and transmitted to the brain by the optical nerve, which generates vision.

3.1.1 Anatomy of the Human Eye

The human eye is approximately spherical, with the distance of the front and posterior pole being about 24mm. As is shown in Figure 3.1, from inward to outward, the eyeball "wall" consists of three layers of membranes: the *fibrous membrane*, the *vascular membrane*, and the *retina*. The front 1/6 part of the fibrous membrane is the transparent *cornea*, whose curvature is higher than the other parts and which serves the major role in light focusing. The posterior 5/6 ivory-white opaque part is the *sclera*. The posterior pole of the sclera is pierced by the optic nerve. The middle layer, i.e. the vascular membrane, is rich of blood vessels and contains melanin. It consists of the *iris*, the *ciliary body* and the *choroid*. The choroid is located at the posterior 2/3 part of the eye, close to the sclera. It serves the role of supplying nutrition to the retina and prevents the intraocular astigmatism. The retina is the photosensitive nerve tissue membrane and is placed at the innermost layer. The front of the retina is circular, with a hole at the center, namely the *pupil*. The vascular membrane at the posterior part of the iris combines the retina and composes the *ciliary body*, in the inner side of the junction between cornea and sclera. The ciliary body consists of three parts: *ciliary processes*, *ciliary zonule*, and *ciliary muscle*. The ciliary zonule is also called *suspensory ligament*, which hangs the transparent lens behind the iris. The part between the cornea and the lens is called the *anterior chamber*, and the part between the iris and the lens is called the *posterior chamber*. Both the anterior and the posterior chambers are filled with limpid liquid which is called the *aqueous humor*. The approximate 4/5 part of the eye from the posterior side of the lens to the retina is filled with transparent colloidal substance, which is called the *vitreous humor* [42].

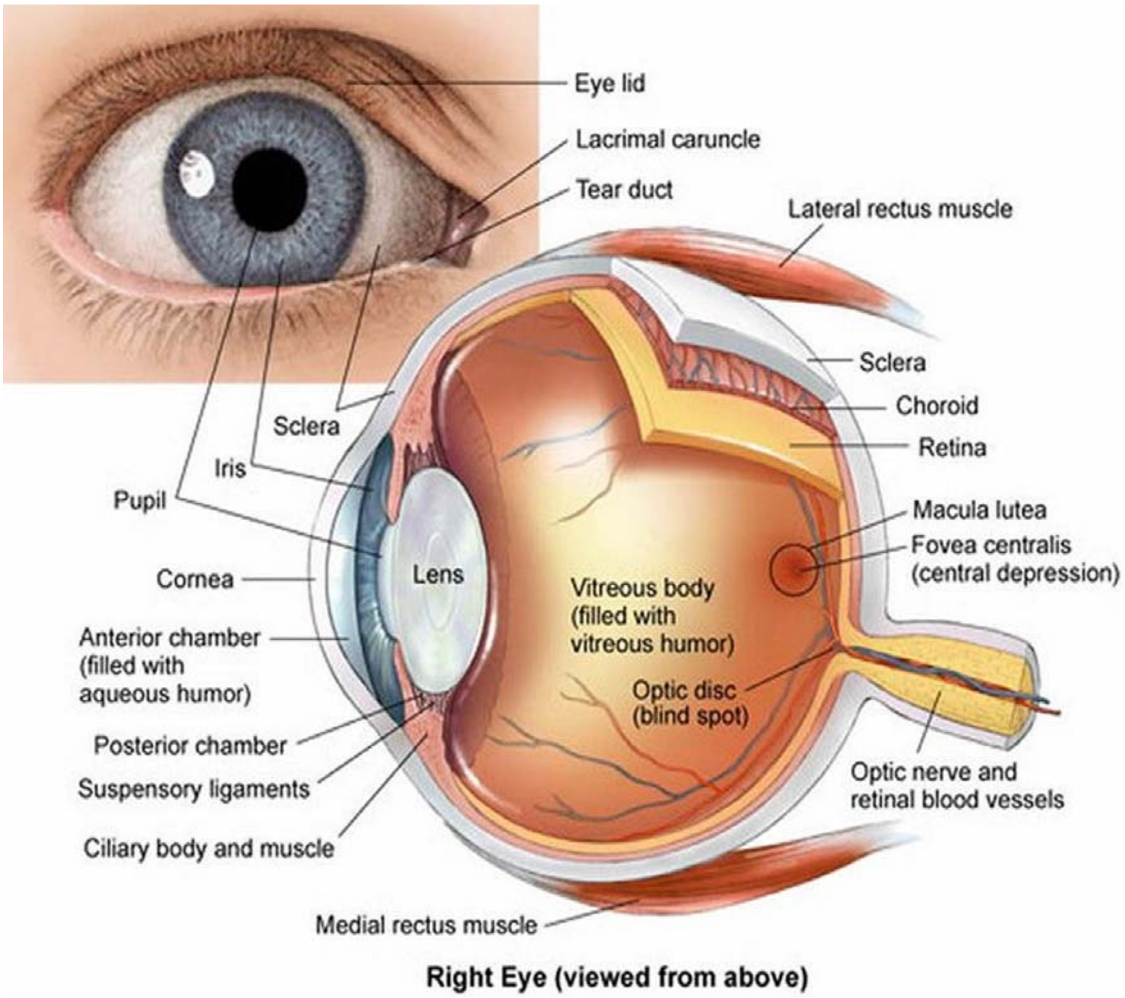

Figure 3.1 The Human Eye Anatomy [43]

3.1.2 Visual Function of the Human Eye

The light goes through three refracting surfaces before finally reaching the retina, which are mainly the *air-cornea interface*, the *aqueous humor-lens interface* and the *lens-vitreous humor interface*. Figure 3.2 shows the functional parameters of the human eye, where *n* represents the refractive index in each part of the eye. If the refractive index of the air is 1.00, that of the cornea is approximately 1.376, and that of the anterior chamber is approximately being 1.336, that of the posterior chamber is approximately 1.337, and that of the lens is approximately between 1.386 and 1.406. Due to the approximate sphericity shape of the cornea, the light is most refracted between the air-cornea interface. The human eye changes its refractive index by adapting the curvature of the front of the lens, thus making the image always fall on the retina.

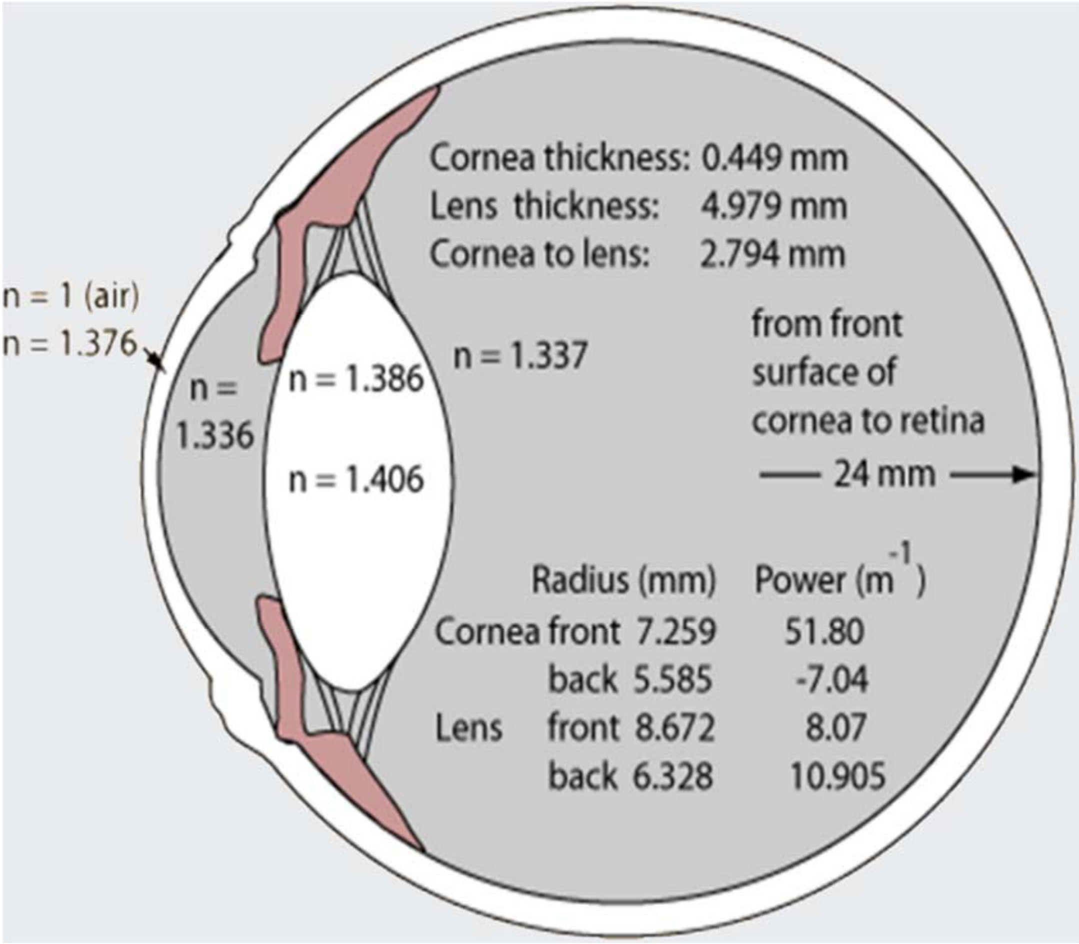

Figure 3.2 The Functional Parameters of the Human Eye [44]

The retina consists of the *pigment layer* (the outer layer) and the *photosensitive nervous tissues* (the inner layer). During the embryo development, it is developed from a brain tissue process. The anatomical structure of the retina is shown in Figure 3.3, which shows that there are not only neuron photosensitive cells, i.e. *rods* and *cones*, which compose the primary neuron of the visual system, but bipolar cells and ganglion cells as well, which are respectively the secondary and tertiary neurons. Besides, rods and cones are widely connected with horizontal cells, while bipolar cells and ganglion cells are indirectly connected with amacrine cells. Overall, the retina can be regarded as both the photosensitive system and part of the central nervous system.

For night-active animals such as mice and owls, the retina is mainly composed of cones, while for day-active animals such as chickens, the retina is almost composed of rods only. At the central fovea of the human retina, there are no rods but only cones, where one cone is connected with a bipolar cell, and a bipolar cell is connected with only one ganglion cell, which forms the special connection from the cone cell to the brain and is adaptive to the high resolving ability of the central fovea. At the other part of the retina, with the location getting closer to the edge, the rods are increasing while the cones are presenting less, where multiple rods or cones connect with one bipolar cell, while multiple bipolar cells connect with only one ganglion. Occasionally, a ganglion can be connected with 250 photosensitive cells, which composes a converging loop of the excitations. Generally, in the human retina there are more photosensitive cells (about 6 million cones and 1200 million rods) and less ganglion cells (about 1 million).

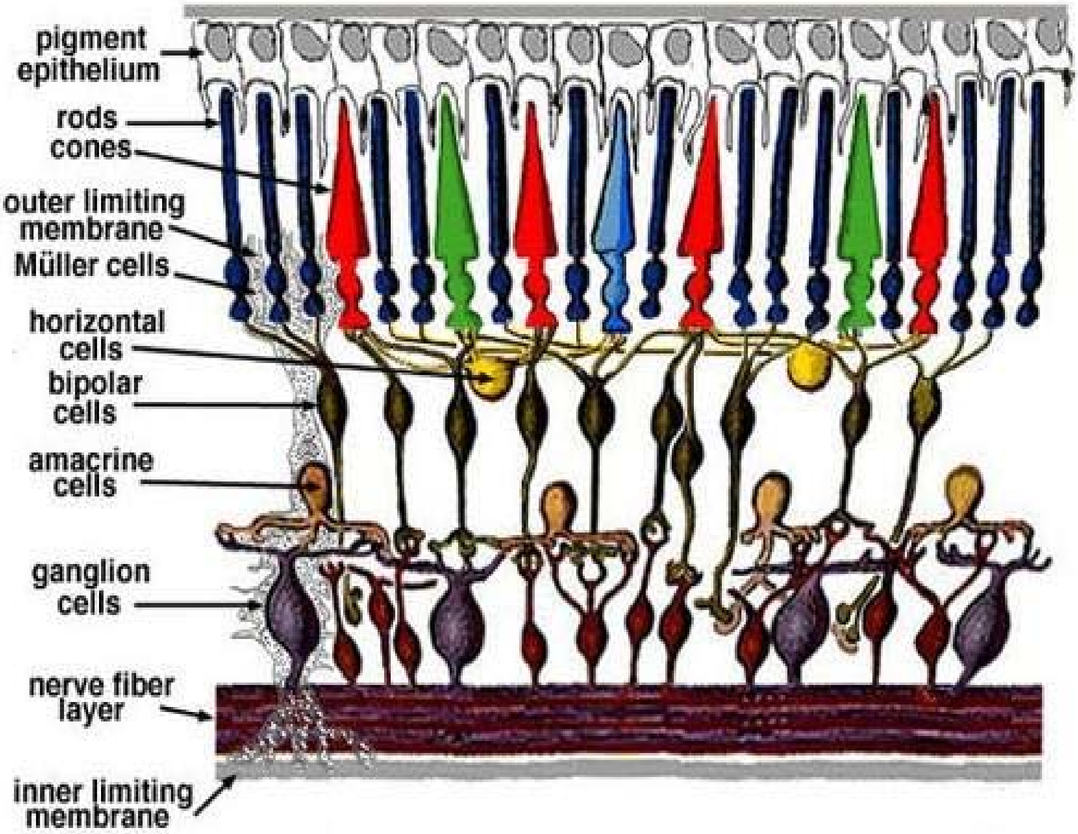

Figure 3.3 The Anatomical Structure of the Retina [45]

3.2 Eye Tracking Technology

Eye tracking is the technology to measure eye movements. The device implementing eye tracking is the *eye tracker*. Techniques to obtain eye movement data can be subdivided into two categories: methods to measure the eye position relative to the head, and techniques to measure the eye orientation in space, namely "point of regard" [46] (widely used for interactive applications). Nowadays, the most widely adopted eye trackers measure the point of regard by means of video-based corneal reflection, which is one of the four broad categories of eye movement measurement methodologies, namely Electro-OculoGraphy (EOG), Scleral Contact Lens/ Search Coil tracking, optical tracking, and video-based combination of pupil/corneal reflection, as above [47].

The EOG method was the main eye tracking technique about 40 years ago [46]. It relies on the electric potential of the skin surrounding the ocular cavity. Therefore, eye trackers based on EOG measurement were usually designed as wearable devices. This method demands low computational power and it can as well monitor eye movements while sleeping. The main disadvantage of this approach is its low accuracy (it can hardly provide precise data about gaze direction [48]).

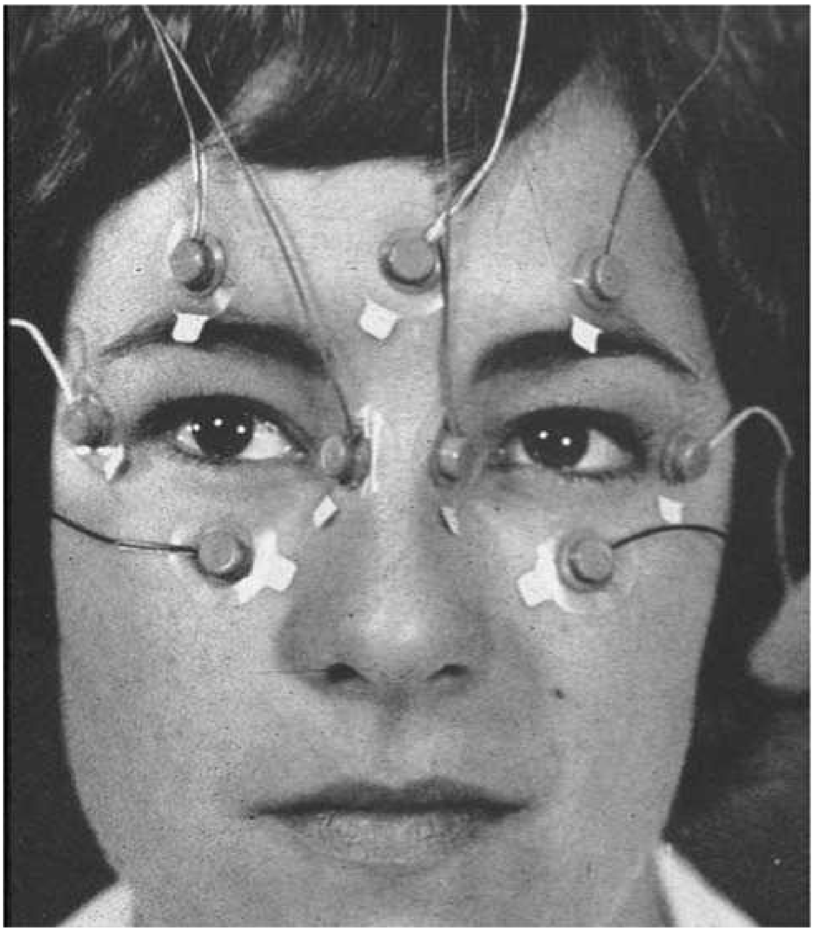

Figure 3.4 Example of an EOG measurement device [49]

Eye-contact tracking techniques, based on scleral contact lenses or search coils attached to the eyes, is a precise method to measure eye movements [46], although it is rather invasive and will cause discomfort to the tester.

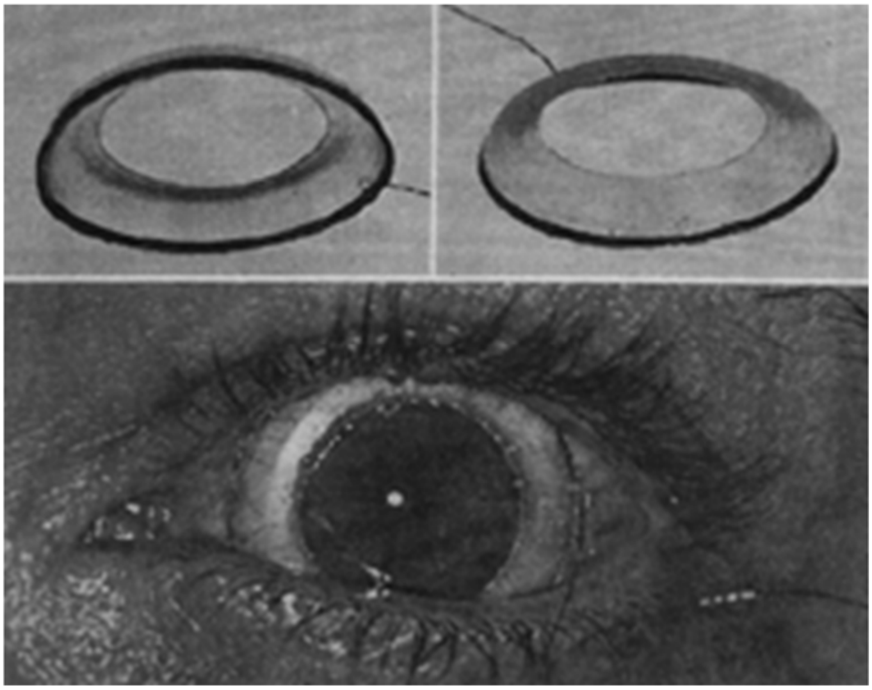

Figure 3.5 Example of a search coil for eye movement measurement [50]

Optical tracking includes the Photo-OculoGraphy (POG) and Video-OculoGraphy (VOG) methods, which exploit the light (typically infrared light) reflected by the eye and sensed by a video camera or some other optical sensor. This method demands testers to keep their head fixed [46].

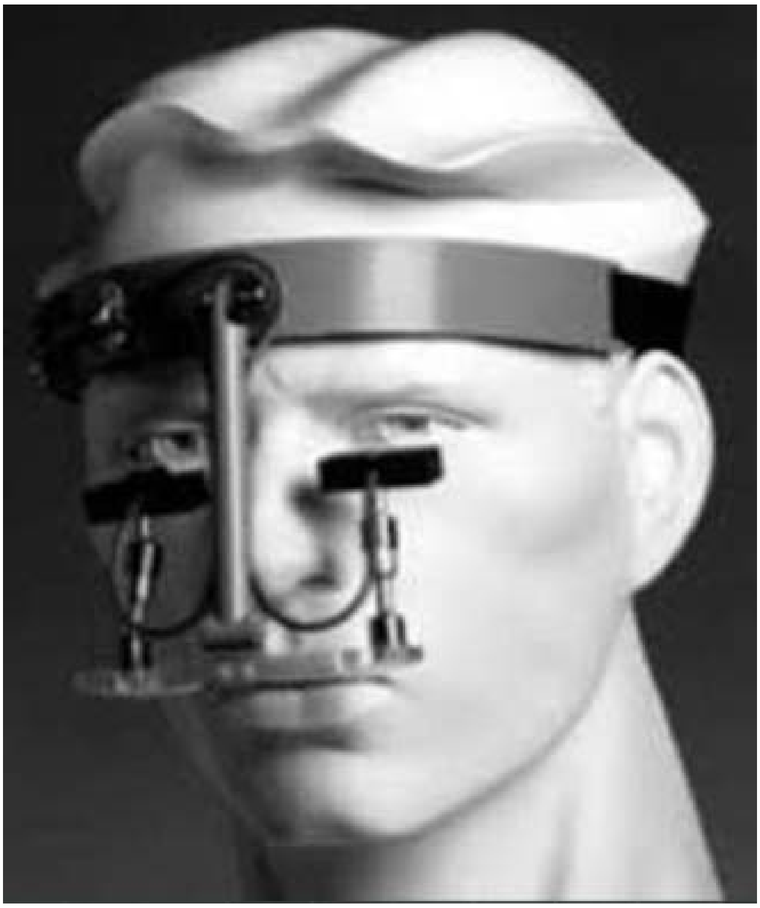

Figure 3.6 Example of an infrared reflection measurement through a wearable device [51]

The video-based combined pupil/corneal reflection method provides two key features, i.e. corneal reflection (a.k.a. Purkinje reflections [52]) and the pupil center, to monitor the point of regard. This technique is adoptable for low-priced cameras to compute the regard point in real-time.

Figure 3.7 Example of the usage of a video-based eye tracker [53]

The current mainly adopted video-based corneal reflection eye trackers can be divided into two types: *wearable eye tracker*s and *remote eye trackers* (a.k.a. "*external eye trackers*") [54].

Wearable eye trackers allow users to freely move their head and body, which makes them feel less constrained or uncomfortable. Since the eye tracker is "worn" by the user, its relative position to the head is always fixed. Some limitations affect such devices, such as the need to adapt to their weight, volume and ergonomics [55][56].

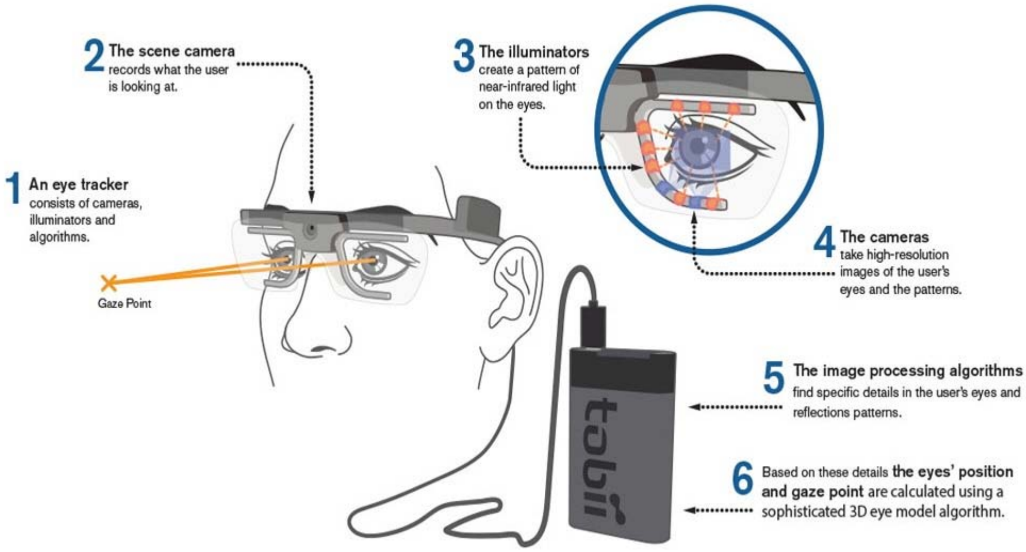

Figure 3.8 Example of a wearable eye tracker [57]

Remote eye trackers are those tracking systems that operate without contact with the user and permit free head movement within reasonable limits without losing tracking [58]. They allow estimating gaze direction in a limited workspace with single or multi camera systems [59]. The main advantage of remote eye trackers is that they are less invasive for users, but this implies that the direction of the gaze will depend on the user's head direction. [54].

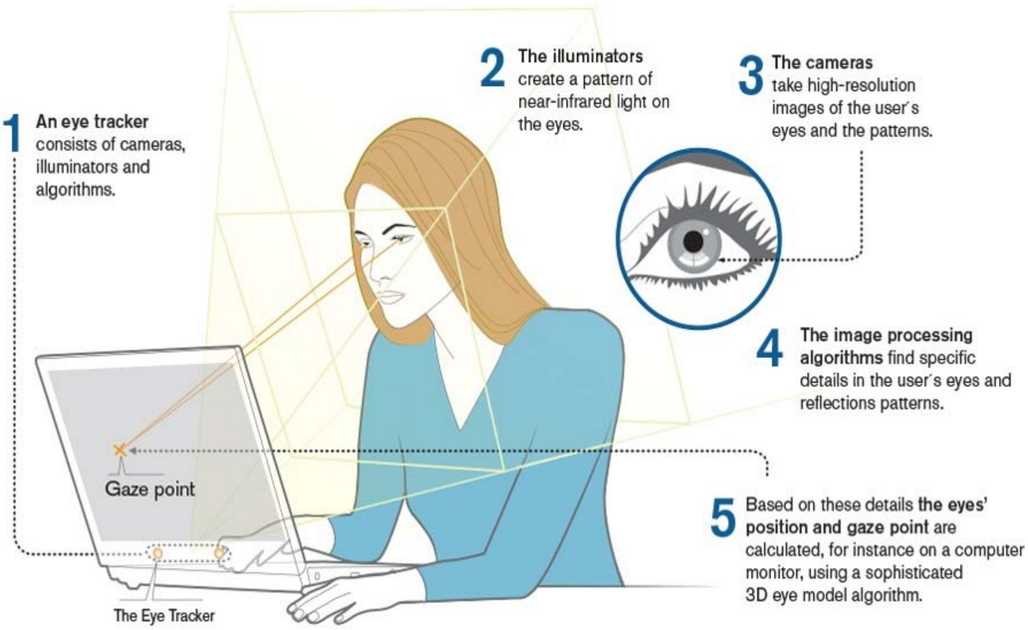

Figure 3.8 Example of a remote eye tracker [57]

3.3 The Eye Tracker Used in Our Case Studies

Nowadays there are several eye tracker manufactures, such as Tobii, SMI, Eye Link, Eye Tribe, Gaze Point, EyeTech, etc. (actually, Eye Tribe and SMI have been acquired by Oculus and Apple in the recent couple of years, respectively [60][61]). The eye tracker which I selected to support my research is the *Eye Tribe* eye tracker, which has the following positive features [62]. Firstly, it occupies a small volume: it is probably the smallest eye tracker device in the world, measuring 20×1.9×1.9 cm; secondly, it does not require a separate power source, which makes it portable. The device uses a USB 3.0 connection, allowing it to run with most computers. Besides, developers can use simple software development kits based on C++, C# and Java programming platforms to develop applications. In the end, the cost of an Eye Tribe eye tracker was low (99 USD), but it is proved to be sufficient for the scientific research and can perform well with sufficiently precise data [63] [64]. Here are the major characteristics of the Eye Tribe eye tracker:

Table 3.1 Major features of the Eye Tribe Eye Tracker [65]

Sampling rate	30 Hz and 60 Hz mode
Accuracy	0.5°(average)
Spatial resolution	0.1°(RMS)
Latency	<20 ms at 60 Hz
Calibration	5, 9, 12 points
Operating range	45 cm – 75 cm
Tracking area	40 cm × 30 cm at 65 cm distance
Screen sizes	Up to 24 inches
API/ SDK	C++, C# and Java included
Data output	Binocular gaze data

Dimensions	(W/H/D)20 × 1.9 × 1.9 cm
Weight	70g
Connection	USB 3.0 Super speed

Chapter 4 *Related Works on Gazed-based Biometrics*

The first eye tracker in the world was probably manufactured in 1908 [66] and, since then, research on eye tracking has been gradually developed over the past century. In 1962, L. Stark G. Vossius and L. R. Young demonstrated that controlling eye movements by changing the position of a target signal would cause important effects on the nature of the biological servomechanism thus indicating the predictive control of eye tracking movements [67]. A. L. Yarbus determined how the human eye examines complex objects and what principles govern this process [68]. After that, eye tracking research flourished during the period from 1970s to 1980s. Especially from the 2000s, research on scientific and business applications of eye tracking started to be developed [69]. Figure 4.1 shows a timeline of eye tracking evolution.

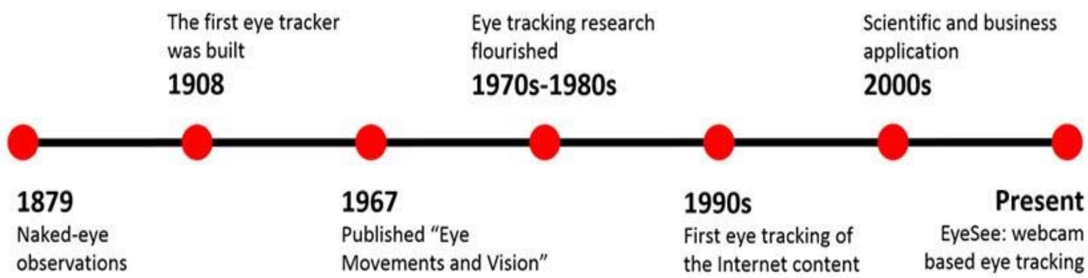

Figure 4.1 A brief history of eye tracking [70]

The following is a brief summarization of the most relevant works on gaze-based biometrics in the past decade, which are listed and summarized by year.

In 2008, A. D. Luca et al. researched on an eye-gesture-based verification method in which the testers moved the eyes to "draw" particular patterns on the screen. The main idea comes from the supposition that it is easier to remember complex shapes than long passwords or PINs. In the experiments, the testers had to perform eye gestures as if "drawing" patterns on the screen by moving their gaze in specific ways. The eye gestures consisted of arbitrary combinations of eight basic strokes starting from or arriving at the screen center along the eight directions given by the four poles and the mid-point of each edge [71]. P. Dunphy, A. Fitch and P. Olivier presented an approach of gaze-contingent interaction as a solution to combat the threat of "shoulder surfing" by demanding the user to identify a sequence of face images in a 3×3 grid, in which one assigned face and eight decoy faces are comprised [72].

In 2009, K. K. E. Ellis, in his master thesis, researched the eye tracking metrics being specific for workload estimation in flight deck operations. The tests were conducted in a flight simulator which comprised a fully functional flight deck and standard Boeing 737 controls. In the experiments, a series of flight scenarios were developed through which the eye tracking metrics were collected. This study provided a simple insight into the trends of eye movement metrics as they responded to induced workload in a cockpit performing a landing task [73]. Nacke et al. researched the feasibility of objective evaluation of player experience in games with biometric evaluation techniques including electromyography, electroencephalography and eye tracking methods. The research was carried out through some experimental game playing cases with the "Left 4 Dead" published by Valve Corporation, through which the eye tracking metrics about saccades, fixations and pupil dilations were measured and analyzed during the enemy contact, fighting and team mate communicating occasions [74].

In 2010, Fookes et al. presented some authentication methods which allowed the users to freely observe stimuli such as a photograph or a video instead of requiring the users to explicitly watch certain screen areas in sequence [75]; T. Kinnunen et al. described eye movements using a histogram of all angles traveled by the eyes during a certain time interval, in which the local velocity direction of the gaze was calculated through trigonometric identities and transformed into a normalized histogram [76].

In 2011, J. Weaver et al. developed a virtual keyboard with symbols displayed as targets to be watched instead of the traditional physical keyboard. The authentication occurred by watching the symbols on the virtual keyboard. Different from typical solutions in which keys were pressed by being looked at for a certain dwell time, in this research gaze points were grouped and automatically analyzed to find out the selected symbols [77]. F. Deravi and S.P. Guness explored the idea that the way an individual looks at external stimuli can be regarded as a distinctive behavioral biometric [78]; C. Holland and O.V. Komogortsev presented an objective evaluation of various gaze-based biometric features and their ability of precise individual identification. Thirty-two participants were involved in the dataset acquisition with a high-frequency eye tracker (1000 Hz). Various features about eye movements were exploited such as fixation count, average fixation duration, average vectorial saccade amplitude, average horizontal saccade amplitude and average vertical saccade amplitude. Besides, the aggregated scan path data were calculated as well, among which scan path length, scan path area, regions of interest, inflection count, slope coefficients of the amplitude-duration and main sequence relationships. Similarity was then measured with the Gaussian Cumulative Distribution Function [79].

In 2012, M. Porta et al. described their ongoing research primarily directed to the development of an e-learning platform in which information about the student's emotional state was obtained by exploiting eye data [80]; O. V. Komogortsev et al. presented that the combined ocular traits had the potential to enhance the accuracy and counterfeit-resistance of existing and future biometric systems. Three fundamentally different traits obtained by the same camera sensor were used in the research. The three traits were the Oculomotor Plant Characteristics (OPC), Complex Eye Movement (CEM) patterns and the unique physical structure of the iris. 87 testers were involved to collect the eye movement and iris data. It was shown in the experimental results that the accuracy could be improved by the combined ocular traits [81]. N. V. Cuong, V. Dinh and L. S. T. Ho proposed an approach using Mel-frequency cepstral coefficients (MFCCs) to encode various features such as eye position, eye difference, and eye velocity for the classification model, with experimental results showing good potential of this method. Eye movement data were obtained using a jumping point as the stimulus. MFCCs were used to encode the useful information as features for the classifier. [82]; I. Rigas, G. Economou and S. Fotopoulos presented a novel approach to exploit eye movements for biometrical identification which revealed the efficiency of the method and encouraged analogous research efforts. Ten photos of faces were employed as stimuli with the sampling frequency of the eye tracker being 50 Hz. Fifteen participants attended eight sessions, and the classification accuracy, which was between 67.5% and 70.2%, was obtained using the KNN and Support Vector Machine classifiers. [83].

In 2013, M. Juhola, Y. Zhang and J. Rasku developed a computational verification method by adopting saccadic eye movements of healthy subjects and otoneurological patients using a dot moving along a black bar as a stimulus. This work compared eye movements detected by two eye trackers with quite different sampling rates, i.e. 400 Hz vs. 30 Hz. Four features were derived from saccadic movements, namely amplitude, accuracy, latency and maximum velocity, which were used in three classifiers namely Linear Discriminant Analysis, Quadratic Discriminant Analysis and Naive Bayesian Rule, with the leave-one-out data split technique. Classification results showed that high frequency data achieved 90% of correct recognition, while low frequency data provided a 70%-90% recognition rate. [84]; A. Darwish and M. Pasquier adopted eye movement features, iris constriction and dilation parameters to investigate the feasibility of dynamic features of the eyes for biometric identification. Data acquisition was done with a 120 Hz eye tracker and involved 22 participants. Experiments were conducted in four sessions, arranged at least twice a week. Different stimuli, among which a 4×4 dot matrix, were used. The following features were calculated: angular velocity (left, right), angular acceleration (left, right), and velocity (left, right). Moreover, for each one of these features, the aggregate values (mean, standard deviation, and the maximum) were calculated. These data were then used to train a Random Forest classifier, with 10-fold cross validation. The best results provided an average HTER (Half Total Error Rate) of 5%. [85].

In 2014, Cymek et al. explored the possibility to exploit the smooth pursuit eye movement for authentication tasks by employing a moving PIN-pad as a stimulus [86]; Nugrahaningsih and Porta researched and proved the possibility to exploit pupil size as an effective distinctive characteristic for gaze-based soft biometrics, with a series of experiments exploiting some basic attributes of pupil size and over one dozen statistics attributes based on them, such as means, standard deviation, median, etc. [87].

In 2015, V. Cantoni et al. proposed a novel Gaze ANalysis Technique ("GANT") by exploiting a graph-based representation of fixation points obtained by an eye tracker during human-computer interaction. In the experiments, the model was constructed from 111 testers of the first study and 24 from the second study. The remaining data were divided into two groups to form the test dataset. Trials using single and combined features were conducted, and the combination of all features (density, total fixation duration, and weighted path) yielded the best ERR of 28%, which demonstrated the conjecture that the way an individual looked at an image might be a soft biometric application [88].

In 2016, D. Cazzato et al. conducted a first attempt to perform biometric identification of individuals by acquiring data with a low-cost, non-invasive, safe and calibration-free gaze estimation framework instead of expensive and unsafe devices which demanded the user's strict cooperation [89]; A. George and A. Routray proposed a novel framework of biometric applications by using eye movement patterns which demonstrated the potential feasibility of a robust anti-counterfeit individual identification system by using eye movement dynamics along with iris recognition. The dataset consisted of three session recordings from 153 participants. Sessions 1 and 2 were separated by a 30 minute interval, while session 3 was recorded one year later. An eye tracker with the high-frequency of 1000 Hz and down-sampled to 250 Hz was employed to collect the data. A jumping point and text formed the stimuli. The eye data were categorized into fixation features (fixation duration, standard deviation for the X coordinate, standard deviation for the Y coordinate, scanpath length, angle with the previous fixation, skewness X, skewness Y, kurtosis X, kurtosis Y, dispersion, and average velocity) and saccade features (saccadic duration, dispersion, mean, median, Skewness). With a procedure of feature selection, the chosen features were input to a RBF-Artificial Neural Network. In the experiments, Session 1 data became the training dataset, and Session 2 and Session 3 data served as a testing dataset. The study yielded a 98% success rate with both stimuli with the 30 minute interval, and 94% as the best accuracy for the 1-year data acquisition interval [90]. C. Galdi et al. presented a critical survey of three existing gaze analysis methods by on the same dataset which consisted of 112 subjects. Sixteen black-and-white face images were displayed in random order as the stimuli. The survey indicated that richer stimuli based on task-dependent user profiles would generate performance improvement [91].

In 2017, F. Vella, I. Infantino and G. Scardino exploited a system which improved the human-computer interaction which identified users according to eye movement without invasive measurements. The process was based on the evaluation of the fixation points while the user was using a browser interface and storing the sequence of the detected points. The recognition of the user was performed through a clustering process employing the Mean-Shift algorithm. The possibility to identify the user and tune the applications towards the needs and characteristic of a single user provides a viable approach to build new and more intelligent interfaces in human machine interaction. Experiments have been led with a set of users typical of a domestic application and results are promising [92]; A. V. Lyamin and E.N. Cherepovskaya proposed a biometric identification method which provided accurate results by using a low-frequency eye tracker. The data were collected with 30 Hz eye tracker and included the following parameters: time (in ms), gaze type (fixation or saccade), gaze point coordinates (in mm) in $A^0x^0y^0$ coordinate system, eye position (in mm) in $A^0x^0y^0z^0$ coordinate system of the monitor, eye pupil diameter (in mm), etc. The proposed algorithms were based on k- Nearest Neighbors and Naïve Bayes classifiers, performing the EER from 15.44% to 16.18% [93].

In early 2018, Q. Wu et al. exploited a multi-task EEG-based individual verification system combining eye blinking signals which could achieve high precision and robustness thus improved the unsatisfactory accuracy and stability of the current EEG-based person authentication systems in practical application [94]; P. Kasprowski and K. Harezlak proposed the first attempt to fuse mouse dynamics and eye movement biometrics as reliable behavioral biometrics. The Eye Tribe eye tracker was used to collect eye movement data of 32 participants with 387 trials at a frequency of 30 Hz. During the experiment, the tester was asked to click some circles with numbers inside to enter a sequenced PIN number. Both the mouse positions and eye gaze positions were recorded during this activity. An SVM classifier was used for training, with the best results obtained for the RBF kernel with gamma = 2^{-9} and $C = 2^{15}$, when the EER of is 6.82% and the classification accuracy is 92.86% [95].

Chapter 5 *Machine Learning and Related Methods*

5.1 Introduction to Machine Learning

Machine Learning techniques address the problem of how to make computers improve automatically their through experience [96]. Machine Learning is a subset of artificial intelligence that uses statistical techniques to give computers the ability to "learn" with data without being explicitly programmed [2]. It is an interdisciplinary subject which involves probability theory, statistics, approximation theory, convex analysis, theory of algorithmic complexity, etc.

Before the term "Machine Learning" was firstly mentioned in 1959 by Arthur Samuel [97], the related researches were mainly focused on the discovery and refinement of traditional statistical methods. In 1950, Alan Turing proposed a "learning machine" which had the ability to learn, which foreshadowed genetic algorithms [98]. In 1951, the first neural network machine, named "SNARC", was built [99]. In 1958, Frank Rosenblatt invented the *perceptron* [100] which gave a basis for many future researches such the Support Vector Machine method. During the early period since the concept of Machine Learning was coined, pioneering researches were conducted using simple algorithms.

In 1964, Bayesian methods were introduced for probabilistic inference in Machine Learning [101]. In 1967, the Nearest Neighbor algorithm was created, which basically started the field of Pattern Recognition [102]. After a period of reduced funding and interest in artificial intelligence research in 1970s (the so-called "AI winter" [103]), a resurgence in machine learning research was led by series of researches in 1980s. In

1980, work the neocognition which is a type of artificial neural network, was published [104]. Furthermore, neocognition later inspired convolutional neural networks, which is quite an important method in the Deep Learning technology [105]. In 1982, the Hopfiled networks, a type of recurrent neural network, was proposed [106]. In 1986, backpropagation was presented and became soon popular [107]. In 1994, the Support Vector Machines method was defined and published [108]. In 1995, the Random Decision Forests technique [109] was proposed. During this period, the computer began to be able to analyze large amounts of data and draw conclusions a.k.a. to "learn" from the results [110]. It is worth mentioning that, in 1997, IBM's AI chess player- “Deep Blue”- beat the world chess champion [111]. In the next decade, i.e. in the 2000s, methods based on Support Vector Machines and different kinds of kernel methods [112] became widespread [113]. From the beginning of 2010s up to now, Deep Learning [105] techniques gradually turned to be feasible and proved tremendous power in some practical cases such as beating two human champions of Go [114], which led a new direction of the development of Machine Learning techniques.

Depending on the type of task of a learning problem, Machine Learning can be subdivided into two categories: *supervised learning* and *unsupervised learning* [115]. Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs [116]. In Supervised learning, the learner receives a set of labeled examples as training data and makes predictions for all unseen points [117]. Unsupervised learning is the machine learning task that learns patterns in the input even though no explicit feedback is supplied [118]. In unsupervised learning, the learner exclusively receives unlabeled training data, and makes predictions for all unseen points. Since in general no labeled example is available in that setting, it can be difficult to quantitatively evaluate the performance of a learner [117]. In studies on biometrics, both the identification and authentication tasks have explicit target labels; thus, these are supervised learning problems.

In the following, some related classification methods which are adopted in this research will be introduced.

5.2 K-Nearest Neighbors

The basic algorithm of Nearest Neighbors model is to consider exactly one nearest neighbor which is the closest training data point to the target point for the prediction. The prediction is then simply the known output for the training point.

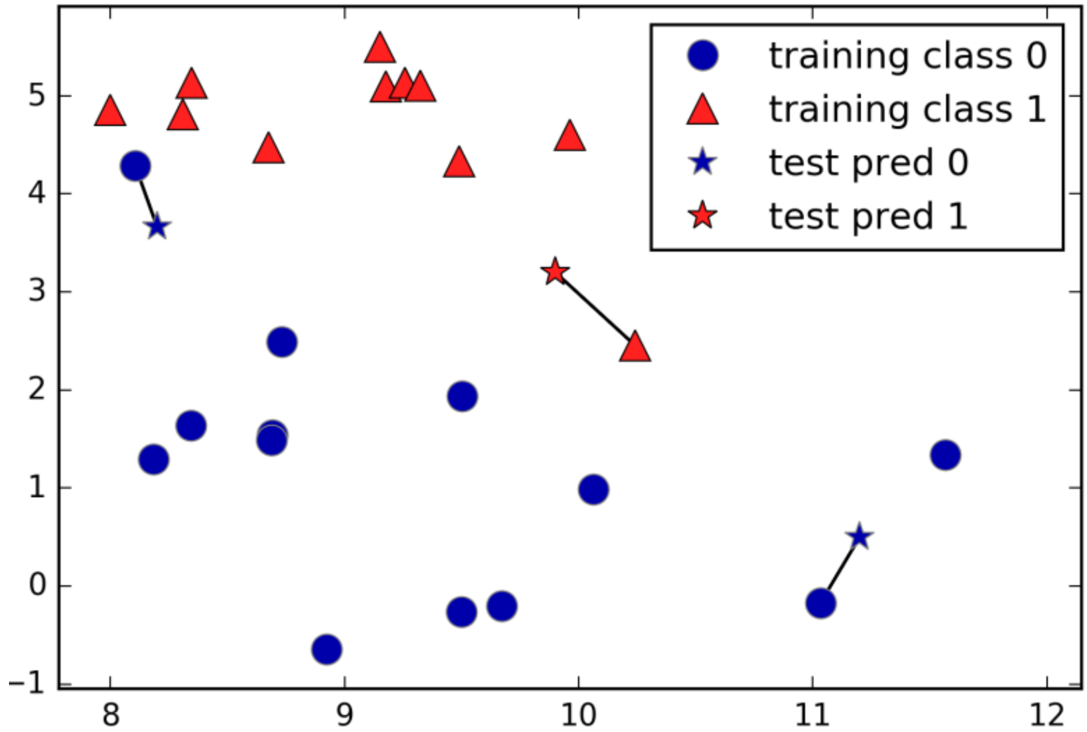

Figure 5.1 Example of predictions made by a one-nearest neighbor model [119]

Consider an arbitrary number, *k*, of closest neighbors instead of only one, and we need to use "voting" to assign a label. For each test point, the number of neighbors which belong to class 0 and those which belong to class 1 are counted. The class that is more frequent i.e. the majority class among the *k*-nearest neighbors, is then assigned. Figure 5.2 shows an example where *k*=3 [119].

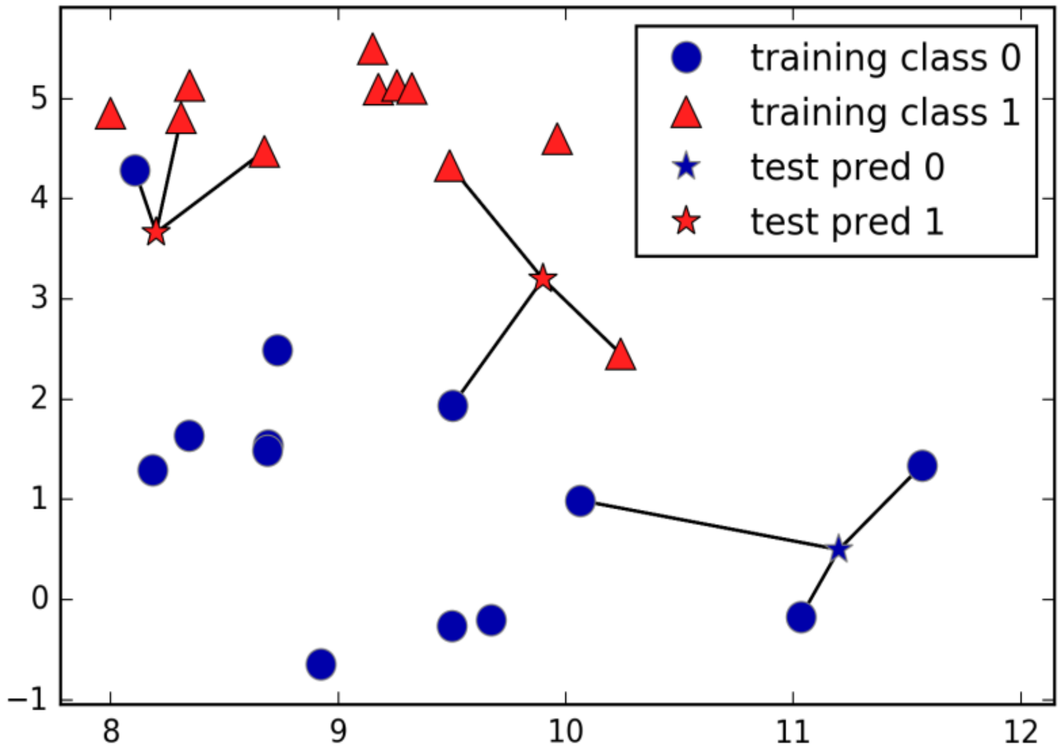

Figure 5.2 Example of predictions made by three-nearest neighbor model [119]

5.3 Classification Trees

The Decision Tree method was firstly introduced by Breiman et al. in 1984 [120]. Decision trees are widely used for classification and regression tasks. Essentially, they learn a hierarchy of if/else questions, leading to a "decision". In classification tasks, decision trees are also called *Classification Trees*.

Classification Trees classify the data according to some features. The data set is separated into two classes according to the answers to some yes/no questions. The answer to each question generates a node or a "leaf". The same procedure continues with the new classes until reaching the ending condition. The questions are based on the learning result of the existing data, and when new data is introduced, it can be categorized into proper leaves by the questions. This is how Classification Trees basically work [121]. Figure 5.3 shows an example of data classification by a Decision Tree.

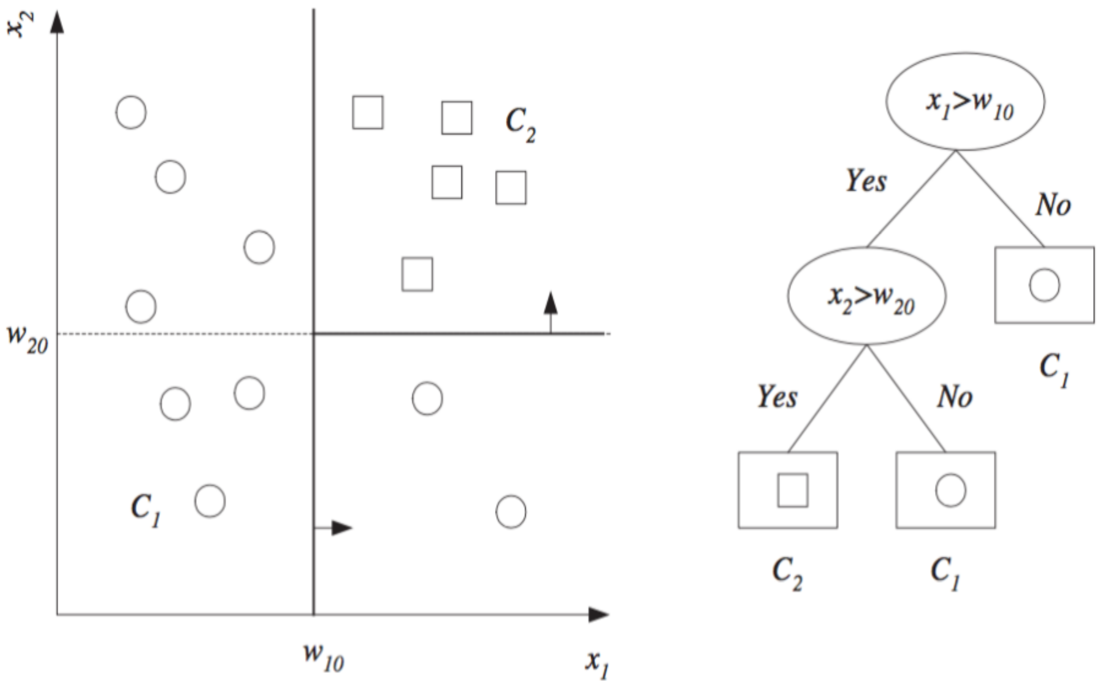

Figure 5.3 Example of data classification by a Decision Tree [122]

5.4 Random Forests

Random Forests are an ensemble learner using the technique of bagging which is a general-purpose procedure used for reducing the variance [123]. Random Forests operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes of the individual trees [109] [124]. The main advantage compared with Classification Trees is correcting the occasionally-occurring overfitting to the training set of Decision Trees [125].

During the training procedure of Random Forests, the technique of bagging, a.k.a. bootstrap aggregating, is employed. For a given training set $X = x_1, x_2 ..., x_n$ with responses $Y = y_1, y_2 ..., y_n$, by B times, the bagging randomly selects a sample to replace the training set and fits decision trees to the selected sample, which can be expressed in the form below:

For b = 1, 2, ..., B:

1. n training examples X_b, Y_b from X, Y as the replacement sample;
2. Train a Decision Tree classifier f_b on X_b, Y_b.

After training, the prediction for any unknown samples x' can be made by taking the majority vote in the Decision Trees. "Bagging" is defined as the procedure above and it produces better performance of data modelling due to the reduction of the variance of the model without the bias being raised. Figure 5.4 shows a conceptual scheme of the Random Forest algorithm, where on the left, trees are trained by partitioning of a sample of the input data, and, on the right, test data pass through each tree and the response is the average of all the single predictions in the forest.

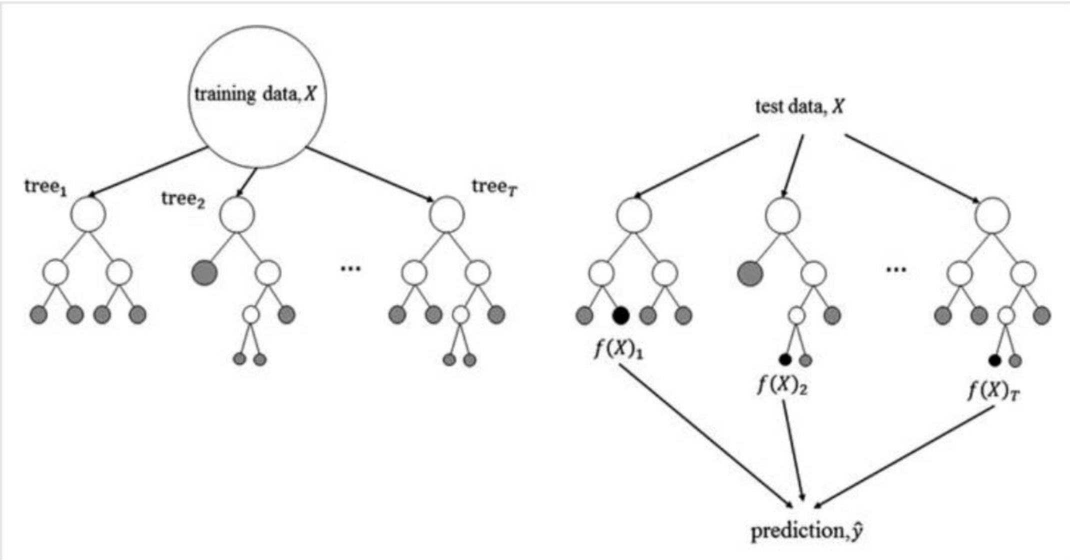

Figure 5.4 Conceptual scheme of the Random Forest algorithm [126]

5.5 Support Vector Machines

The Support Vector Machines (SVM) classification method was proposed by Vapnic et al. in mid-1990s, and its most fundamental principle is based on convex optimization. By a selected nonlinear mapping, it maps the input data set into a high-dimensional feature space where an optimal hyperplane is constructed for the classification, as Figure 5.5 illustrates, where the input data set is presented on the left with the classification result on the right.

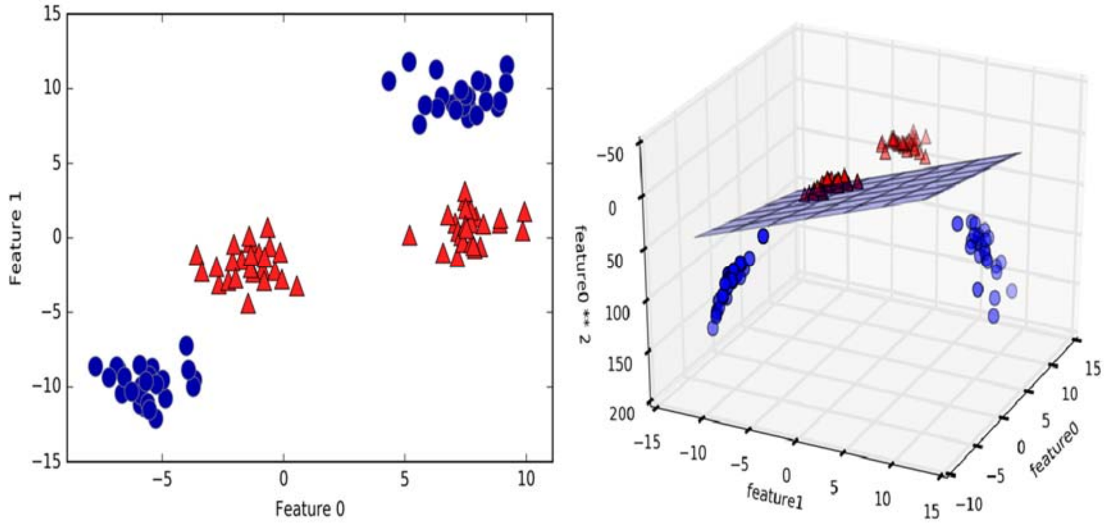

Figure 5.5 Illustration of the SVM classification [127]

The core problem of SVM classification is to maximize the margin from the hyperplane to the target classes. Suppose a training dataset of n points: $(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n)$, where y_i are either 1 or -1 to indicate the class which the point $\vec{x}_i$ belongs. The hyperplane can be expressed as a set of points $\vec{x}$ satisfying $\vec{w} \cdot \vec{x} - b = 0$, where $\vec{w}$ is the normal vector to the hyperplane. If the training data is linearly separable, two parallel hyperplanes can be selected to separate the two data classes to maximize the distance between them and the hyperplane with the maximum margin is the one lying halfway. The two hyperplanes can be expressed mathematically as:

$\vec{w} \cdot \vec{x} - b = 1$, indicating that values on or beyond this threshold are mapped to one class which is labelled as 1;

and

$\vec{w} \cdot \vec{x} - b = -1$, indicating that values beneath this threshold are mapped to the other class, which is labelled as -1.

The distance between the hyperplanes is $\frac{2}{\|\vec{w}\|}$. Therefore, maximizing the distance between the hyperplanes is equivalent to minimizing $\|\vec{w}\|$.

5.6 Naïve Bayes

The Naïve Bayes classifier is based on the following principle: assume that the value of a particular feature is independent of the value of any other feature, given the class variable [128], which is mathematically expressed as [30]:

$$P(C_k|X) = \frac{P(C_k) \cdot P(X|C_k)}{P(X)}$$

where:

- X is a vector $(x_1, x_2, \dots, x_n)$ with n being the number of features;

-$P(C_k)$ stands for the probability to identify class C_k calculated by dividing the frequency in which class C_k appears by the total number of elements;

- $P(X)$ stands for the probability to have instance X;

-$P(C_k|X)$ stands for a conditional probability of having an instance X in class C_k.

The classifier chooses the class C_k that maximizes $P(C_k|X)$. This classifier is called "Naïve" due to the simplistic independence assumption. However, it has been used for years, as its performances have been good enough even when competing with more sophisticated algorithms [129]. This technique is efficient because it learns the parameters using the single features and computes the statistics for each class. There are three typical kinds of Naïve Bayes: GuassianNB, applied to any continuous data, stores the average and standard deviation values of each feature for each class; BernoulliNB, assuming binary data, counts how often every feature of each class is different from zero; MultinomialNB, assuming integer count data, calculates and analyses the average value of each feature for each class [130].

5.7 AdaBoost

AdaBoost, short for *Adaptive Boosting*, is an ensemble learning method firstly proposed in 1995 [131].

AdaBoost is a kind of iteration algorithm, which adds a new weak learning in each round until a pre-defined low enough error rate is reached. Each of the training samples is assigned with a weight to indicate its probability of being selected into the training set by the classifier. If a sample has been accurately categorized, its probability of being selected will be decreased during the creation of the next training set; contrarily, if a sample is not accurately categorized, its weight will be increased. This way, the AdaBoost method can "focus on" those informative (namely more difficult to be categorized) samples. Practically, the weight of each sample was defined equally at the beginning, and as to the k-th iteration, the sample can be selected according to the weights, thus to train the classifier C_k. Based on this classifier, the weights of the wrongly categorized samples are raised and that of correctly categorized samples are decreased. Afterwards, the sample set whose weights have been updated is employed for the training of the next classifier. The whole training procedure iterates in this principle [132]. Figure 5.6 shows the algorithm scheme of AdaBoost.

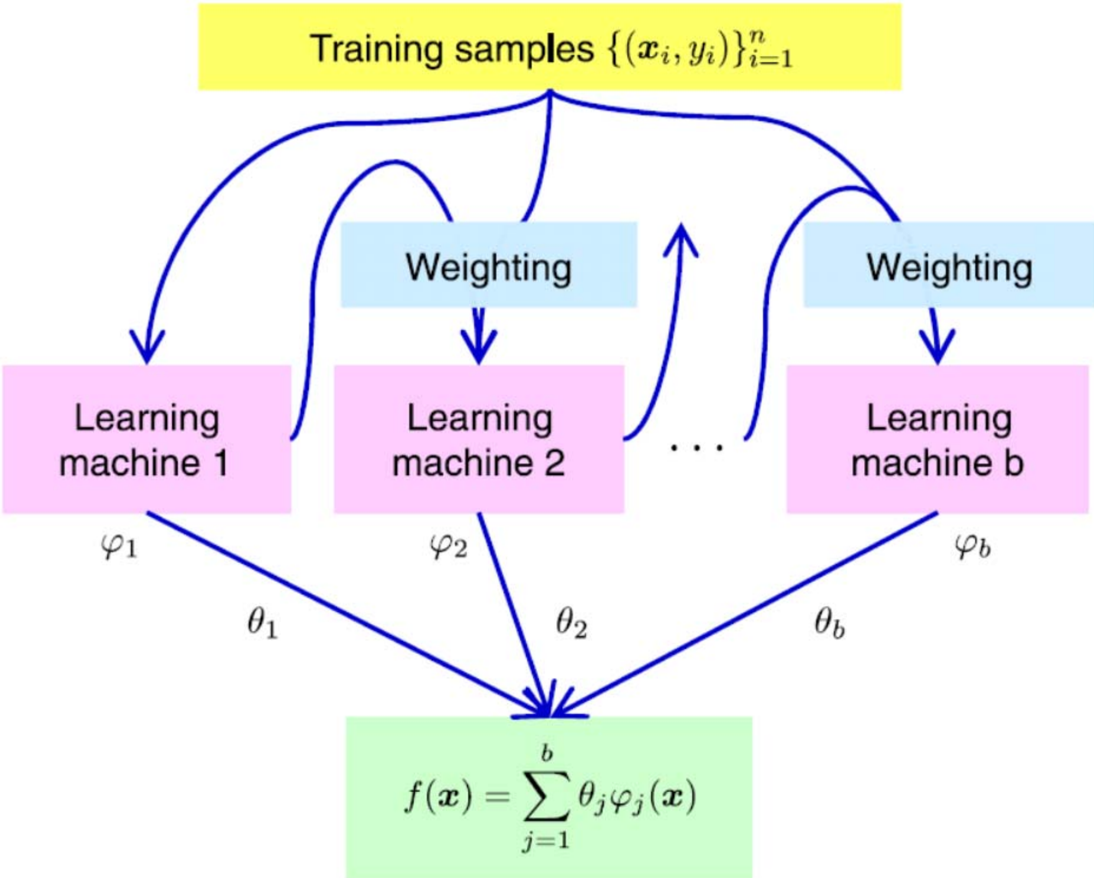

Figure 5.6 Algorithm scheme of AdaBoost [133]

5.8 Neural Networks

Artificial Neural Networks (ANN) are mathematical methodologies which are inspired by human brains and based on biological neural networks that perform multifactorial analysis [134].

An artificial neural network is a network which consists of simple units called “artificial neurons”. A neuron can be regarded as a “black boxes” which receives input, changes its internal state (or “activation”) according to the received input, and produces output depending on the input and state of activation. The entire network is a directed and weighed graph formed up by connecting the output of certain neurons to the input of other neurons. The weights, as well as the functions that compute the activation, can be modified and adjusted by a “learning” process which is governed by a learning rule [135]. Figure 5.7 illustrates the topology of the artificial neuron.

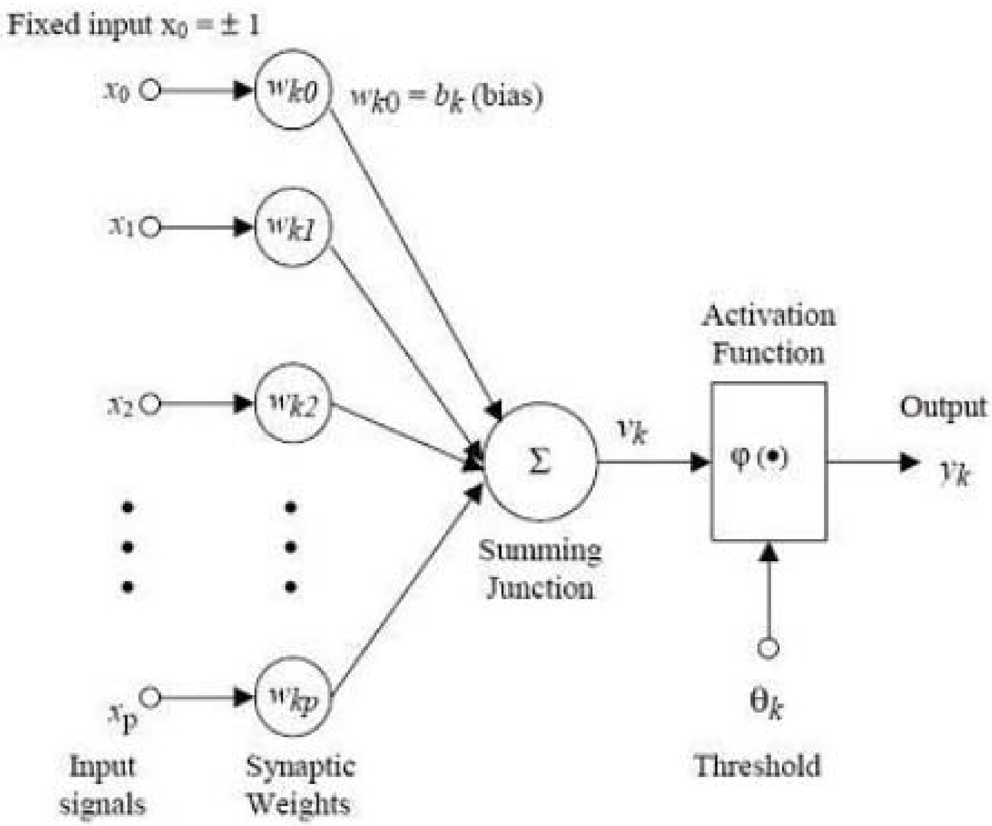

Figure 5.7 Topology of the artificial neuron [136]

A typical multilayer Neural Network consists of three parts:

- input layer, where the neurons receive big quantities of input data vectors.
- output layer, where data are transmitted, analyzed, weighed, and establish the output vectors.
- hidden layer(s), which are the layers formed by neurons and connections. The hidden layer can contain a single layer or multiple layers.

The topology of a Neural Network is illustrated in Figure 5.8:

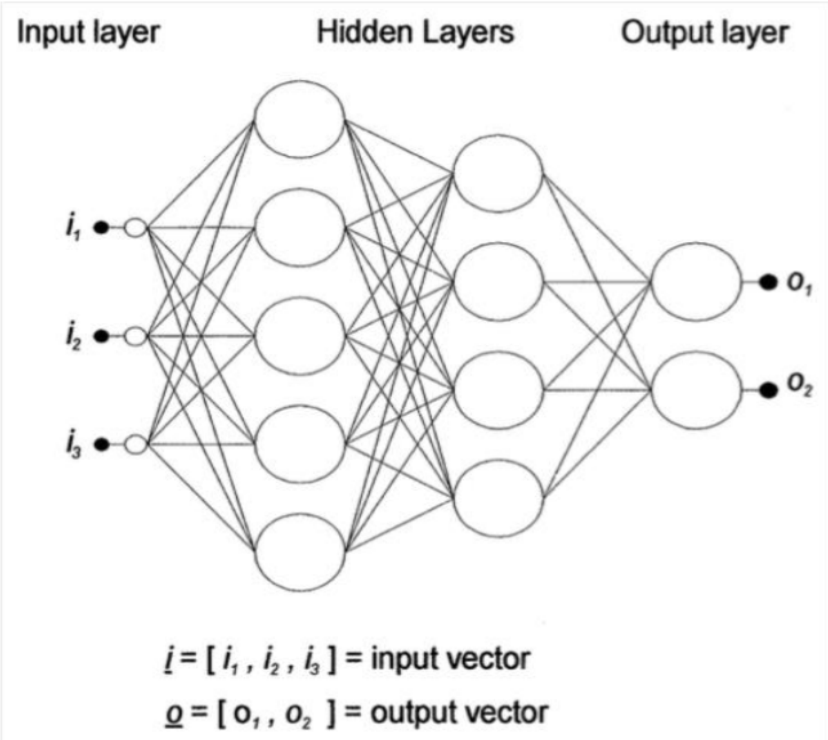

Figure 5.8 Topology of a two-layer Neural Network [137]

Chapter 6 *Case Study on Gaze-based Biometrics for Building Access Security*

6.1 Experiment Design and Data Acquisition

6.1.1 Participants of Experiments

A series of tests were conducted to collect the data. 45 participants aged from 20 to 69 attended the experiments. The participants were composed of 33 males and 12 females. The tests were carried out from December 2017 to January 2018. The purpose was study the gaze behavior when in front of a doorbell board, simulating “first-time”, “frequent”, and “familiar” visitors.

6.1.2 Apparatus

The tests were carried out using an EyeTribe eye-tracker, with a frequency of 60 Hz. Data about eye movements were recorded by means of the Ogama 5.0 software tool [138].

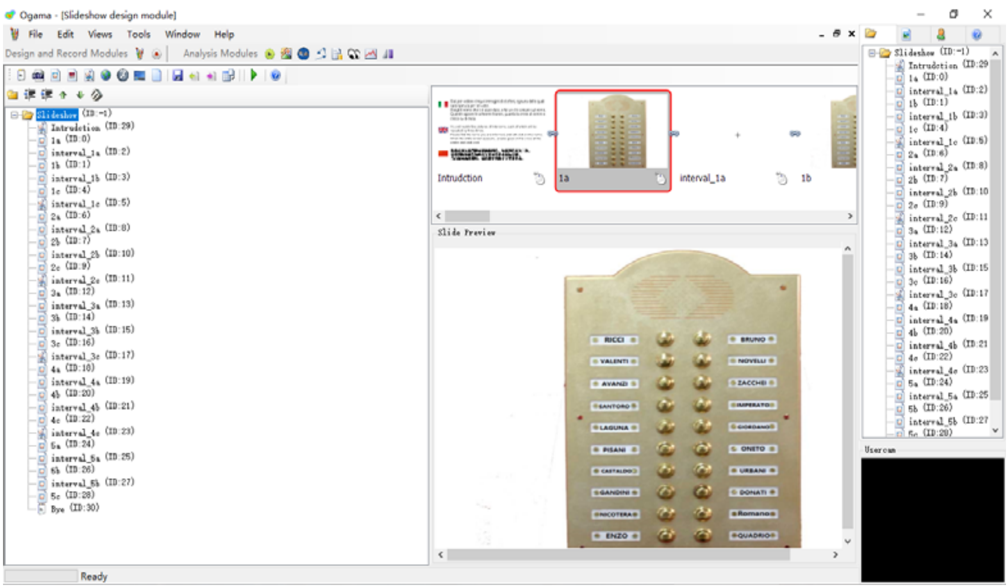

Figure 6.1 Interface of the Ogama software tool (Design Module)

The gaze and fixation data were pre-processed with Microsoft Excel and Python scripts. The familiarity identification stage and the person identification stage were carried out using Orange Canvas [139], and the person authentication stage was carried out through a series of Python scripts.

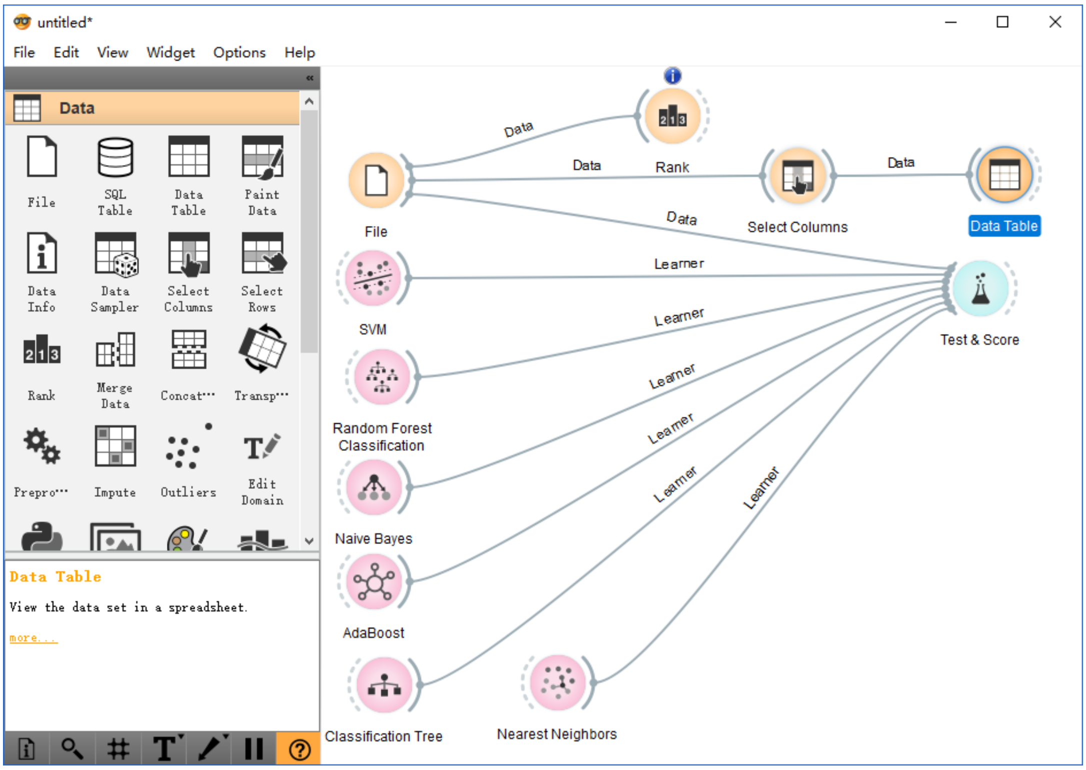

Figure 6.2 Interface of the Orange Canvas

6.1.3 Experiment Procedure

The experiments took place in the Computer Vision and Multimedia Laboratory of the Department of Electrical, Computer and Biomedical Engineering of the University of Pavia (“UniPV-CVML” for short in the following). A white wall was, opposite to the device and the user, offered a quiet and light-friendly environment for the experiments.

Five images of doorbell names, in different sizes, were chosen as observation objects, as Figures 6.1- 6.5 show below.

Figure 6.3 Image of the first doorbell name board

Figure 6.4 Image of the second doorbell name board

Figure 6.5 Image of the third doorbell name board

Figure 6.6 Image of the fourth doorbell name board

Figure 6.7 Image of the fifth doorbell name board

After a short calibration procedure, aimed to adjust the eye tracker's parameters to the specific subject, the experiment started with an instruction image describing the rules of the experiment in three languages: Italian, English and Chinese, as is shown in Figure 6.6.

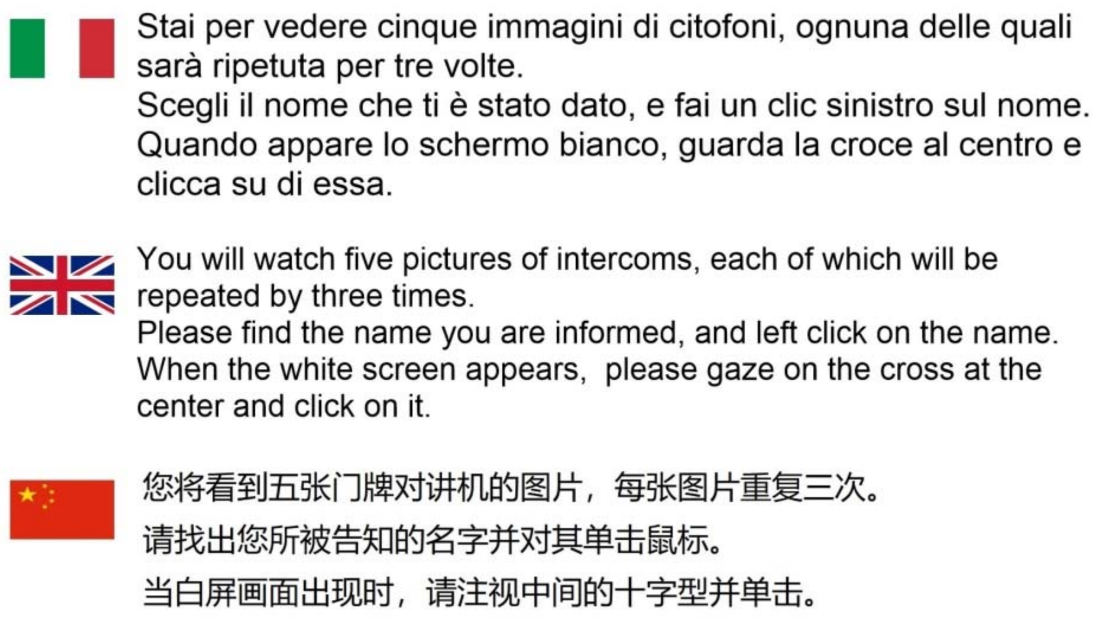

Figure 6.8 Instruction image

Before the tester pressed a key to skip the instruction image, an experiment conductor (also the experiment designer, i.e. me) would choose the name of the first picture and tell it to the tester. Meanwhile, the tester who had just taken the calibration would keep their head as still as possible until completing the test. Subsequently, the tester observed the images with doorbell names presented on the screen. After identifying the correct name, the tester would focus their gaze on it and click with the left mouse button within its area. To simulate the different levels of familiarity of testers with the specific doorbell names, each image was presented three times in three trials, one immediately after the other, as if the tester was a “first-time visitor”, a “frequent visitor”, or a “familiar visitor”. Before the presentation of each image, a blank screen with a white background and a black cross in the center was shown, to make the tester “reset” their gaze position by looking at the cross. A short notification which reminded the tester not to click the mouse until they had been told the next name to be looked for was also displayed after the three repetitions of a doorbell names picture. Figure 6.7 shows the interval image with the notifications.

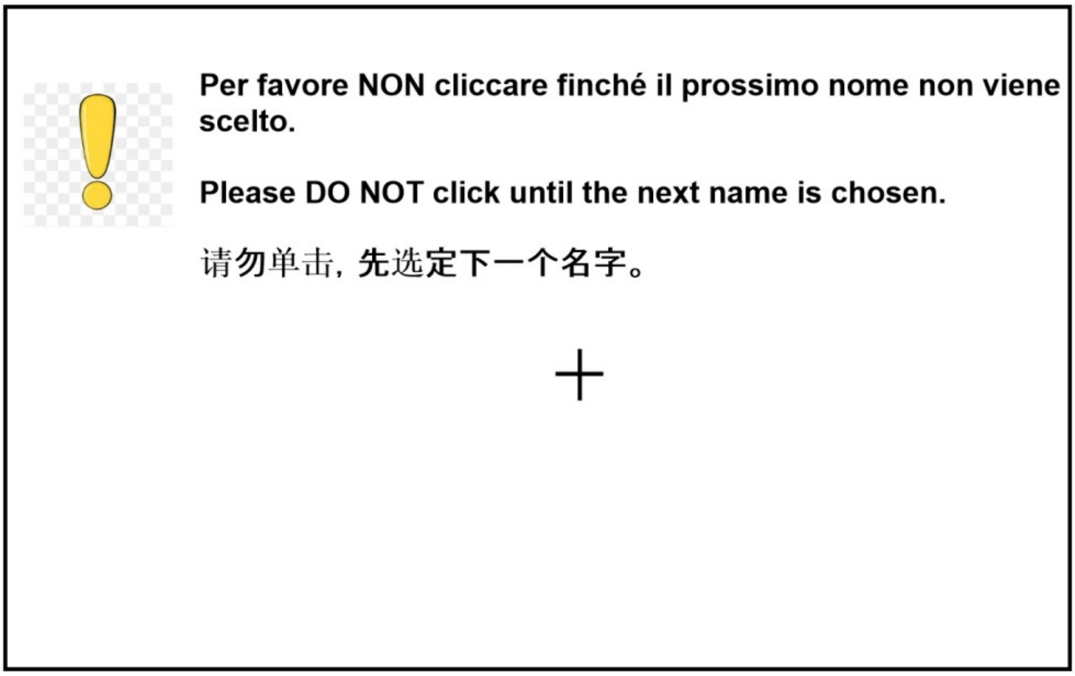

Figure 6.9 Image displayed after each specific sequence of three pictures of doorbell names

6.1.4 Data Pre-Processing and Refinement

The data recorded by Ogama include some features which are not relevant to our gaze-based analysis, such as Mouse Position, Trial Sequence and Handedness. These useless features were simply discarded.

For the analysis, rectangular areas Of Interest (AOI) were defined in the images for each name.

After each test, the raw data generated by Ogama would be checked immediately. Sometimes it was found that, for some testers, the data for the features “left pupil diameter” or the “right pupil diameter” (or sometimes both) had a “0.0000” value. According to the frequency of the device (60 Hz), the “zero data” should indicate that, during the interval of 0.0167 seconds between each gaze sample, the tester had closed their left or right eye, or both, so that the device could not measure the pupil diameter. However, if too many zero values data were present, it is obvious that there was a problem. This might be caused by bad light conditions, device faults or incorrect gaze detection (for example, astigmatism is sometimes problematic).

To make the data more accurate, some refinements were adopted before the data analysis:

1. If all the data of a tester presented non-zero values, the set of data was regarded as perfect and accepted without any modification;
2. If the data of one tester presented several lines of zero data, but less than 10% of the total, the set of data could be accepted after deleting the lines with the zero data.
3. If the quantity of zero data was higher than 10% of the total lines, this set of data was regarded as a bad set, and the tester had to retake the experiment.
4. If a tester still failed to obtain enough good data after retaking the experiment for several times, the tester was not included in the experiments.

Through the above measurements, the data of 43 testers (32 males and 10 females) were employed.

6.1.5 Feature Selection

For the classification task, the feature vector was defined with ten features in total: TFD, NOF, SPL, SPL/TFD, TFD/NOF, AvgDistFix, AvgPX, AvgPY, Avg(PX/PY), Avg(PX-PY). Here are the descriptions of these features:

-TFD stands for “total fixation duration”, calculated by adding up all of the durations of each fixation during the observation of each of the three trials of each picture;

-NOF stands for “number of fixations”, directly obtained from the data generated by Ogama;

-SPL stands for “scan path length”, calculated by adding up the Euclidean distances between every two fixations;

-SPL/TFD and TFD/NOF are two derivative features based on the previous three fixation features, among which SPL/TFD can be regarded as the average saccade speed, and TFD/NOF can be regarded as the average duration of each fixation;

-AvgDistFix stands for “average distance of fixations”, calculated as the average value of the Euclidean distances between every two fixations;

-AvgPX and AvgPY, respectively, stand for “average left pupil diameters” and “average right pupil diameters”, calculated from the average values of the left pupil diameters and the right diameters of each trial of each picture;

-Avg(PX/PY) and Avg(PX-PY) are, respectively, the difference and the quotients of the previous two basic pupil size features.

6.2 Analysis and Results

6.2.1 Analysis stages

The analysis consisted of three stages: (1) familiarity identification (i.e., recognizing the familiarity of an individual about the name location within the doorbell names), (2) person identification (i.e., recognizing an individual using biometric data in a dataset), and (3) person authentication (i.e., verifying an individual's claimed identity using biometric data in a dataset).

6.2.2 Familiarity Identification

The pre-processed data were labeled with the target classes "a", "b", and "c", to indicate the three trials. Six classifiers were used: Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes, AdaBoost, Classification Tree, and k-Nearest Neighbors (KNN). Two types of data sampling methods were adopted: (1) random sampling, with 70% of data for training and 30% for testing; (2) 10-fold cross validation. Table 6.1 shows the Classification Accuracy (CA) obtained with the entire data set.

Table 6.1. Results for familiarity identification using the entire data set

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
SVM	0.557	SVM	0.559
RF	0.552	RF	0.579
Naïve Bayes	0.567	Naïve Bayes	0.557
AdaBoost	0.520	AdaBoost	0.552
Classification Tree	0.503	Classification Tree	0.535
KNN	0.584	KNN	0.614

The entire data set presented very unsatisfactory classification results with all the six classifiers. During both the experiment procedure and the data pre-processing, it was noticed that some testers became perfectly familiar about the position of the chosen name just after the first trial. Consequently, the second and the third trials were not clearly distinguishable. To verify this fact, we considered the data of only the second and third trials. Table 6.2 shows the results, which also in this case are completely unsatisfactory.

Table 6.2. Results for the familiarity identification using the data of trial 2 and trial 3

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
SVM	0.497	SVM	0.488
RF	0.519	RF	0.495
Naïve Bayes	0.528	Naïve Bayes	0.502
AdaBoost	0.509	AdaBoost	0.490
Classification Tree	0.504	Classification Tree	0.498
KNN	0.540	KNN	0.569

The combinations of trial 1 and trial 2 and of trial 1 and trial 3 were also considered. Table 6.3 and Table 6.4 show the results.

Table 6.3. Results for the familiarity identification using the data of trial 1 and trial 2

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
SVM	0.817	SVM	0.807
RF	0.803	RF	0.829
Naïve Bayes	0.833	Naïve Bayes	0.824
AdaBoost	0.748	AdaBoost	0.764
Classification Tree	0.766	Classification Tree	0.752

KNN	0.825	KNN	0.833

Table 6.4. Results for the familiarity identification using the data of trial 1 and trial 3

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
SVM	0.848	SVM	0.848
RF	0.852	RF	0.843
Naïve Bayes	0.863	Naïve Bayes	0.852
AdaBoost	0.782	AdaBoost	0.819
Classification Tree	0.827	Classification Tree	0.831
KNN	0.848	KNN	0.845

As can be seen, the CA values for trial 2 and trial 3 ranged from 0.488 to 0.569, while those for trial 1 and trial 2, and for trial 1 and trial 3, ranged from 0.782 to 0.863. Our initial supposition that trial 2 and trial 3 were not so notably distinguishable was thus confirmed (while trial 1 can be clearly distinguished from trial 2 and trial 3). Because of this, the dataset was reorganized in another way: feature vectors from trial 1 were labeled as "fresh", while feature vectors from trial 2 and trial 3 were combined into one class labelled as "familiar". Table 6.5 shows the results.

Table 6.5. Results for the familiarity identification of "trial 1 vs. trial 2 + trial 3"

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
SVM	0.821	SVM	0.803
RF	0.858	RF	0.851
Naïve Bayes	0.852	Naïve Bayes	0.848
AdaBoost	0.805	AdaBoost	0.792
Classification Tree	0.829	Classification Tree	0.859
KNN	0.860	KNN	0.844

The classification for "trial 1 vs. trial 2 + trial 3" also presented good CA values ranging from 0.792 to 0.860.

6.2.3 Person Identification

For person identification analysis, the "tester ID" label was assigned to each feature vector obtained from the pre-processed data. The same six classifiers used for the familiarity identification analysis were employed, i.e. SVM, RF, Naïve Bayes, AdaBoost, Classification Tree, and KNN (with the same two types of data sampling methods adopted, i.e. 70-30 random sampling and 10-fold cross validation). The following tables show the personal identifications results for the entire data set (Table 6.6), trial 1 and trial 2 (Table 6.7), trial 1 and trial 3 (Table 6.8), and trial 2 and trial 3 (Table 6.9).

Table 6.6. Results for person identification using the entire dataset

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
SVM	0.499	SVM	0.535
RF	0.527	RF	0.546
Naïve Bayes	0.341	Naïve Bayes	0.343
AdaBoost	0.478	AdaBoost	0.522
Classification Tree	0.470	Classification Tree	0.513
KNN	0.040	KNN	0.038

Table 6.7. Results for the person identification using the data of trial 1 and trial 2

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
SVM	0.447	SVM	0.500
RF	0.496	RF	0.581
Naïve Bayes	0.303	Naïve Bayes	0.295
AdaBoost	0.471	AdaBoost	0.500
Classification Tree	0.470	Classification Tree	0.514
KNN	0.028	KNN	0.029

Table 6.8. Results for the person identification using the data of trial 1 and trial 3

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
SVM	0.422	SVM	0.443
RF	0.476	RF	0.476
Naïve Bayes	0.277	Naïve Bayes	0.286
AdaBoost	0.441	AdaBoost	0.450
Classification Tree	0.426	Classification Tree	0.455
KNN	0.035	KNN	0.033

Table 6.9. Results for the person identification using the data of trial 2 and trial 3

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
SVM	0.504	SVM	0.550
RF	0.504	RF	0.550
Naïve Bayes	0.344	Naïve Bayes	0.345
AdaBoost	0.497	AdaBoost	0.505
Classification Tree	0.464	Classification Tree	0.488
KNN	0.049	KNN	0.048

KNN always exhibited extremely low CA values ranging from 0.028 to 0.049. Among the other classification algorithms, the highest score (0.581) was obtained by Random Forest with 10-fold cross validation for the dataset of trial 1 and trial2, while the lowest score (0.277) was obtained by Naïve Bayes with 70-30 random sampling for the dataset of trial 1 and trial 3. All combinations of datasets and trials produced totally unsatisfactory results with all the six classifiers (especially KNN).

6.2.4 Person Authentication

For person authentication analysis, feature vectors were labelled with the binary value of “True” or “False”, to describe “the true target” and the “imposters”. The same features used for familiarity detection and person identification were employed. According to the bad performance of KNN for identification, it was excluded in this stage, while Neural Network (NN) was adopted in addition to the other five classifiers, i.e. SVM, RF, Naïve Bayes, AdaBoost, and Classification Tree. Three types of data sampling methods were applied: 1. random sampling with 70% data; 2. 10-fold cross validation; 3. test on test data (i.e., the first two trials used for training and the third trial for testing). For each method, the six classifiers mentioned above were iteratively applied for ten times, during each of which the “legible target” was randomly selected among the testers, while the rest of the testers were marked as “the others”. Finally, the average scores of the CA values of the ten-time iterations were calculated. Table 6.10 shows the results of the analysis.

Table 6.10 Results for person authentication

Methods	70-30 Random Sampling	10-fold cross validation	Test on Test Data
Naïve Bayes	0.833	0.849	0.504
NN	0.827	0.845	0.502
AdaBoost	0.841	0.867	0.495
RF	0.873	0.888	0.510
Classification Tree	0.832	0.852	0.504
SVM	0.838	0.833	0.494

As can be seen, the highest score (0.888) was obtained with the RF method and 10-fold cross validation, while the lowest score (0.494) was produced by SVM with Test on Test Data. Moreover, the Test on Test Data performance was significantly lower than that of the other two sampling types.

6.3 Summary of the Results

In this case study, the feasibility of using gaze behaviors as a biometric method was investigated by a case about building access security. Specifically, I have considered three aims, namely the detection of the "familiarity level" with doorbell names, and the recognition and authentication of a subject.

For familiarity identification, it was not possible to precisely distinguish the level of familiarity in the three trials, namely "initial visitor" (trial 1), "frequent visitor" (trial 2), or "familiar visitor" (trial 3). However, satisfactory results were obtained when comparing trial 1 with trials 2 and 3 (together or separately).

For person identification, maybe because of the limited number of samples per subject available, the identification results achieved only 60% of classification accuracy.

For person authentication, performance was decidedly better (apart from the "Test on

Test Data” case).

In conclusion, we think that the gaze-based biometric approaches presented in this work can be promising if appropriate classifiers and data sampling methods are adopted, and sufficient data are collected.

Chapter 7 *Case Study on Gaze-based Biometrics through Static Images Observation*

7.1 Experiment Design and Data Acquisition

7.1.1 Participants of experiments

The experiment was designed early in 2017 and carried out from April 2017 to the beginning of May 2017 with three repetition sessions. Forty-three testers, aged from 20 to 68 participated the experiment, 30 of whom took at least two sessions of the experiment, and 18 took the third session. Each two sessions were separated by at least one week at least and two weeks at most.

7.1.2 Apparatus

The tests were carried out using an EyeTribe eye-tracker, with a sampling frequency of 30 Hz. Data about eye movements were recorded by means of the Ogama 5.0 software tool. Gaze and fixation data were pre-processed by Microsoft Excel and a Python script. The identification analysis was carried out by Orange Canvas, and the authentication analysis was carried out by Python scripts with the Orange library.

7.1.3 Experiment Procedure

The experiments took place in the Computer Vision and Multimedia Laboratory of the University of Pavia, where a quiet and light-friendly environment for the eye tracking experiments was offered.

The slides used as experiment materials were played in Ogama system with a full HD resolution of 1920×1080. Twenty images, sorted in 4 categories were displayed as the stimuli in each session of the experiment. The 4 categories were: “Animals”, with 5 images of four-foot animals (side views), which might probably attract the tester’s attention of vision with their specific parts such as paws, head or muzzles etc.; “Cars”, with 5 images of cars (side views), which might probably attract tester’s vision attention with their doors, wheels or headlights; “Scissors”, with 5 images of scissors in different sizes and orientations, which might probably attract the tester to observe their sharp edges and handles ; “Objects”, with 5 images, each of which showing a set of four objects with simple shapes placed in different sequences, which would probably attract the attention with their details. Images of the 4 categories are respectively presented as in Figure 7.1-7.4.

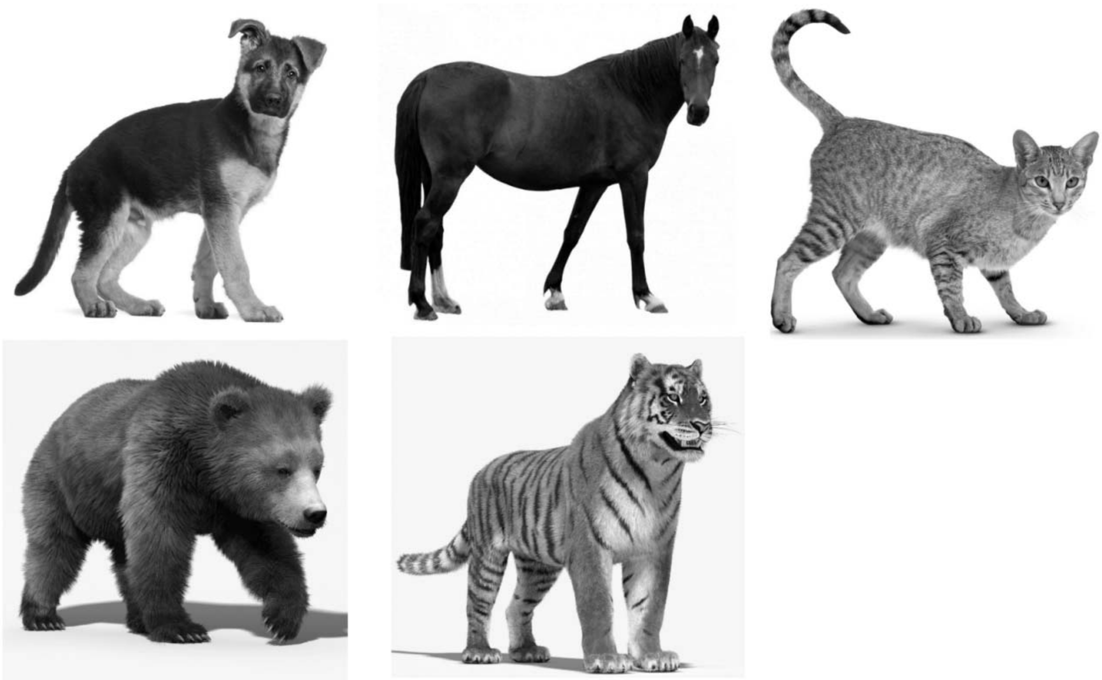

Figure 7.1 Images of the category “Animals”

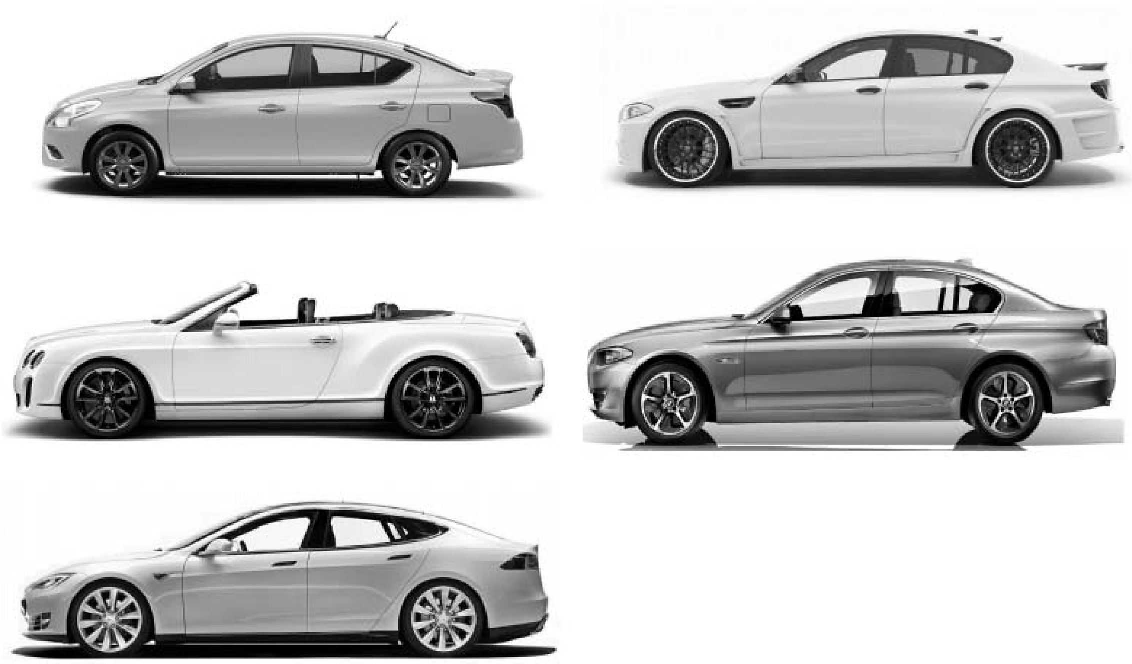

Figure 7.2 Images of the category “Cars”

Figure 7.3 Images of the category "Scissors"

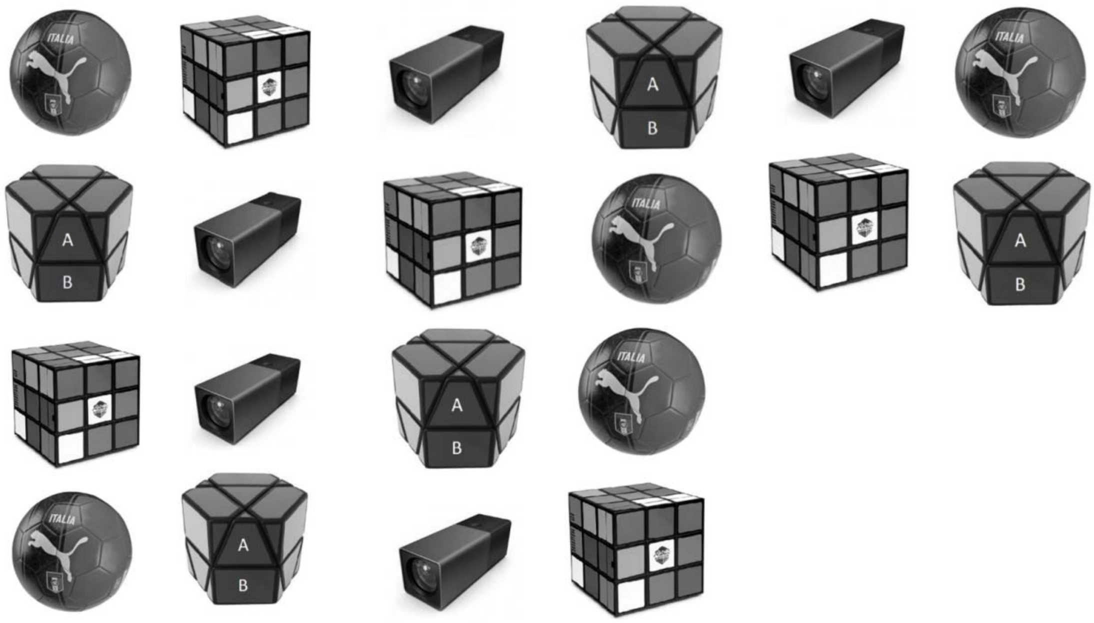

Figure 7.4 Images of the category "Objects"

After taking a short calibration procedure which is aimed to adjust the eye tracker's parameters to the specific subject, the experiment started by randomly displaying the prepared stimuli. The tester had to freely watch the displayed stimuli and keep the head and the body as static as possible. Each of the stimulus images would be displayed for 6 seconds. Between every two stimuli, an interval screen with a cross at the center would be displayed for 2 seconds to guide the tester's initial view point back to the initial location i.e. the screen center.

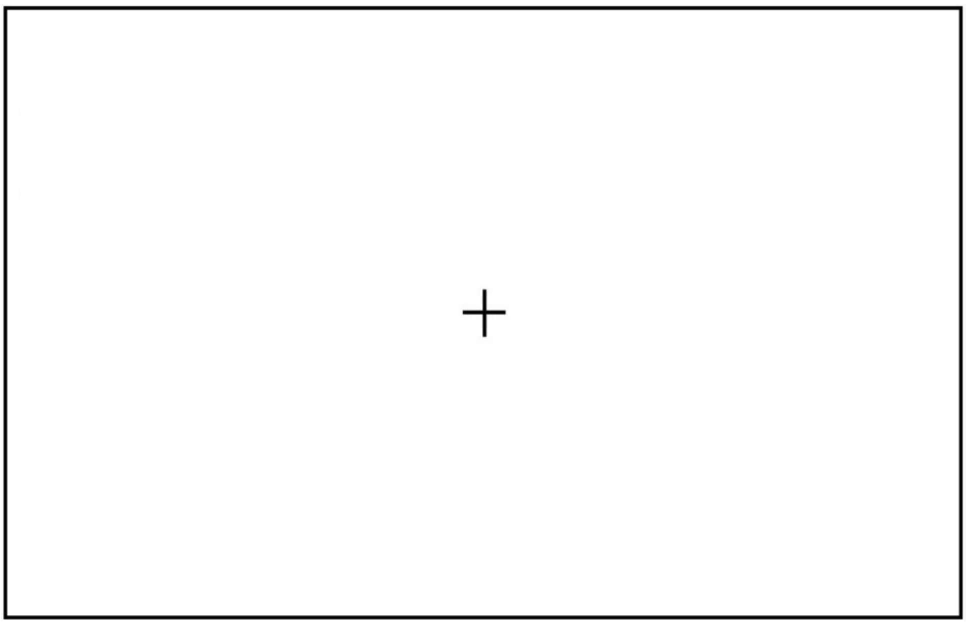

Figure 7.5 The interval screen between each two stimuli

7.1.4 Data Pre-Processing and Feature Selection

While each test was taken, the Ogama software would record the data about the tester's gaze and fixation behaviors and store them in its database. Gaze movement data with the following attributes were extracted and output into Excel sheets: SubjectName, TrialName, TrialCategory, TrialStartTime, Time, PupilDiaX and PupilDiaY. Here are the descriptions of the attributes:

-"SubjectName" stands for the name of the tester, and every tester is named in the format of "Subject"+ number, for example "Subject01";

-"TrialName" stands for the index number of the stimulus, which is a number ranged from 1 to 20;

-"TrialCategory" stands for the category which the current stimulus belongs to;

-"TrialStartTime" stands for the time stamp of the moment when the current stimulus started to display;

-"Time" stands for the time stamp of the moment when the current gaze started;

-"PupilDiaX" stands for the diameter of the left pupil of the tester;

-"PupilDiaY" stands for the diameter of the right pupil of the tester.

In addition to gaze movement data, fixation behavioral data were also extracted from the database. Apart from the basic attributes about the tester's name, the stimulus' category etc., the following attributes were extracted: CountInTrial, StartTime, Length, PosX, PosY. Here are the descriptions of the fixation attributes:

-"CountInTrial" stands for the number of fixations in the current stimulus;

-"StartTime" stands for the time stamp of the current fixation;

-"Length" stands for its duration (in millisecond);

-"PosX" stands for the horizontal parameter of the current fixation;

-"PosY" stands for the vertical parameter of the current fixation.

Before starting the analysis, a pre-process was conducted on the gaze movement data files to remove all the abnormal lines with lost data which were caused by the error of the device at those moments. Based on the primary attributes above, the following features were calculated:

(1) Pupil size features: PX_Mean, PX_StDev, PX_Median, PY_Mean, PY_StDev, PY_Median, ABS(PX-PY)_Mean, ABS(PX-PY)_StDev, ABS(PX-PY)_Median, PX/PY_Mean, PX/PY_StDev, PX/PY_Median, PX*PY_Mean, PX*PY_StDev, PX*PY_Median. Here are the descriptions of the pupil size features:

-"PX_Mean" stands for the mean value of the diameters of the left pupil;

-"PX_StDev" stands for the standard deviation of the diameters of the left pupil;

-"PX_Median" stands for the median value of the diameters of the left pupil;

-"PY_Mean" stands for the mean value of the diameters of the right pupil;

-"PY_StDev" stands for the standard deviation of the diameters of the right pupil;

-"PY_Median" stands for the median value of the diameters of the right pupil;

-"ABS(PX-PY)_Mean" stands for the mean value of the absolute values of the difference between the diameters of the left pupil and that of the right pupil;

-"ABS(PX-PY)_StDev" stands for the standard deviation of the absolute values of the difference between the diameters of the left pupil and that of the right pupil;

-"ABS(PX-PY)_Median" stands for the median value of the absolute values of the difference between the diameters of the left pupil and that of the right pupil;

-"PX/PY_Mean" stands for the mean value of the ratios of the diameters of the left pupil and that of the right pupil;

-"PX/PY_StDev" stands for the standard deviation of the ratios of the diameters of the left pupil and that of the right pupil;

-"PX/PY_Median" stands for the median value of the ratios of the diameters of the left pupil and that of the right pupil;

-"PX*PY_Mean" stands for the mean value of the products of the diameters of the left pupil and that of the right pupil;

-"PX*PY_StDev" stands for the standard deviation of the products of the diameters of the left pupil and that of the right pupil;

-"PX*PY_Median" stands for the median value of the products of the diameters of the left pupil and that of the right pupil.

(2) Fixation features: TFD, FD_Mean, FD_Median, NOF, TFD/NOF, SPL, SPL/NOF.

-“TFD” stands for “Total Fixation Duration”, calculated by adding up all the values of “Length” of one fixation;

-“FD_Mean” stands for the mean value of the fixation duration;

-“FD_Median” stands for the median value of the fixation duration;

-“NOF” stands for “Number of Fixations”, directly obtained from the maximum value of the “CountInTrial” value of one fixation;

-“TFD/NOF” stands for the average fixation duration, calculated from the ratio of the TFD value and the NOF value;

-“SPL” stands for “Scan Path Length”, calculated by computing the sum-up of the Euclidean distances of each pair of PosX and PosY values;

-“SPL/NOF” stands for the average speed of the fixation, calculated from the ratio of the SPL value and the NOF value.

When calculating fixation features, two methods were employed: in the first method, for each of the 20 stimuli, the fixation features of the whole image were calculated; in the second method, some “Areas of Interest” (AOIs) were defined in the images by Ogama’s “AOI” module, and each of the four categories of stimuli follows the same way to define the AOI. Thus, fixation features would be calculated for the fixations which fell in each of the AOIs. The way of the AOI definitions of the four categories are presented in Figure 7.6-7.9.

In Category “Animals”, five AOIs (A, B, C, D and E, respectively) enclosed the areas of the head, the fore legs, the hind legs, the tail and the torso of the animal, as shown in Figure 7.6.

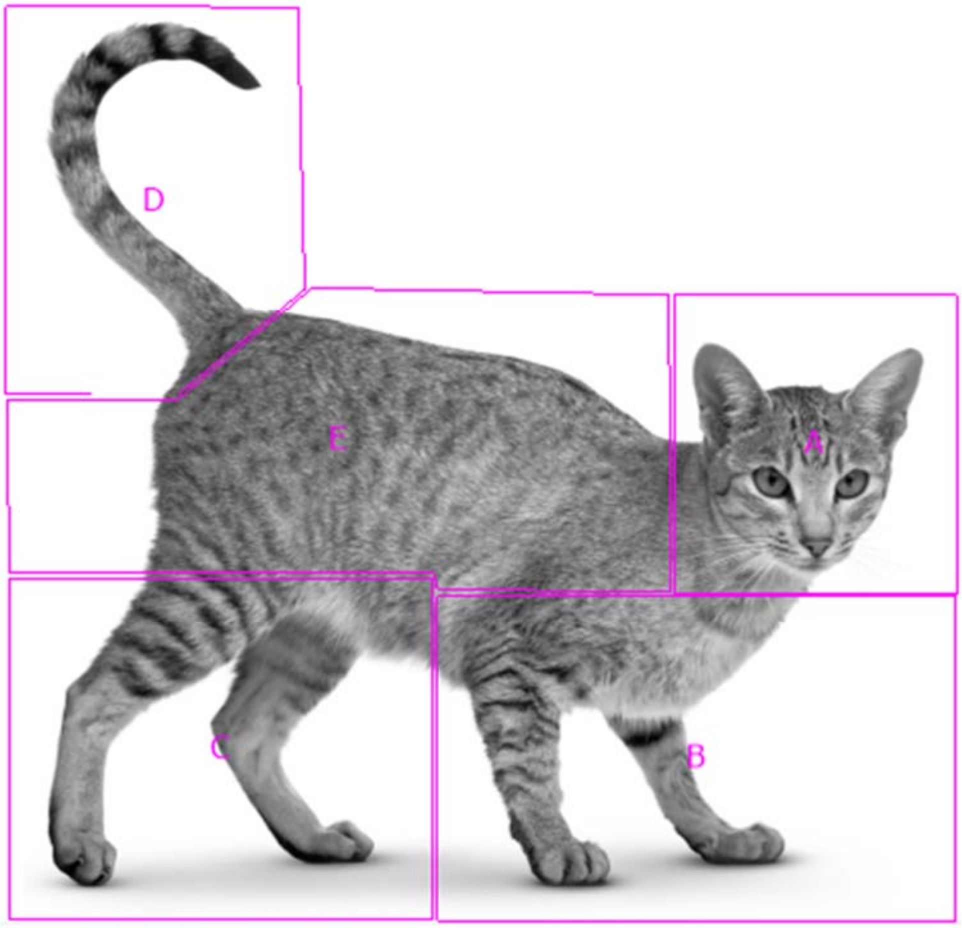

Figure 7.6 AOI definition of Category "Animals"

In Category "Cars", six AOIs (A, B, C, D, E and F, respectively) enclosed the areas of the fore wheel, the hind wheel, the headlights, the front seat and door, the back seat and door, and the taillights of the car, as shown in Figure 7.7.

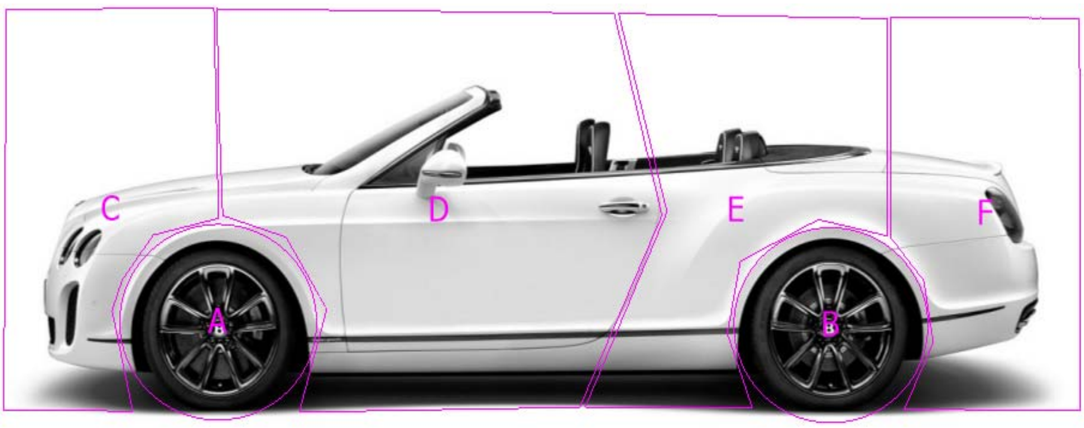

Figure 7.7 AOI definitions of Category "Cars"

In Category “Scissors”, five AOIs (A, B, C, D, and E, respectively) enclosed the areas of the two sharp edges, the axis, and the two handles of the scissors, which is shown in Figure 7.8.

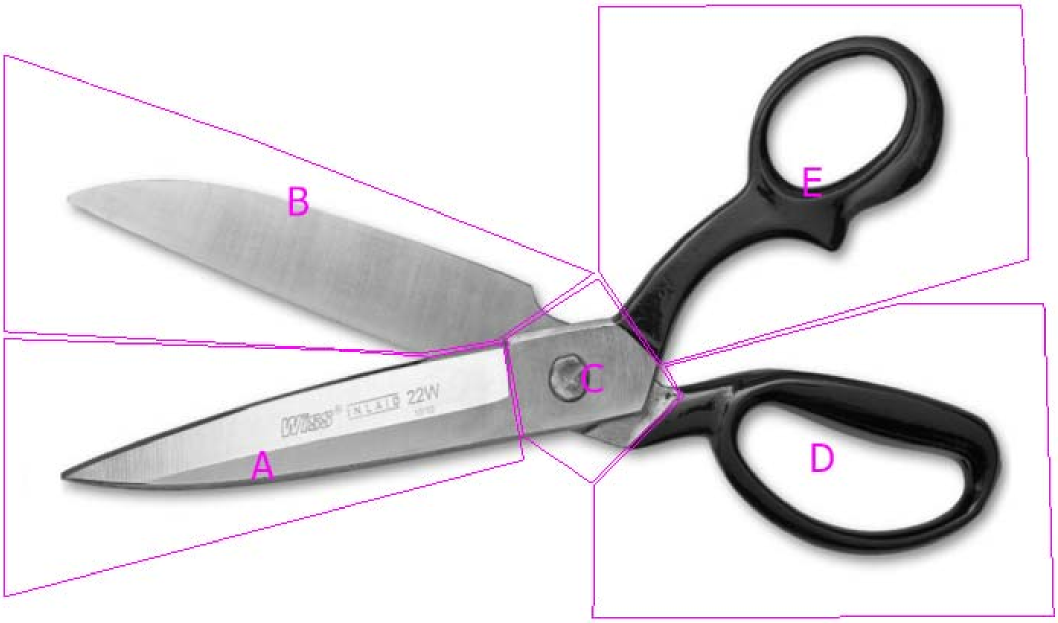

Figure 7.8 AOI definitions of Category “Scissors”

In Category “Objects”, four AOIs (A, B, C and D, respectively) enclosed the four objects presented in the image, as shown in Figure 7.9.

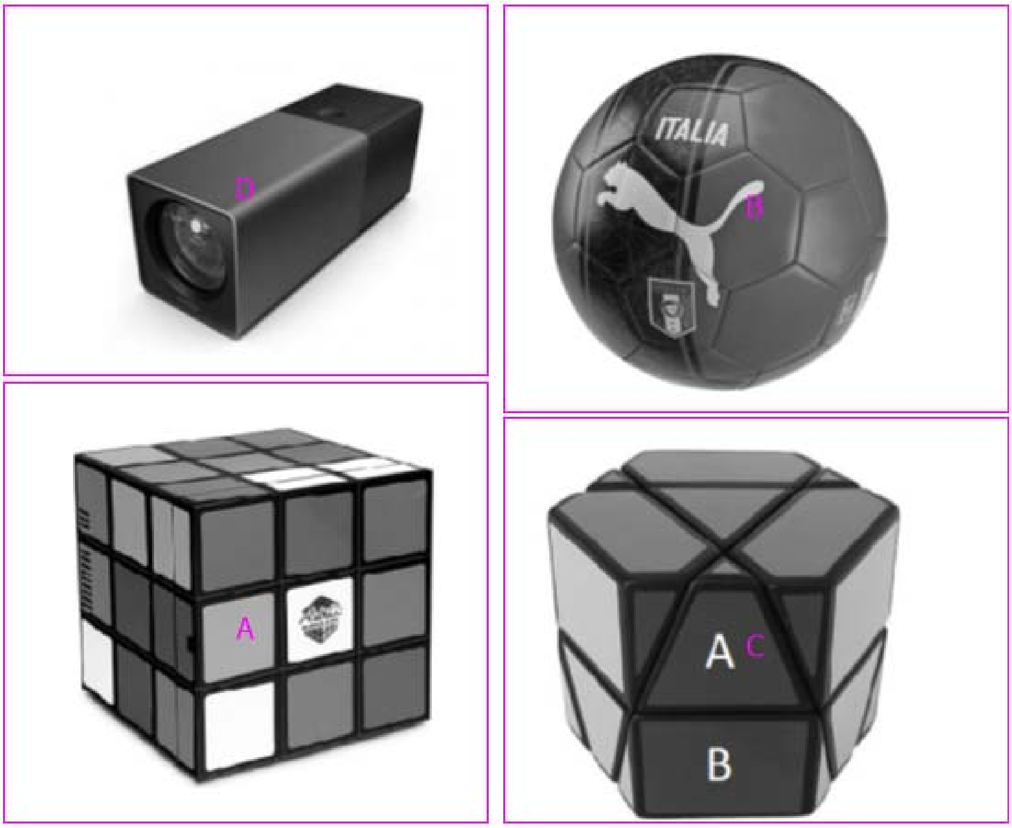

Figure 7.9 AOI definitions of Category “Objects”

7.2 Analysis and Results

7.2.1 Identification

In the identification stage, the “SubjectName” label was set as the classification target, which was assigned to each feature vector obtained from the pre-processed data. For both of the two analysis methods, the same five classifiers used for the identification analysis were employed, i.e. Classification Tree, Random Forest, SVM, Naïve Bayes and AdaBoost. The same two types of data sampling methods were adopted, i.e. 70-30 random sampling and 10-fold cross validation. Table 7.1 shows the results obtained with the first analysis method, i.e. considering fixation features of the whole images.

Table 7.1 Results for identification considering the whole images

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Classification Tree	0.636	Classification Tree	0.659
Random Forest	0.766	Random Forest	0.801
SVM	0.797	SVM	0.820
Naïve Bayes	0.618	Naïve Bayes	0.634
AdaBoost	0.643	AdaBoost	0.636

As can be seen from the results, when considering the whole images, both in random sampling and cross validation, the highest CAs were provided by SVM respectively with the value of 0.797 (in random sampling) and 0.820 (in cross validation) while the lowest CA were provided by Naïve Bayes respectively with the value of 0.618 (in random sampling) and 0.634 (in cross validation). Among the other classifiers, Random Forest was also good with slightly lower CA than SVM, respectively 0.766 in random sampling and 0.801 in cross validation; the other two classifiers namely Classification Tree and AdaBoost performed mediocrely with slightly higher CA than Naïve Bayes.

Table 7.2 shows the CA obtained with the second analysis method, i.e. considering the fixation features in the defined AOIs.

Table 7.2 Results for identification considering the AOIs

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Classification Tree	0.701	Classification Tree	0.700
Random Forest	0.845	Random Forest	0.832
SVM	0.839	SVM	0.851
Naïve Bayes	0.682	Naïve Bayes	0.674
AdaBoost	0.709	AdaBoost	0.716

The results demonstrate that when considering each of the AOIs, Random Forest provided the highest CA in random sampling with the value of 0.845, and SVM provided the highest CA in cross validation with the value of 0.851. Naïve Bayes provided the lowest CA in both random sampling (with the value of 0.682) and cross validation (with the value of 0.674). Among all the results of the identification stage, SVM produced the best CA with the value 0.851 in 10-fold cross validation by the second analysis method. Comparing the results of the two analysis methods, it can be seen that all the classifiers achieved a slightly increased CA when considering the fixation features in each of the defined AOIs.

7.2.2 Authentication

In the authentication stage, feature vectors were labelled with the binary value of "True" or "False", to describe "the true target" and the "imposters". The same features used for identification were employed. The Neural Network (NN) method was adopted in addition to the other five classifiers. Three types of data sampling were applied: 70-30 random sampling and 10-fold cross validation.

The authentication procedure iterates for ten times: for each tester T_i, feature vectors from the other testers are randomly selected, so that the same numbers of feature vectors for the "legible target" (i.e. tester T_i) and for "the others" are employed. The CA for T_i is given by the average CA calculated over the ten iterations. The same is done for the other testers, and the final CA for all testers is given by the average CA calculated for each tester.

Table 7.3 presents the results obtained with the first analysis method, i.e. to considering fixation features of the whole images.

Table 7.3 Results for authentication considering the whole images

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Naïve Bayes	0.803	Naïve Bayes	0.811
Random Forest	0.889	Random Forest	0.898
Neural Network	0.836	Neural Network	0.886
AdaBoost	0.821	AdaBoost	0.843
Classification Tree	0.843	Classification Tree	0.839
SVM	0.884	SVM	0.920

When considering the whole images, Random Forest provided the highest CA in random sampling with the value of 0.889, while SVM provided the highest CA in cross validation with the value of 0.920. Both in random sampling and cross validation, Naïve Bayes provided the lowest CA, with the respective value of 0.803 and 0.811. Table 7.4 shows the results obtained by considering the fixation features of each of the defined AOIs.

Table 7.4 Results for authentication considering the AOIs

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Naïve Bayes	0.854	Naïve Bayes	0.846
Random Forest	0.907	Random Forest	0.904
Neural Network	0.890	Neural Network	0.881
AdaBoost	0.873	AdaBoost	0.867
Classification Tree	0.844	Classification Tree	0.846
SVM	0.933	SVM	0.925

As is shown in the table, both in random sampling and cross validation, SVM was the best, with the CA value of 0.933 (random sampling) and 0.925 (cross validation). In random sampling, Classification Tree produced the lowest CA of 0.844, while in cross validation, Naïve Bayes and Classification Tree produced approximately equal values, with the lowest value of 0.846. Among all the results of the authentication stage, SVM produced the highest value of 0.933 with random sampling when considering fixation features in each of the AOIs. Besides, considering the AOIs produces slightly better results than considering the whole images.

7.3 Summary of the Results

In this case study, the feasibility of using gaze behaviors as a biometric method was investigated through a case of static image observation. Two analysis methods, i.e. considering the fixation features of the whole images and considering that of each defined AOI in each image, were conducted to detect the aims of identification and authentication of a subject.

In identification stage, when considering the whole images, some acceptable results could be obtained by adopting proper sampling methods and classifiers; when considering the AOIs, results presented to be generally better.

In authentication stage, with both analysis methods, most of the results obtained were beyond acceptance. When considering the whole image, excellent results over 0.9 could be achieved by adopting proper sampling methods and classifiers; when considering the AOIs, the CA presented to be totally improved and excellent values could be obtained with more probabilities than that in the first analysis method.

Chapter 8 *Case Study on Gaze-based Biometrics through Moving Target Observation*

8.1 Experiment Design and Data Acquisition

8.1.1 Participants of experiments

The experiment was designed late in 2016 and carried out in early 2017 with three repetitions. Forty-three testers aged from 19 to 68 participated the experiments, 32 of whom took at least two sessions of the experiment, and 18 of whom took the third session as well. Each two sessions were separated with at least two weeks.

8.1.2 Apparatus

The experiment material was generated by Microsoft C#, which offered the stimuli for the testers to observe. The tests were carried out using an Eye Tribe eye tracker, with a frequency of 30 Hz. The raw data about eye movements were extracted by exploiting a C# program. Gaze and fixation data were pre-processed by Microsoft Excel and a Python script. The identification stage was carried out by means of Orange Canvas, and the authentication stage was carried out through a series of Python scripts.

8.1.3 Experiment Procedure

The experiments took place in the Computer Vision and Multimedia Laboratory of the University of Pavia, in a quiet and light-friendly environment.

Animations generated by the C# program were employed as the stimuli, displayed on a screen with a full HD resolution (1920×1080). The whole animation lasted 27 seconds, and was composed of three 9-second phases. A short calibration was preliminarily carried out to adjust the eye tracker's parameter to the specific subject.

Initially, an orange circle with a diameter of 150 pixels (Figure 8.1) appeared and moved according to a random path; after 9 seconds, a blue square with an edge of 150 pixels would appear (Figure 8.2), and it too moved randomly, with the circle continuing its motion; in the last 9-second phase, a green triangle would appear (Figure 8.3), and move randomly like the first two shapes.

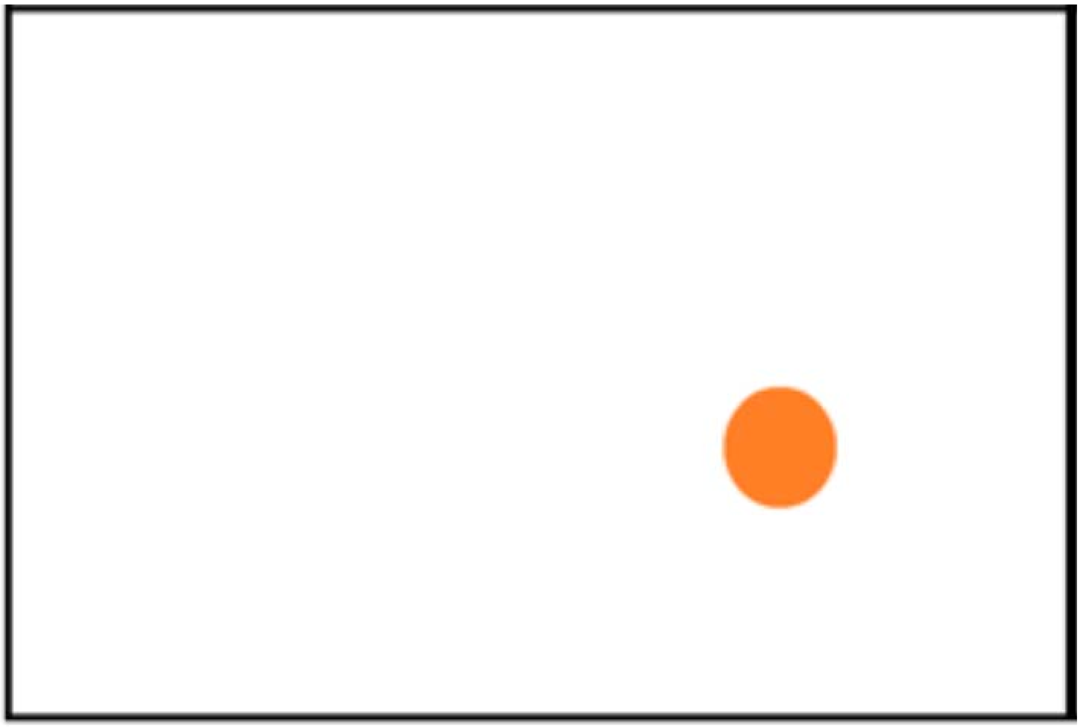

Figure 8.1 Phase 1: circle

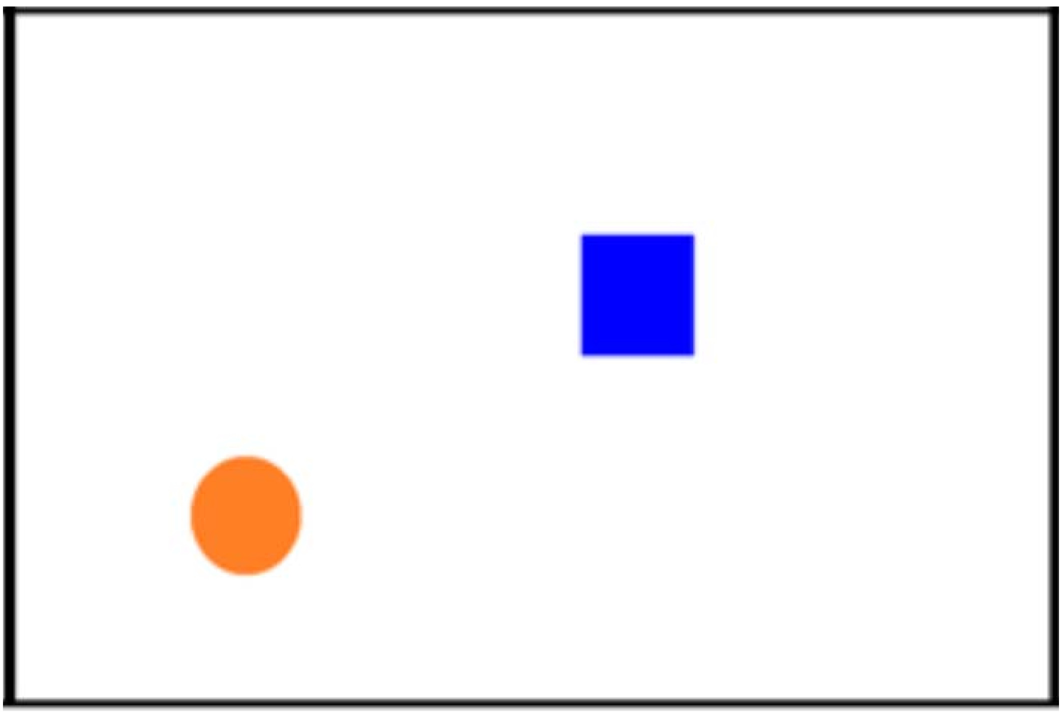

Figure 8.2 Phase 2: circle + square

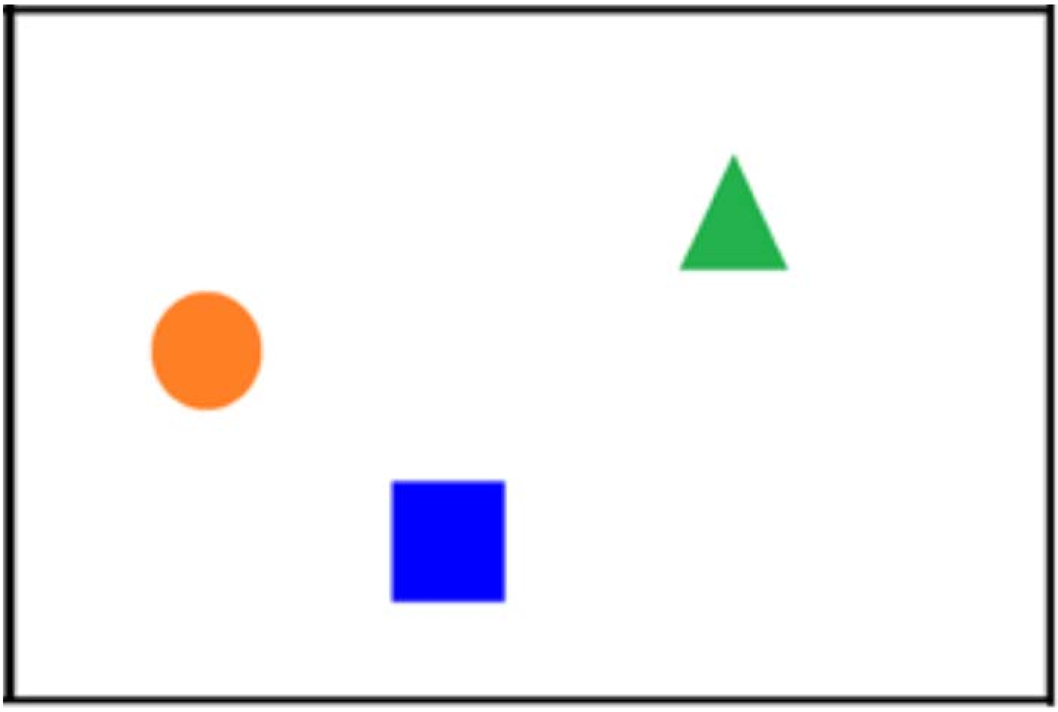

Figure 8.3 Phase 3: circle + square + triangle

8.1.4 Data Pre-Processing and Feature Selection

During each test, a C# program would record the raw data, which included the following attributes: Subject, LeftPupilSize, RightPupilSize, circle, square, triangle, distCircleCenter, distSquareCenter, distTriangleCenter, TimeStamp, RawGazeX, RawGazeY. Each of the attributes are described as follows:

-“Subject” stores the designation of each tester;

-“LeftPupilSize” stands for the diameter of the left pupil size;

-“RightPupilSize” stands for the diameter of the right pupil size;

-“circle” is a binary flag to mark if the current gaze sample falls in the circle;

-“square” is a binary flag to mark if the current gaze sample falls in the square;

-“triangle” is a binary flag to mark if the current gaze sample falls in the triangle;

-“distCircleCenter” stands for the distance of the current gaze sample position from the circle center;

-“distSquareCenter” stands for the distance of the current gaze position from the square center;

-“distTriangleCenter” stands for the distance of the current gaze position from the circle center;

-“isFixated” is a binary flag marking whether the current gaze sample belongs to a fixation;

-“TimeStamp” records the time stamp of the current gaze sample;

-“RawGazeX” stands for the horizontal coordinate of the current gaze sample;

-“RawGazeY” stands for the vertical coordinate of the current gaze sample.

Based on the basic raw data listed above, suppose PX is the left pupil size and PY is the right pupil size. Some measurements about pupil size were calculated: ABS(PX-PY), stands for the absolute value of the difference between the both pupil sizes; PX/PY, stands for the ratio between pupil sizes; PX*PY, stands for the product of pupil sizes. Mean value, standard deviation and median of PX and PY, as well as the above measurements were used as pupil features.

An algorithm was implemented for identifying fixations. The algorithm startsed by searching for the first three consecutive gaze samples, $S_1(X_1, Y_1)$, $S_2(X_1, Y_2)$, $S_3(X_3, Y_3)$, which were within a circle whose radius was 20 pixels. Then, the average X and Y values of S_1, S_2 and S_3 were computed, and a new point $C(\overline{X}, \overline{Y})$ was defined. When a next sample $S_4(X_4,Y_4)$ was acquired, if the distance between $C(\overline{X}, \overline{Y})$ and $S_4(X_4,Y_4)$ was less than or equal to 20 pixels, the gaze sample $S_4(X_4,Y_4)$ was considered part of the fixation and a new point $C_1(X_{C_1},Y_{C_1})$ was computed. X_{C_1} was obtained as the average of X_1, X_2, X_3, X_4, and Y_{C_1} was obtained as the average of Y_1, Y_2, Y_3, Y_4. The procedure was repeated this way. Instead, in case S_4 was not part of the fixation, the procedure started again considering samples S_2, S_3, S_4, always checking whether they were inside a circle with radius 20 pixels.

Based on the algorithm above, the value of the attribute “isFixated” of each gaze sample was obtained. In addition, some other features about fixations were calculated: FixationDuration, FixationCount, FixationInCircle, FixationInSquare, FixationInTriangle, distCircleCenter, distSquareCenter, distTriangleCenter. Here are their descriptions:

-“FixationDuration” is the duration of each fixation, which is calculated by the difference between the lst and first “TimeStamp” within a fixation. The mean, standard deviation and median of the “FixationDuration” were calculated as features.

-“FixationCount” stands for the number of fixations, which is counted based on the algorithm of the fixation stated above.

-“FixationInCircle” stands for the number of fixations which fall in the circle.

-“FixationInSquare” stands for the number of fixations which fall in the square.

-“FixationInTriangle” stands for the number of fixations which fall in the triangle.

Besides, the mean value, standard deviation and median value of “distCircleCente”, “distSquareCenter” and “distTriangleCenter” were calculated as features.

A Python script was written to pre-process the raw data, calculate the features, and establish the train data document. Occasionally, in the raw data there were some bad or lost data, . The script excluded these data before feature calculation.

8.2 Analysis and Results

Three types of analysis were performed, according to which the data were divided into some intervals and the feature vectors would be calculated for each of these intervals. In the first method, the 27-second data were subdivided into 6 intervals, each of which lasted 4.5 seconds, with 135 gaze samples; in the second method, the data were subdivided into 9 intervals, each of which lasted 3 seconds, with 90 gaze samples; in the third method, lastly, the data were subdivided into 18 intervals, each of which lasted 1.5 seconds, with 45 samples.

8.2.1 Identification

In the identification stage, the “Subject” label was set as the classification target, which was assigned to each feature vector obtained from the pre-processed data. For each of the three analysis methods, the same five classifiers used for the identification analysis were employed, i.e. Classification Tree, Random Forest Classification, SVM, Naïve Bayes and AdaBoost. The same two types of data sampling methods were adopted, i.e. 70-30 random sampling and 10-fold cross validation. Table 8.1 shows the

Classification Accuracy (CA) obtained with the first analysis method, i.e. 6 intervals lasting 4.5 seconds each.

Table 8.1 Results for identification with 6-interval data

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Classification Tree	0.690	Classification Tree	0.666
Random Forest	0.719	Random Forest	0.750
SVM	0.597	SVM	0.670
Naïve Bayes	0.456	Naïve Bayes	0.527
AdaBoost	0.695	AdaBoost	0.765

When subdividing the gaze samples with 6 intervals, Random Forest provided the highest CA, with the value of 0.719 with random sampling, and AdaBoost provided the highest CA with the value of 0.765 with cross validation. The other classifiers performed mediocrely, with values below 0.7, among which Naïve Bayes provided the lowest CA both in random sampling (as 0.456) and cross validation (0.527). With cross validation, Random Forest provided a CA value of 0.750.

Table 8.2 shows the CA results obtained with the second analysis method, i.e. subdividing gaze samples every 3 seconds, with 9 intervals.

Table 8.2 Results for identification with 9-interval data

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Classification Tree	0.668	Classification Tree	0.679
Random Forest	0.737	Random Forest	0.770
SVM	0.626	SVM	0.741
Naïve Bayes	0.445	Naïve Bayes	0.507
AdaBoost	0.695	AdaBoost	0.670

When subdividing the gaze samples with 9 intervals, Random Forest provided the highest CA in both random sampling (with the value of 0.737) and cross validation (with the value of 0.770); Naïve Bayes provided the lowest CA, with the value of 0.445 in random sampling and 0.507 in cross validation, which were even lower than the results when considering the 6-interval data. For most of the other classifiers, CA increased slightly compared to the first analysis method. For few classifiers, the CA was even lower than that of the first analysis method, besides Naïve Bayes, such as Classification Tree in random sampling (from 0.690 to 0.668) and AdaBoost in cross validation (from 0.765 to 0.670). Table 8.3 shows the CA results obtained with the third analysis method, i.e. subdividing gaze samples every 1.5 seconds, with 18 intervals.

Table 8.3 Results for identification with 18-interval data

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Classification Tree	0.657	Classification Tree	0.746
Random Forest	0.803	Random Forest	0.834
SVM	0.721	SVM	0.737
Naïve Bayes	0.615	Naïve Bayes	0.619
AdaBoost	0.664	AdaBoost	0.728

It can be seen from the table that Random Forest performed with the highest and acceptable CA in both random sampling (with the value of 0.803) and cross validation (with the value of 0.834), which were fairly acceptable. Naïve Bayes performed with the lowest CA in both random sampling (with the value of 0.615) and cross validation (with the value of 0.619), however notably increased compared to the first and second analysis method. For the other classifiers, the CA in the third method were generally improved compared to the second analysis method, while few classifiers decreased slightly, such as Classification Tree in random sampling (from 0.668 to 0.657), AdaBoost in random sampling (from 0.695 to 0.664), and SVM in cross validation (from 0.741 to 0.737). Among all the results of the identification stage, Random Forest provided the highest CA in cross validation when adopting the third analysis method, i.e. subdividing gaze samples every 1.5 seconds with 18 intervals.

8.2.2 Authentication

In the authentication stage, feature vectors were labelled with the binary values of "True" or "False", to describe "the true target" and the "imposters". The same features used for identification were employed. The Neural Network (NN) method was adopted in addition to the other five classifiers. Three types of data sampling were applied: 70-30 random sampling, 10-fold cross validation, and test on test data (i.e., the data of the first and second sessions used for training and that of the third session for testing).

The following procedure was iterated for ten times: for each tester T, feature vectors from the other testers were randomly selected, so that the same numbers of feature vectors for the "legible target" (i.e., the tester T) and for “the others” were employed. The CA for T was given by the average CA calculated over the ten iterations. The same was done for the other testers, and the final CA for all testers was given by the average CA calculated for each tester. Table 8.4 shows the results obtained with the first analysis method, i.e. subdividing the gaze samples every 4.5 seconds, with 6 intervals.

Table 8.4 Results for authentication with 6-interval data

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Classification Tree	0.753	Classification Tree	0.726
Random Forest	0.784	Random Forest	0.818
Neural Network	0.746	Neural Network	0.757
SVM	0.752	SVM	0.731
Naïve Bayes	0.697	Naïve Bayes	0.674
AdaBoost	0.758	AdaBoost	0.835

When subdividing the gaze samples with 6 intervals, Random Forest provided the highest CA, with the value of 0.784 in random sampling, and AdaBoost provided the highest CA in cross validation, with the value of 0.835. Naïve Bayes performed with the lowest CA both in random sampling (0.697) and cross validation (0.674). Most of the other classifiers provided CA between 0.7 and 0.8, except Random Forest, which performed with the CA of 0.818 in cross validation. Table 8.5 shows the results obtained with the second analysis method, i.e. to subdividing the gaze samples every 3 seconds, with 9 intervals.

Table 8.5 Results for authentication with 9-interval data

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Classification Tree	0.759	Classification Tree	0.741
Random Forest	0.804	Random Forest	0.840
Neural Network	0.754	Neural Network	0.762
SVM	0.783	SVM	0.808
Naïve Bayes	0.705	Naïve Bayes	0.693
AdaBoost	0.768	AdaBoost	0.781

When subdividing the gaze samples with 9 intervals, Random Forest performed with the best CA in both random sampling (with the value of 0.804) and cross validation (with the value of 0.840). Naïve Bayes performed with the lowest CA in both random sampling (with the value of 0.705) and cross validation (with the value of 0.693). The other classifiers provided CA between 0.7 and 0.8 except SVM, which provided the CA of 0.808 in cross validation. Besides, in this analysis, the CA of all the classifiers were improved compared to that of the first method with 6-interval data. Table 8.6 shows the results obtained with the third analysis method, i.e. subdividing the gaze samples every 1.5 seconds, with 18 intervals.

Table 8.6 Results for authentication with 18-interval data

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Classification Tree	0.767	Classification Tree	0.753
Random Forest	0.876	Random Forest	0.890
Neural Network	0.807	Neural Network	0.815
SVM	0.787	SVM	0.814
Naïve Bayes	0.771	Naïve Bayes	0.776
AdaBoost	0.784	AdaBoost	0.794

When considering the 18-interval data, Random Forest provided the highest CA in both random sampling (0.876) and cross validation (0.890). Classification Tree provided the lowest CA in random sampling with value of 0.767, and Naïve Bayes provided the lowest CA in cross validation with the value of 0.776. Most of the other classifiers performed with CA between 0.75 and 0.8, except SVM, which provided the CA of 0.814 in cross validation. Among all the results of the authentication stage, Random Forest provided the best CA in cross validation when adopting the third analysis method, i.e. subdividing the gaze samples every 1.5 seconds with 18 intervals, with the value of 0.890, which can be regarded as "decidedly good". Besides, the CA of all classifiers were slightly improved compared to the second method and certainly to the first method as well.

8.2 Summary of the Results

In this case study, the feasibility of using gaze behaviors as a biometric method was investigated by considering moving target observation. Two aims, namely identification and authentication of a subject, were carried out by three analysis methods, i.e. subdividing gaze samples every 4.5 seconds with 6 intervals, every 3 seconds with 9 intervals, and every 1.5 seconds with 18 intervals.

In the identification stage, considering the 18-inverval data, good results above 0.8 could be obtained when adopting proper classifiers; in the authentication stage, performances were better for all classifiers in both sampling methods, with all the three analysis methods.

Chapter 9 *Case Study on Gaze-based Biometrics Through Eye-Driven PIN Input*

9.1 Experiment Design and Data Acquisition

9.1.1 Participants of experiments

The experiments was designed early in 2017 and carried out from April 2017 to the beginning of May 2017 with three repetitions. 45 testers who were aged from 21 to 68 participated in the experiment, 33 of whom took at least two sessions of the experiment, and 23 took the third session. Consecutive sessions were separated by at least two weeks.

9.1.2 Apparatus and Tools

The experiment material was generated with Microsoft C#, which offers the targets for the testers to observe and interact. The tests were carried out using an Eye Tribe eye tracker, with a frequency of 30 Hz. Raw data about eye movements were extracted by a C# program. Gaze and fixation data were pre-processed with Microsoft Excel and Python scripts. The identification stage was carried out with Orange Canvas, while the authentication stage was carried out using Python scripts.

9.1.3 Experiment Procedure

Experiments were carried out in the UniPV-CVML, in a quiet and light-friendly environment for eye tracking experiments.

A virtual PIN board interface was generated with C# and set up as the experiment material with the resolution of full HD (1920×1080). Figure 9.1 shows the PIN interface, composed of nine single-digit buttons (0-9) and the functional buttons “Cans” (as “Cancel”) and “Done”.

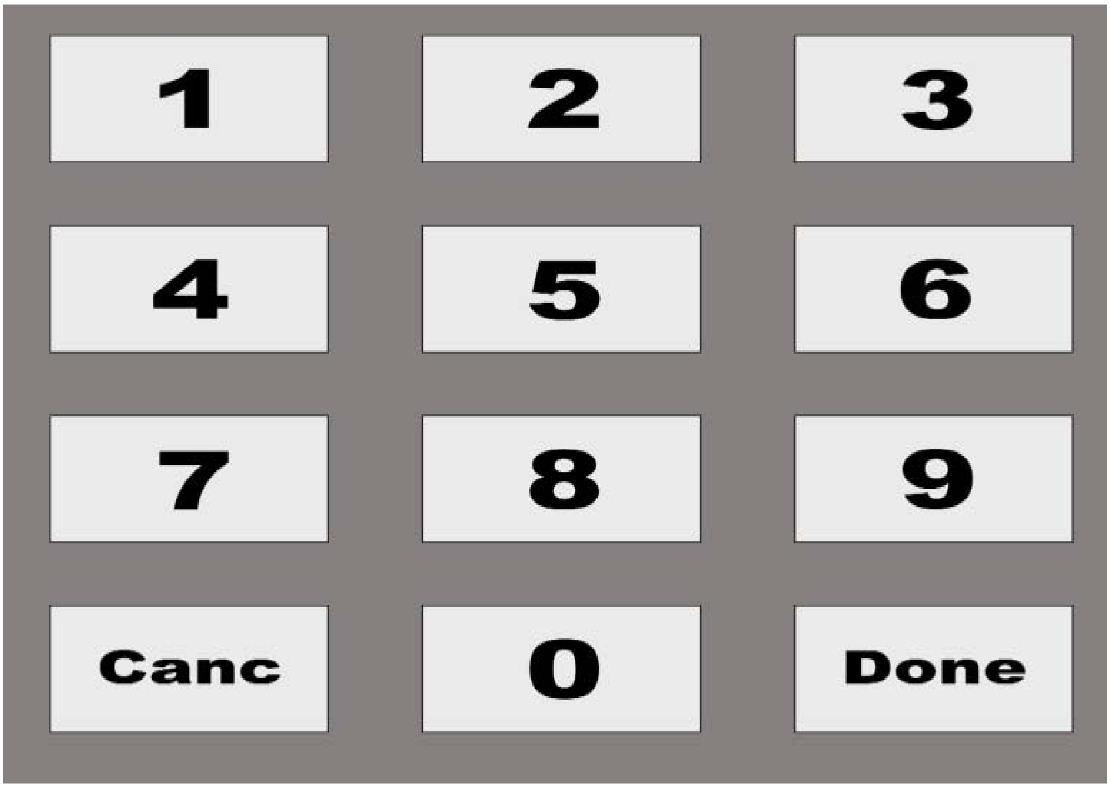

Figure 9.1 PIN board interface

A button on the PIN board could be "pressed" by gazing at it for two seconds. When a button was successfully pressed, a beep sound was played as a feedback signal. After a short calibration to adjust the eye tracker's parameters to the specific subject, the interface of the soft PIN board was displayed and the tester was instructed to input a six-digit PIN by gaze pressing. The "Canc" key could be pressed to delete the last digit entered when the user had pressed a wrong key, and the "Done" key would be pressed after finishing the input of the whole PIN. Each tester had to enter four PINs by dictation in one session of the experiment.

9.1.4 Data Pre-Processing and Feature Selection

During the test, a C# program recorded the following attributes: Timestamp, RawX, RawY, Fix, LPDiam, RPDiam. Here are their description:

-"Timestamp" stands for the current time, in milliseconds;

-"RawX" stands for the horizontal coordinate of the current gaze sample;

-"RawY" stands for the vertical coordinate of the current gaze sample;

-"Fix" stands for the tag which marks if a gaze sample belongs to a fixation as a

Boolean value;

-“LPDiam” stands for the diameter of the left pupil at the current gaze sample;

-“RPDiam” stands for the diameter of the right pupil at the current gaze sample.

Firstly, a pre-processing operation that removed all the abnormal lines with lost data was performed. Based on the primary attributes above, the following features were then calculated:

(1) Fixation features: FixCount, FixDurAvg, FixDurStDev. Here are the descriptions of the fixation features:

-“FixCount” stands for the number of fixations;

-“FixDurAvg” stands for the average value of fixation durations;

-“FixDurStDev” stands for the standard deviation of fixation durations.

F

(2) Pupil size features: PX, PY, StDevPX, StDevPY, MaxMinPX, MaxMinPY, NormPX, NormPY, PX/PY. Here are the descriptions of the pupil size features:

-“PX” stands for the mean value of the diameter of the left pupil;

-“PY” stands for the mean value of the diameter of the right pupil;

-StDevPX” stands for the standard deviation of the diameter of the left pupil;

-StDevPY” stands for the standard deviation of the diameter of the right pupil;

-“MaxMinPX” stands for the difference value between the maximum and minimum diameters of the left pupil;

-“MaxMinPY” stands for the difference value between the maximum and minimum diameters of the right pupil;

-“NormPX” stands for a normalized average value of the diameters of the left pupil, calculated by:

$$NormPX = \frac{Current(PX) - \min(PX)}{\max(PX) - \min(PX)}$$

-“NormPY” stands for a normalized average value of the diameters of the right pupil, calculated by:

$$NormPY = \frac{Current(PY) - \min(PY)}{\max(PY) - \min(PY)}$$

-“PX/PY” stands for the ratio of the left pupil average diameter and the right pupil average diameter.

(3) Saccade features. When two consecutive gaze samples were too distant from each other to belong one same fixation, they would be regarded as a "saccade". Three features related to saccades were calculated: SacCount, SacAvgDist, SacStDevDist. Here are the descriptions of the saccade features:

-“SacCount” stands for the number of saccades;

-“SacAvgDist” stands for the average value of the distance covered by saccades;

-“SacStDevDist” stands for the standard value of the distance covered by saccades.

(4) Time features: TotalTime, Reaction. Here are the descriptions of time features:

-“TotalTime” stands for the total time which the tester took to finish entering a six-digit PIN, and its value was directly extracted from the last “Timestamp” value generated after the PIN was entered;

-“Reaction” stands for the average reaction time that the tester needed to take after he was told the PIN and before he pressed the number button. It is calculated by:

$$Reaction = \frac{TotalTime - MinimumTime}{NumberOfButtons}$$

where “NumberOfButtons” stands for the six numeric buttons plus the “Done” button, and “MinimumTime” stands for the least time (in milliseconds) necessary for the tester to finish entering the whole PIN. Since the key press operation was defined as 2 seconds, the “MinimumTime” value should be 14 seconds.

Besides the features above, several features related to head behavior and physiological traits of the tester were also extracted and calculated. Considering a segment connecting both pupil centers, the midpoint of this segment is supposed to be the “head position”. Accordingly, the following head behavioral and physiological features were computed: HeadAngle, HeadXmaxmin, HeadXStDev, HeadYmaxmin, HeadYStDev, StdevLPXP, StdevLPYP, StdevRPXP, StdevRPYP, EyeDist. Here are their descriptions.

- HeadAngle stands for the angle formed by the segment connecting pupils and the horizontal axis, being positive when the head moves leftward and negative when the head moves rightward;

- HeadXmaxmin stands for the difference between the maximum and minimum values of the horizontal parameters of the head position;
- HeadXStDev stands for the standard deviation of the horizontal position of the head;
- HeadYmaxmin stands for the difference between the maximum and minimum values of the vertical position of the head;
- HeadYStDev stands for the standard deviation of the vertical position of the head;
- StdevLPXP stands for the standard deviation horizontal position of the left pupil;
- StdevLPYP stands for the standard deviation of the vertical position of the left pupil;
- StdevRPXP stands for the standard deviation of the horizontal position of the right pupil;
- StdevRPYP stands for the standard deviation of the vertical position of the right pupil.
- EyeDist stands for the average Euclidean distance between the eyes.

9.2 Analysis and Results

For data analysis, two types of methods were used to organize the data for both the identification and authentication stage. The first method is to take the whole procedure of the PIN entering as the data set, while the second method is to investigate each of the key separately.

9.2.1 Identification

In the identification stage, the “Tester” label was set as the classification target, which was assigned to each feature vector obtained from the pre-processed data. For both of the two analysis methods, the same five classifiers were employed, i.e. Classification Tree, Random Forest Classification, SVM, Naïve Bayes and AdaBoost. The same two types of data sampling methods were adopted, i.e. 70-30 random sampling and 10-fold cross validation. Table 9.1 shows the Classification Accuracy (CA) obtained with the first analysis method, i.e. considering the entire PIN entering procedure.

Table 9.1 Results for identification considering the entire PIN

70-30 Random Sampling	10-fold Cross Validation

Method	CA	Method	CA
Classification Tree	0.550	Classification Tree	0.586
Random Forest	0.690	Random Forest	0.736
SVM	0.799	SVM	0.816
Naïve Bayes	0.559	Naïve Bayes	0.609
AdaBoost	0.640	AdaBoost	0.664

When considering the entire PIN, in both 70-30 random sampling and 10-fold cross validation, SVM provided the best result, with the values of 0.799 for random sampling and 0.816 for cross validation; Classification Tree provided the lowest result, with the value of 0.550 for random sampling and 0.586 for cross validation.

Table 9.2 shows the CA obtained with the second analysis method, i.e. considering each of the keys separately.

Table 9.2 Results for identification considering each key separately

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Classification Tree	0.787	Classification Tree	0.827
Random Forest	0.773	Random Forest	0.796
SVM	0.826	SVM	0.856
Naïve Bayes	0.609	Naïve Bayes	0.636
AdaBoost	0.787	AdaBoost	0.830

When considering each of the keys separately, in both 70-30 random sampling and 10-fold cross validation SVM provided the highest CA, with the values of 0.826 for random sampling and 0.856 for cross validation; Naïve Bayes provided the lowest CA, with the value of 0.609 for random sampling and 0.636 for cross validation. Among all the analyses in the identification stage, the SVM method produced the highest CA value (0.856) with the 10-fold cross validation when considering each of the PIN button separately, and the Classification Tree method produced the lowest CA value (0.550) with the random sampling when considering the entire PIN. Besides, for all the classification methods, the CA values were higher in the second analysis method compared to that in the first analysis method.

9.2.2 Authentication

In the authentication stage, feature vectors were labelled with the binary values of “True” or “False”, to describe “the true target” and the “imposters”. The same features used for identification were employed. The Neural Network (NN) method was adopted in addition to the other five classifiers. Three types of data sampling were applied: 70-30 random sampling, 10-fold cross validation, and test on test data (i.e,. the data of the first and second sessions used for training and that of the third session for testing).

The following procedure iterates for ten times: for each tester T, feature vectors from the other testers are randomly selected, so that the same numbers of feature vectors for the "legible target"(i.e., tester T) and for “the others” are employed. The CA for T is given by the average CA calculated over the ten iterations. The same is done for the other testers, and the final CA for all testers is given by the average CA calculated for each tester. Table 9.3 shows the results of the analysis on the authentication considering the entire PIN.

Table 9.3 Results for authentication considering the entire PIN

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Naïve Bayes	0.762	Naïve Bayes	0.787
Random Forest	0.781	Random Forest	0.766
Neural Network	0.710	Neural Network	0.705
AdaBoost	0.822	AdaBoost	0.842
Classification Tree	0.707	Classification Tree	0.706
SVM	0.836	SVM	0.876

As can be seen from the table, when considering the entire PIN, SVM presented the best accuracy of classification in both random sampling (0.836) and cross validation (0.876); Classification Tree provided the lowest CA for random sampling with the value of 0.707, and Neural Network provided the lowest CA for cross validation with the value of 0.705.

Table 9.4 shows the CA obtained with the second analysis method, i.e. considering each of the PIN keys input separately.

Table 9.4 Results for authentication considering each key separately

70-30 Random Sampling		10-fold Cross Validation	
Method	CA	Method	CA
Naïve Bayes	0.836	Naïve Bayes	0.791
Random Forest	0.868	Random Forest	0.841
Neural Network	0.742	Neural Network	0.792
AdaBoost	0.897	AdaBoost	0.906
Classification Tree	0.814	Classification Tree	0.837
SVM	0.901	SVM	0.935

When considering each of the PIN input separately, in both random sampling and cross validation, SVM provided the highest result, with the value of 0.901 for random sampling and 0.935 for cross validation; Neural Network provided the lowest result in random sampling with the value of 0.742 and Naïve Bayes provided the lowest result in cross validation with the value of 0.801. Among all the analyses in the authentication stage, SVM provided the highest CA with the value of 0.935 for cross validation when considering each of the PIN key input separately, while Neural Network provided the lowest CA for cross validation with the value of 0.705 when considering the entire PIN input. Besides, for all the classification methods, the CA values were higher in the second analysis method compared to that in the first analysis method.

9.2 Summary of the Results

In the case study, the feasibility of using gaze behaviors as a biometric method was investigated by a case about eye-driven PIN input. Two aims namely the detection of the identification and authentication of a subject were carried out based on the two analysis methods i.e. to consider the entire PIN input and to separately consider each of the PIN keys input.

In identification stage, when considering the whole PIN input, good result above 0.8 could be obtained when adopting proper sampling method and classifier; when considering each PIN button separately, the performances were totally raised for all the classifiers in both sampling methods.

In authentication stage, performances were decidedly good compared to that of identification stage. Generally, the performance obtained when considering each PIN button separately yielded notably better than considering the entire PIN, with the best result above 0.9 for proper classifiers and sampling methods.

Chapter 10 *Conclusions*

In this thesis, the feasibility of implementing biometrics using eye tracking was studied and evaluated, for both the tasks of subject identification and authentication.

Drawing inspiration from studies about gaze-based biometrics carried out during the past decade, four case studies were designed, exploring eye behaviors as biometrics in different types of activities and scenes, such as building access control, static image observation, moving target observation, and soft PIN input.

In Chapter 6, the feasibility of using gaze behaviors as a biometric method was investigated by a case about building access security. Three aims characterized the study, namely the detection of the “familiarity level” with doorbell names, and the recognition and authentication of a subject. For familiarity identification, the best classification accuracy (CA) considering all the three trials of the experiment was not higher than 0.614, and therefore it was not possible to precisely distinguish the level of familiarity in the three trials, namely “initial visitor” (trial 1), “frequent visitor” (trial 2), or “familiar visitor” (trial 3). However, satisfactory results (0.86) were obtained when comparing trial 1 with trials 2 and 3 (together or separately). For person identification, maybe because of the limited number of samples per subject available, the identification results achieved only 60% of classification accuracy. For person authentication, the performance was decidedly better (apart from the “Test on Test Data” case), with the best CA of 0.888. According to this case study, we think that the gaze-based biometric approaches presented in this work can be promising if appropriate classifiers and data sampling methods are adopted, and sufficient data are collected.

In Chapter 7, a case study on static image observation was carried out to investigate the feasibility of using gaze behaviors as a biometric method. Two aims, namely subject identification and authentication, were pursued by means of twenty randomly displayed pictures which belonged to four categories. Two analysis methods were also used, i.e. considering fixation features of the whole images and considering only those of each defined AOI in each image. In identification stage, when considering the whole images, some acceptable results (with a best CA of 0.82) could be obtained by adopting proper sampling methods and classifiers; when considering the AOIs separately, results were generally better, with a best CA of 0.851. In the authentication stage, with both analysis methods, most of the results obtained were fairly acceptable. When considering the whole images, excellent results, over 0.9, could be achieved by adopting proper sampling methods and classifiers; when considering AOIs, the CA was improved and excellent values (with a maximum of 0.933).

In Chapter 8, a case study investigated the feasibility of using gaze behaviors as a biometric method with moving target observation. The identification and authentication of a subject were carried out through the observation of a 27-second animation with three colorful shapes moving according to a random path. Three methods were used for the analysis: subdividing gaze samples every 4.5 seconds with 6 intervals, subdividing gaze samples every 3 seconds with 9 intervals, and subdividing gaze samples every 1.5 seconds with 18 intervals. In the identification stage, considering the 18-inverval data, good results, with CA above 0.8, could be obtained when adopting proper classifiers， while the other two methods performed with lower CAs. In the authentication stage, the performance decidedly increased for all classifiers in both sampling methods with all three analysis methods, with a best CA of 0.890.

In Chapter 9, the feasibility of using gaze behaviors as a biometric method was investigated with a case about soft PIN input. Subject identification and authentication were pursued based on two analysis methods, i.e. considering the entire PIN input and separately considering each of the PIN keys. In the identification stage, when considering the whole PIN input, good results, with a CA above 0.8, could be obtained when adopting proper sampling methods and classifiers; when considering each PIN button separately, the performance decidedly increased for all classifiers in both sampling methods, with a best CA of 0.856. In the authentication stage, performances were very good compared to those of the identification stage. Generally, the performance obtained when considering each PIN button separately yielded notably better results than considering the entire PIN, with a best result of 0.935 for proper classifiers and sampling methods.

In summary, through the case studies presented in this thesis, we have demonstrated that eye behaviors recorded by an eye tracker during different observation tasks can be potentially exploited as biometric traits.

The results obtained in our experiments can suggest further research on gaze-based biometrics. Firstly, larger groups of the testers can guarantee better training procedures. Secondly, the training effectiveness can be improved by involving testers in multiple test sessions (possibly, with time spans between sessions of at least one week). Moreover, the use of more sophisticated eye trackers (with a higher frequency) would probably allow obtaining better results.

Further development can be focused on the application of gaze-based biometrics in commercial and public security fields. The advance of technology will likely provide smaller, portable and stable eye trackers, which could be incorporated into personal electronic devices such as laptop, tablet, smart phones, etc. Gaze-based biometrics has thus the potential to become a reliable and popular authentication method.

References

[1] Anil K. Jain, Patrick Flynn, Arun A. Ross. Handbook of Biometrics. Springer.2008. pp. 1.

[2] Koza J R, Bennett F H, Andre D, et al. Automated design of both the topology and sizing of analog electrical circuits using genetic programming[M]//Artificial Intelligence in Design'96. Springer, Dordrecht, 1996: 151-170.

[3] Jain A. K., Nandakumar K., and Ross A., 50 years of biometric research: Accomplishments, challenges, and opportunities, Pattern Recognit. Lett., vol. 79, pp. 80–105, 2016.

[4] Zhang D., Song F., Xu Y. and Liang Z., Advanced Pattern Recognition Technologies with Applications to Biometrics, Information Science Reference (IGI Global), First Edition, 2009.

[5] Angle S., Bhagtani R., Chheda H., Biometrics: A further echelon of security. Paper presented at the First UAE International Conference on Biological and Medical Physics, 2005, March 27-30.

[6] Dugelay J. L., Junqua J. C., Kotropoulos C., Kuhn R., Perronnin F., Pitas I., Recent advances in biometric person authentication. Paper presented at the IEEE Int. Conf. on Acoustics Speech and Signal Processing (ICASSP), Special Session on Biometrics, Orlando, FL, 2002.

[7] Lee K., Park H. A new similarity measure based on intraclass statistics for biometric systems. 2003: ETRI Journal, 25(5), 401–406. doi:10.4218/etrij.03.0102.0017.

[8] Lumini A. and Nanni L., Overview of the combination of biometric matchers, *Inf. Fusion*, vol. 33, pp. 71–85, 2017.

[9] Wang L., Behavioral Biometrics for Human Identification : Intelligent

Applications. 2010. pp.121
[10] Jain A. K., Ross A., and Prabhakar S.. An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, 14(1):4–20, January 2004.
[11] Jain A. K., Flynn P., Ross A. A. Handbook of Biometrics. Springer.2008. pp. 2.
[12] Jain A. K., Flynn P., Ross A. A. Handbook of Biometrics. Springer.2008. pp. 3.
[13] J. Ashbourn, Biometrics: Advanced Identity Verification: The Complete Guide, Springer, 2000.
[14] Jain A. K., Ross A. A., Nandakumar K. Introduction to Biometrics. Springer Science+Business Media, LLC.2011. pp. 12.
[15] Jain A. K., Ross A. A., Nandakumar K. Introduction to Biometrics. Springer Science+Business Media, LLC.2011. pp. 10.
[16] Unar J. A., Seng W. C., and Abbasi A., A review of biometric technology along with trends and prospects, Pattern Recognit., vol. 47, no. 8, pp. 2673–2688, 2014.
[17] Wang L., Behavioral Biometrics for Human Identification: Intelligent Applications. 2010. pp.121- pp.122
[18] The Database of Faces.[Online].
https://www.cl.cam.ac.uk/research/dtg/attarchive/facedata
base.html
[19] Figure From. [Online].
https://www.saudedica.com.br/wp-content/uploads/2017/08/
syndrome-do-triploX1.jpg
[20] Figure From. [Online].
https://www.techemergence.com/wp-content/uploads/2018/01/ai-in
-biometrics-and- security-current-business-applications.png
[21] Figure From. [Online].
https://cacm.acm.org/system/assets/0002/5621/111516_Messageto
EagleCom_eyescanning.large.jpg?1479234288&1479234287
[22] Wang L., Behavioral Biometrics for Human Identification: Intelligent Applications. 2010. pp.1- pp.3.

[23] Yampolskiy R. V, Human computer interaction based intrusion detection. Paper presented at the 4th International Conference on Information Technology: New Generations (ITNG 2007), Las Vegas, NA (pp. 837-842).
[24] Bioprivacy.org. FAQ's and definitions. International Biometric Group, LLC. Retrieved on October 2, 2005, from http://www.bioprivacy. org/bioprivacy_text.htm.
[25] Yampolskiy R. V., Govindaraju V., Dissimilarity functions for behavior- based biometrics. Paper presented at the Biometric Technology for Human Identification IV. 2007, SPIE De- fense and Security Symposium, Orlando, FL.
[26] Denning D. E, An intrusion-detec- tion model. IEEE Transactions on Software Engineering, 1987, 13(2), 222–232. doi:10.1109/ TSE.1987.232894
[27] Ilgun K., Kemmerer R. A., Porras P. A., State transition analysis: A rule-based intrusion detection approach. Software Engineering, 1995, 21(3), 181–199.
[28] Ghosh A. K., Schwatzbard A., Shatz M., Learning program behavior profiles for intrusion detection. Paper presented at the 1st USENIX Workshop on Intrusion Detection and Network Monitoring, 1999, Santa Clara, CA (pp. 51-62).
[29] Apap F., Honig A., Hershkop S., Eskin E., Stolfo S., Detecting malicious software by monitoring anomalous windows registry accesses. Paper presented at the Fifth International Symposium on Recent Advances in Intrusion Detection (pp. 16-18), 2002.
[30] Pennington A. G., Strunk J. D., Griffin J. L., Soules C. A. N., Goodson G. R., Ganger G. R., Storage-based intrusion detection: Watching storage activity for suspicious behavior. (No. CMU--CS-02-179). 2002, Carnegie Mellon University.
[31] Feng H. H., Kolesnikov O. M., Fogla P., Lee W., Gong W., Anomaly detection using call stack information. Paper presented at the IEEE Symposium on Security and Privacy (pp. 62-78), 2003.
[32] A. Garg, R. Rahalkar, S. Upadhyaya, K. Kwiat. (2006, June 21-23). Profiling users in GUI based systems for masquerade detection. Paper presented at The 7th IEEE Information Assurance Workshop (IAWorkshop 2006), West Point, NY.
[33] Pusara M., Brodley C. E., User reau- thentication via mouse movements. Paper presented at the ACM Workshop on Visualization and Data Mining for Computer

Security, 2004, Washington, D.C. (pp. 1-8).
[34] Yampolskiy R. V., Motor- skill based biometrics. Paper presented at the 6th Annual Security Conference, 2007, Las Vegas, NV.
[35] Caslon.com.au. (2005). Caslon-analytics. Retrieved on October 2, 2005, from http://www.caslon.com.au/biometricsnote8.htm
[36] http://careyhead.me/wp-content/uploads/2018/06/the-best-way-to-make-a-cool-signature- wikihow-3-letter-names.jpeg
[37] Figure From. [Online]. http://theeyetribe.com/dev.theeyetribe.com/lh5.googleusercontent. com/F3n0tQ1HWuJV8e_oA4pFx0dhhN3cAdDDRCg_JkwLOkhSL3Xw1goy6TVRE0V7at6CLw%3ds2000.png
[38] Figure From. [Online]. https://userscontent2.emaze.com/images/f12d56d4-6a68-4329-b79f-31fd4d300ed0/db9c5798-3818-4095-a85c-b49f876a768a.jpg
[39] Figure From. [Online]. https://cfatech-jdntqjj5tat2yt1zpj9.netdna-ssl.com/wp-content/ uploads/2015/12/keystroke_logging.jpg
[40] Choubisa T., Prasanna S. R. M., Sahoo S. K., Multimodal Biometric Person Authentication: A Review, IETE Technical Review, vol. 29, no. 1, pp. 54-75, 2012.
[41] Cherifi F., Hemery B., Giot R., Pasquet M., and Rosenberger C., Performance Evaluation of Behavioral Biometric Systems, in Behavioral Biometrics for Human Identification: Intelligent Applications.: Medical Information Science Reference, 2010.
[42] Widmaier E P, Raff H, Strang K T. Vander's human physiology: the mechanisms of body function[M]. McGraw-Hill Higher Education, 2008.
[43] Human Eye. [Online]. https://healthjade.com/human-eye/
[44] Figure From. [Online]. https://archive.li/lx5wy/a0add3518a76ac5cd20336b6104388b4441300 fe.gif
[45] Retina Layers Diagram. [Online].

http://anatomysystem.com/retina-layers-diagram/
[46] Young, L. R., & Sheena, D. (1975). Survey of Eye Movement Recording Methods. Behavior Research Methods & Instrumentation, 7(5), 397–439.
[47] Duchowski A T. Eye tracking techniques[M]//Eye Tracking Methodology. Springer, Cham, 2017: 51.
[48] Hess C W, Muri R, Meienberg O. Recording of horizontal saccadic eye movements: methodological comparison between electro-oculography and infrared reflection oculography[J]. Neuro-ophthalmology, 1986, 6(3): 189-197.
[49] Overview of Major EyeMovement Recording Technologies. [Online].
http://aibolita.com/sundries/23641-overview-of-major-eyemovement-recording-technologies.html
[50] Chapter 4 Measurement of Eye Position and Motion. [Online].
http://www.opt.indiana.edu/v665/CD/CD_Version/CH4/CH4.HTM
[51] Duchowski A T. Eye tracking techniques[M]//Eye Tracking Methodology. Springer, Cham, 2017: 55.
[52] Crane H D. The Purkinje image eyetracker, image stabilization, and related forms of stimulus manipulation[J]. Visual science and engineering: Models and applications, 1994: 15-89.
[53] (Fig.3.7 Updated: Refining Eye-Tracking Performance with the Chameleon3 USB 3.1. [Online]. https://www.ptgrey.com/case-study/id/11038
[54] Noris B, Keller J B, Billard A. A wearable gaze tracking system for children in unconstrained environments[J]. Computer Vision and Image Understanding, 2011, 115(4): 476-486.
[55] Babcock J S, Pelz J B. Building a lightweight eyetracking headgear[C]//Proceedings of the 2004 symposium on Eye tracking research & applications. ACM, 2004: 109-114.
[56] Ryan W J, Duchowski A T, Birchfield S T. Limbus/pupil switching for wearable eye tracking under variable lighting conditions[C]//Proceedings of the 2008 symposium on Eye tracking research & applications. ACM, 2008: 61-64.
[57] How do Tobii Eye Trackers work? [Online].

https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/
Eye Tracking Helps Improve Accuracy in Radiology. [Online].
https://www.photonics.com/Articles/Eye_Tracking_Helps_Improve_Accuracy_in_Radiology/a43855)
[58] Bohme M, Meyer A, Martinetz T, et al. Remote eye tracking: State of the art and directions for future development[C]//Proc. of the 2006 Conference on Communication by Gaze Interaction (COGAIN). 2006: 12-17.
[59] Morimoto, Carlos H., and Marcio RM Mimica. "Eye gaze tracking techniques for interactive applications." Computer vision and image understanding 98.1 (2005): 4-24.
[60] Apple acquires SMI eye-tracking company. [Online].
https://techcrunch.com/2017/06/26/apple-acquires-smi-eye-tracking-company/
[61] Oculus acquires eye-tracking startup The Eye Tribe. [Online].
https://techcrunch.com/2016/12/28/the-eye-tribe-oculus/
[62] The Eye Tribe. [Online].
https://en.wikipedia.org/wiki/The_Eye_Tribe#cite_note-9
[63] Dalmaijer, Edwin. Is the low-cost EyeTribe eye tracker any good for research?. No. e585v1. PeerJ PrePrints, 2014.
[64] Ooms K, Dupont L, Lapon L, et al. Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental setups[J]. Journal of eye movement research, 2015, 8(1).
[65] The Eye Tribe Tracker. [Online]. https://theeyetribe.com/products/
CHAP4
[66] Huey E B. The psychology and pedagogy of reading [M]. The Macmillan Company, 1908.
[67] Stark L, Vossius G, Young L R. Predictive control of eye tracking movements [J]. IRE Transactions on Human Factors in Electronics, 1962 (2): 52-57.
[68] Yarbus A L. Eye movements during perception of complex objects [M]// Eye movements and vision. Springer US, 1967: 171-211.
[69] Jacob R J K, Karn K S. Eye tracking in human-computer interaction and usability

research: Ready to deliver the promises [M]// The mind's eye. 2003:573-605.
[70] Eye Tracking Through History. [Online].
http://eyesee-research.com/blog/eye-tracking-history/
[71] De Luca A, Weiss R, Hußmann H, et al. Eyepass-eye-stroke authentication for public terminals[C]//CHI'08 Extended Abstracts on Human Factors in Computing Systems. ACM, 2008: 3003-3008.
[72] Dunphy P, Fitch A, Olivier P. Gaze-contingent passwords at the ATM[C] //4th Conference on Communication by Gaze Interaction (COGAIN). Citeseer, 2008: 59-62.
[73]Ellis K K E. Eye tracking metrics for workload estimation in flight deck operations[J]. Theses and Dissertations, 2009: 288.
[74]Nacke L, Ambinder M, Canossa A, et al. Game metrics and biometrics: the future of player experience research[C]//Conference on Future Play. Algoma University, 2009.
[75] Fookes C, Maeder A, Sridharan S, et al. Gaze based personal identification[M]//Behavioral Biometrics for Human Identification: Intelligent Applications. IGI Global, 2010: 237-263.
[76] Fookes C, Maeder A, Sridharan S, et al. Gaze based personal identification[M]//Behavioral Biometrics for Human Identification: Intelligent Applications. IGI Global, 2010: 237-263.
[77] Weaver J, Mock K, Hoanca B. Gaze-based password authentication through automatic clustering of gaze points[C]//Systems, Man, and Cybernetics(SMS), 2011 IEEE International Conference on. IEEE, 2011: 2749-2754.
[78] Deravi F, Guness S P. Gaze Trajectory as a Biometric Modality [C}// Biosignals. 2011:335-341.
[79] Holland C, Komogortsev O V. Biometric identification via eye movement scanpaths in reading[C]// Biometrics (IJCB), 2011 International Joint Conference on. IEEE, 2011: 1-8.
[80] Porta M, Ricotti S, Perez C J. Emotional e-learning through eye tracking[C]//Global Engineering Education Conference (EDUCON), 2012 IEEE. IEEE, 2012: 1-6.

[81] Komogortsev O V. Karpov A, Holland C D, et al. Multimodal ocular biometrics approach: A feasibility study[C]//Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on. IEEE, 2012: 209-216.

[82] Cuong N V, Dinh V, Ho L S T. Mel-frequency cepstral coefficients for eye movement identification[C]//Tools with Artificial Intelligence(ICTAI), 2012 IEEE 24th International Conference on. IEEE, 2012, 1: 253- 260.

[83] Rigas I, Economou G, Fotopoulos S. Human eye movements as a trait for biometrical identification[C]//Biometrics: Theory, Applications and systems (BTAS), 2012 IEEE Fifth International Conference on. IEEE, 2012: 217-222.

[84] Juhola M, Zhang Y, Rasku J. Biometric verification of a subject through eye movements[J]. Computers in biology and medicine, 2013, 43(1): 42-50.

[85] Darwish A, Pasquier M. Biometric identification using the dynamic features of the eyes [C]//Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on . IEEE, 2013: 1-6.

[86] Cymek D H, Venjakob A C, Ruff S, et al. Entering PIN codes by smooth pursuit eye movements[J]. Journal of Eye Movement Research, 2014, 7(4).

[87] Nugrahaningsih N, Porta M. Pupil size as a biometric trait[C]//International Workshop on Biometric Authentication. Springer, Cham, 2014: 222-233.

[88] Cantoni V, Galdi C, Nappi M, Porta M, Riccio D. GANT: Gaze analysis technique for human identification [J]. Pattern Recognition, 2015, 48(4): 1027- 1038.

[89] Cazzato D, Evangelista A, Leo M, et al. A low-cost and calibration-free gaze estimator for soft biometrics: An explorative study [J]. Pattern Recognition Letters, 2016, 82: 196-206.

[90] George A, Routray A. A score level fusion method for eye movement biometrics [J]. Pattern Recognition Letters, 2016, 82: 207-215.

[91] Galdi C, Nappi M, Riccio D, et al. Eye movement analysis for human authentication: a critical survey [J]. Pattern Recognition Letters, 2016, 84: 272-283.

[92] Vella F, Infantino I, Scardino G. Person identification through entropy oriented mean shift clustering of human gaze patterns [J]. Multimedia Tools and Applications, 2017, 76(2): 2289-2313.

[93] Lyamin A V, Cherepovskaya E N. An Approach to Biometric Identification by Using Low-Frequency Eye Tracker [J]. IEEE Transactions on Information Forensics and Security, 2017, 12(4): 881-891.
[94] Wu Q, Zeng Y, Zhang C, et al. An EEG-based person authentication system with open-set capability combining eye blinking signals [J]. Sensors, 2018, 18(2): 335.
[95] Kasprowski P, Harezlak K. Fusion of eye movement and mouse dynamics for reliable behavioral biometrics [J]. Pattern Analysis and Applications, 2018, 21(1): 91-103.
CHAP5
[96] Jordan M I, Mitchell T M. Machine learning: Trends, perspectives, and prospects[J]. Science, 2015, 349(6245): 255-260.
[97] Samuel A L. Some studies in machine learning using the game of checkers[J]. IBM Journal of research and development, 1959, 3(3): 210-229.
[98] Turing A M. Computing machinery and intelligence[M]//Parsing the Turing Test. Springer, Dordrecht, 2009: 23-65.
[99] Crevier D. AI: the tumultuous history of the search for artificial intelligence[M]. Basic Books, 1993.
[100] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological review, 1958, 65(6): 386.
[101] Solomonoff R J. A formal theory of inductive inference. Part II[J]. Information and control, 1964, 7(2): 224-254.
[102] Cover T, Hart P. Nearest neighbor pattern classification[J]. IEEE transactions on information theory, 1967, 13(1): 21-27.
[103] History of artificial intelligence. [Online].
https://en.wikipedia.org/wiki/History_of_artificial_intelligence
[104] Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition[M]//Competition and cooperation in neural nets. Springer, Berlin, Heidelberg, 1982: 267-285.
[105] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436.
[106] Hopfield J J. Neural networks and physical systems with emergent collective

computational abilities[J]. Proceedings of the national academy of sciences, 1982, 79(8): 2554-2558.
[107] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. nature, 1986, 323(6088): 533.
of the third international conference on. IEEE, 1995, 1: 278-282.
[108] Vapnik V, Levin E, Cun Y L. Measuring the VC-dimension of a learning machine[J]. Neural computation, 1994, 6(5): 851-876.
[109] Ho T K. Random decision forests[C]//Document analysis and recognition, 1995., proceedings
[110] Marr B. A short history of machine learning–every manager should read[J]. Forbes. URL: https://www. forbes. com/sites/bernardmarr/2016/02/19/a-shorthistory-of-machine-learning-every-manager should-read, 2016.
[111] Deep Blue, IBM's supercomputer, defeats chess champion Garry Kasparov in 1997. [Online]. http://www.nydailynews.com/news/world/kasparov-deep-blues-losingchess-champ-rooke-article-1.762264
[112] Hofmann T, Schölkopf B, Smola A J. Kernel methods in machine learning[J]. The annals of statistics, 2008: 1171-1220.
[113] Bennett J, Lanning S. The netflix prize[C]//Proceedings of KDD cup and workshop. 2007, 2007: 35.
[114] Google achieves AI 'breakthrough' by beating Go champion. [Online]. https://www.bbc.com/news/technology-35420579
[115] Müller A C, Guido S. Introduction to machine learning with Python: a guide for data scientists[M]. " O'Reilly Media, Inc.", 2016: 2-4.
[116] Russell S J, Norvig P. Artificial intelligence: a modern approach[M]. Malaysia; Pearson Education Limited,, 2016: 695.
[117] Mohri M, Rostamizadeh A, Talwalkar A. Foundations of machine learning[M]. MIT press, 2012:7.
[118] Russell S J, Norvig P. Artificial intelligence: a modern approach[M]. Malaysia;

Pearson Education Limited,, 2016: 694.
[119] Müller A C, Guido S. Introduction to machine learning with Python: a guide for data scientists[M]. " O'Reilly Media, Inc.", 2016: 35-36.
[120] Leo B, Friedman J H, Olshen R A, et al. Classification and regression trees[J]. Wadsworth International Group, 1984.
[121] Müller A C, Guido S. Introduction to machine learning with Python: a guide for data scientists[M]. " O'Reilly Media, Inc.", 2016: 70-74.
[122] Alpaydin E. Introduction to machine learning[M]. MIT press, 2009: 186.
[123] James G, Witten D, Hastie T, et al. An introduction to statistical learning[M]. New York: springer, 2013: 316-317.
[124] Ho T K. The random subspace method for constructing decision forests[J]. IEEE transactions on pattern analysis and machine intelligence, 1998, 20(8).
[125] Friedman J, Hastie T, Tibshirani R. The elements of statistical learning[M]. New York, NY, USA:: Springer series in statistics, 2001: 587-588.
[126] Mennitt D, Sherrill K, Fristrup K. A geospatial model of ambient sound pressure levels in the contiguous United States[J]. The Journal of the Acoustical Society of America, 2014, 135(5): 2746-2764.
[127] Müller A C, Guido S. Introduction to machine learning with Python: a guide for data scientists[M]. " O'Reilly Media, Inc.", 2016: 93-95.
[128] Naive Bayes classifier. [Online]. https://en.wikipedia.org/wiki/Naive_Bayes_classifier.
[129] Rish I. An empirical study of the naive Bayes classifier[C]//IJCAI 2001 workshop on empirical methods in artificial intelligence. New York: IBM, 2001, 3(22): 41-46.
[130] Müller A C, Guido S. Introduction to machine learning with Python: a guide for data scientists[M]. " O'Reilly Media, Inc.", 2016: 68-69.
[131] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of computer and system sciences, 1997, 55(1): 119-139.
[132] Duda R O, Hart P E, Stork D G. Pattern Classification and Scene Analysis Part 1:

Pattern Classification[J]. Wiley, Chichester, 2000.

[133] M. Sugiyama, Introduction to Statistical Machine Learning, Morgan Kaufmann, First Edition, 2016

[134] Dayhoff J E, DeLeo J M. Artificial neural networks: opening the black box[J]. Cancer: Interdisciplinary International Journal of the American Cancer Society, 2001, 91(S8): 1615-1635.

[135] Hopfield J J. Artificial neural networks[J]. IEEE Circuits and Devices Magazine, 1988, 4(5): 3-10.

[136] Bajpai S, Jain K, Jain N. Artificial neural networks[J]. International Journal of Soft Computing and Engineering (IJSCE), 2011, 1(NCAI2011).

[137] S. R. Dorling, M. W. Gardner, "Artificial neural networks (the multilayer perceptron) — a review of applications in the atmospheric sciences", Atmospheric Environment, vol. 32, no. 14-15, pp. 2627-2636, 1998

[138] OGAMA (OpenGazeAndMouseAnalyzer): An open source software designed to analyze eye and mouse movements in slideshow study designs. [Online].http://www.ogama.net/

[139] Data Mining Fruitful and Fun. [Online].https://orange.biolab.si/

Acknowledgements

The four-year career of my Ph.D. study has passed as if in a blink. Hereby I could scarcely express every mood about all the experiences filled the four years which had been probable imperfect but enriched in a few paragraphs. Everything in memory is as fresh as if they have just happened, and my fervency is how deep it was when the moment I firstly stepped on the land of Italy four years ago. For sure that in the many four-year sections of the life, this has been one of the most remembered and cherished of mine. What the experience of my Ph.D. study during these four years has connected was not merely from the mother country I used to grow up to Italy, the beautiful country which is seven time zones and over 8000 kilometers away, but as well from the sight in front of my eyes to a grander world. The most important principle I learnt from the four-year experience in Italy is that, both personally and nationally, a rich history and tradition is not just the source of pride, but also a treasure that always keeps you remembering who you are, where you are from and how you will conductyourself.

I would like to present my highest respect and gratitude to my supervisors, Prof. Virginio Cantoni and Prof. Marco Porta. My research would never go fluently and successfully without your patient trainings and kind help. Moreover, your hearts of gentle and charm of personality of have impressed me deeply. If one day I have fortunately follow the path of your career, your ways of doing the research, treating the students and constructing the atmosphere of the laboratory will my tutorial of how to be a good professor.

With the same love and expectation as the previous years over one score, my parents have given their support and love to me along these four years. And with the same results as the previous stages of my education, I am to bring back a new obtainment in my life to present you. Thank you once more for the changeless love these years and being the harbor of my heart, wherever as far as I have sailed.

It will be the end of the distance separating my girlfriend, who was doing her Ph.D. study in University of Kyoto and me. No matter the time of four years and the distance from Japan to Italy has never changed our love, passion and faith. During the several days I saw you again every half year, no matter the reflection of the old bridge on River Ticino, the evening wind above River Katsuragawa, or the neon lamps along River Songhua in our hometown, wherever I have you with me, where I feel as “the home”, even if it lies at the edge of the world. I would just like to thank your company in these four years, as well as everything since thirteen years ago when we were teenagers.

To my dearest colleagues in Computer Vision and Multimedia Laboratory of University of Pavia, I would like to thank you for all the joy and support you gave to me. Mrs. Alessandra Setti, with your generous and warm help, everything with my study and living in Italy was so fluent and pleasant; Prof. Mauro Mosconi and Prof. Giuseppe Lisanti, I am so glad to obtain friendship from you beyond the gap of the position of “professor and student”; my dear colleague Dr. Mirto Musci and Gianluca Gerard, and dear colleagues who had graduated and left the lab, Dr. Waqqas Ur Rehman Butt and Dr. Nahumi Nugrahaningsih, I cherish the days we spent in the laboratory and I wish everything goes well with your future career; Prof. Luca Lombardi, Prof. Marco Piastra, Prof. Roberto Marmo, Dr. Piercarlo Dondi, thank you so much for warm greet and smile and you are inseparable parts of my memories of our laboratory. In memories of Mrs. Maria Grazia (“Lella”) Semenza, who had once gave her warm smile and hard work to the laboratory, and had gone on 13 March, 2018.

Best wishes to the all the colleagues and friends I met in Pavia.

Special gratitude to China Scholarship Council, which had been offering me with the scholarship as my financial support for my Ph.D. study in Italy.

致谢

四年的博士生涯转瞬即逝。对于四年来那些未必完美但很充实的体验的每一份心情，此刻我很难在这简短的篇幅间表达详尽。记忆中的一点一滴都新鲜如初，而我的热情如同四年前我刚刚踏上意大利这片土地的那一刻时一样深切。我确信，在我人生中的若干个四年构成的片段中，这一段是我所最难忘和最珍视的之一。这四年的博士经历所联结的，不止是从我生长所在的祖国到意大利这个跨越七个时区位于 8000 多公里以外的美丽国度，更是从我目力所及的视野到一个更广阔的世界。这四年在意大利的经历所教给我最重要的道理是，无论对个人还是对国家，拥有一份深厚的历史和传统，这不仅仅是自豪感的源泉，更是一份财富，它永远提醒着你，身为谁人，何以自持，何去何从。

我谨向我的导师，弗吉尼奥·坎东尼教授和马可·波塔教授，致以我最崇高的敬意和感谢。没有二位教授的耐心教诲和悉心帮助，我将无法顺利成功地完成我的研究。更重要的是，二位教授的温润品行和人格魅力令我深切感怀，如果有一天我能够有幸追随二位教授的职业生涯足迹，你们如何科研治学、对待学生以及营造实验室氛围的方式都将成为指导我如何做一名好教授的教科书。

四年来我的父亲母亲一如既往地给着我支持和爱，正如过去的二十年一直以来他们所抱以的关爱和期望。于是如同过去我每次学成毕业一样的结果，这次我又为你们带回了我人生的新收获。再一次感谢你们这些年来未曾变过的爱，感谢你们一直是我内心的港湾，无论我已经航行到哪里，航行了多远。

这将是我和我远在京都大学攻读博士的女友异地分离的日子的结束。无论是四年的时光还是从日本到意大利的距离都从未改变我们的爱意，热情和信念。在每半年我们重逢的短短几天里，无论是提契诺河上的古桥倒影，桂川上的徐徐晚风，还是家乡松花江畔的霓虹灯火，只要有你在的地方，我就会觉得“这里就是家”，即便它远在世界的尽头。我只想感谢你这四年来的陪伴，以及十三年前当我们青葱年少的时候便已开始的一切。

致我在帕维亚大学计算机视觉与多媒体实验室的最亲爱的同事们，我谨在此感谢所有我从你们相处的日子中获得的支持和快乐。亚历山德拉·塞媞女士，您

慷慨和热情的帮助使我在意大利的学习和生活种种事务都如此顺利和愉快；玛乌罗 • 莫思科尼教授和朱塞佩 • 利桑蒂教授，我很高兴能够超越“教授与学生”的身份之别收获同你们的友谊；我亲爱的同事米尔托 • 穆希博士和札安卢卡 • 杰拉德，以及亲爱的已毕业离开实验室的同事，瓦卡斯 • 乌尔 • 拉赫曼 • 布特博士和娜胡米 • 努格拉宁茜博士，我珍惜和你们共事与实验室的时光并祝愿你们未来生涯一切顺利；卢卡 • 伦巴第教授，马可 • 皮亚斯特拉教授，罗伯托 • 玛莫教授，皮埃尔卡洛 • 东迪博士，感谢你们每次的温暖问候与微笑，你们是我有关实验室的记忆中不可或缺的一部分。缅怀玛莉娅 • 格拉齐娅 •（“蕾拉”）• 塞梦查女士，她曾将她的温暖微笑和辛勤工作献给实验室，与 2018 年 5 月 13 日与世长辞。

向我在帕维亚认识的所有同学与朋友致以最美好的祝愿。

特别感谢中国留学基金委，对我在意大利的博士学习生涯提供了奖学金作为资金支持。