A practical Face Recognition System using a Game with a Purpose

Willer Caxias Travassos

Master of Science

School of Computer Science

McGill University

Montréal, Québec

2010-02-12

A dissertation submitted in partial fulfillment of
the requirements for the degree of Master of Science

DEDICATION

To my parents and my sister for their support,

care, encouragement, and patience

during this academic endeavor.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Xue Liu, for believing in my potential and guiding me through this past year of academic research. Through the span of his knowledge in the field of computer science, he introduced me to new and exciting research areas that are not only interesting but cater to my taste in research. Also, I would like to thank all of those whom I have met and helped me, in one way or another, shape myself as a student as well as a person. Finally, thank you to my family for all of the support you have given to me, during this journey of almost a decade in Canada.

ABSTRACT

A facial recognition system is a computer application built to automatically identify or verify the identity of a person from a digital source. The quality of the source and environment from which digital information is retrieved pose problems to a face recognition system (FRS) that lead to erroneous results. Even though there is a necessity for systems that are capable of performing facial recognition on the fly, current systems, in order to be accurate, try to completely or partially control the environment of its digital sources. In this thesis, we approach the problems that afflict these systems by using a “game with a purpose” (GWAP). In our GWAP-based approach, we create an online game that uses human vision to perform facial recognition. We use the aid of humans because we still hold the edge over FRSs at recognizing faces and their features. It is through a GWAP that we channel our superior visual skills to deal with problems that affect current FRSs.

ABRÉGÉ

Un système de reconnaissance faciale est une application informatique créée pour identifier ou vérifier automatiquement l'identité d'une personne provenant d'une source numérique. La qualité de la source et l'environnement duquel l'information numérique est extrait posent des problèmes à un système de reconnaissance faciale (FRS) et peuvent causer la production de résultats erronés. Même s'il y a une nécessité pour les systèmes qui sont capables d'effectuer la reconnaissance des visages à la volée, les systèmes actuels, pour être précis, essayent de contrôler complètement ou partiellement l'environnement de ses sources numériques. Dans cette thèse, nous abordons les problèmes qui affligent ces systèmes en utilisant un "game with a purpose" (GWAP). Dans notre GWAP, nous créons un jeu en ligne qui utilise la vision humaine, pour effectuer la reconnaissance du visage. Nous utilisons l'aide des êtres humains, car nous tenons toujours l'avantage sur les systèmes de reconnaissance faciale à la reconnaissance des visages et ses caractéristiques. C'est par une GWAP que nous canalisons nos compétences visuelles supérieures pour faire face aux problèmes qui affectent actuellement les FRSs.

TABLE OF CONTENTS

LIST OF FIGURES

1 Introduction

1.1 Game with a Purpose

A game with a purpose (GWAP) is a game developed to channel human brain power, i.e., to capture the human capacity to solve, analyze, assimilate, and interpret stimulus from the outside world. These types of stimulus can be things such as music, movies, art, or challenges of a daily routine. Even though computers and software have evolved greatly in recent decades, they are still not capable of performing better or at least, imitating the abilities that are innate to us such as intelligence, sight, and hearing, [1, 2, 3].

The idea behind a GWAP is to utilize our human computational power to solve problems that are still complex to a machine and trivial to us. Games were chosen as the manner of harnessing such power because they provide people with a tool to perform tasks that would be otherwise daunting and tedious. Also, engaging games and a well-known method of distribution can allow the creation of a distributed system with massive computational power (i.e., several people), which can solve open problems and provide insight into our abilities and how we use them.

The use of GWAPs was first discussed and presented by Luis Von Ahn, who used GWAPs to increase Internet accessibility to users with special needs, in particular blind Internet users. However, the potential of GWAPs is vastly unexplored, and there are still problems at which computers do not excel or have yet to be tackled. As such, we can use the power of distributed human computing to undertake the task of solving complex problems for today's computers while teaching them to understand how we perceive and judge things that surrounds us on a daily basis.

The goal of this research and GWAP is to use the amassed computational power of people all around the world to solve problems that plague facial recognition systems (FRS). Today's FRSs suffer from a lack of accuracy due to the viewing angles of a face and dependence on environmental conditions such as illumination and the absence of objects in a picture.

The most important aspect of these weaknesses is that these systems are less effective when they handle faces in a picture that contains facial expressions. But human beings, on the other hand, are not easily influenced by these details and are considered better at facial recognition than most FRSs. Thus, by siphoning the skills of humans to detect faces and their details through a game, we can overcome the shortcoming of FRSs and, at the same time help FRSs improve and even start replicating our capacity for facial recognition.

1.2 Organization of the Thesis

This thesis is organized as follows. Chapter 2 introduces background and related work that has inspired this thesis. Chapter 3 includes a discussion on the current methods used in facial recognition. Chapter 4 talks about employing a GWAP to better facial recognition and the byproduct information that originates from designing and using a GWAP to tackle facial recognition. Chapter 5 describes the internal details of the GWAP, the technical information related to it, a walkthrough of the game, and its execution. Chapter 6 discusses the evaluation period of the GWAP. Finally, Chapter 7 presents the conclusion and potential directions for future work.

2 Related Work

2.1 Origins of Games with a Purpose

Using GWAPs as a helper for solving complicated computer tasks is a new research area. A GWAP serves as an interface that collects data that can be fed to a machine learning algorithm. It aims to aid in the development of truly intelligent systems that are capable of mimicking innate abilities of human beings such as logical thinking, hearing, and vision, without the difficulties that these systems face today. It does this by using human-based computation, i.e., by outsourcing certain computation steps, in an algorithm, to humans.

The main idea behind a GWAP is similar to the Internet-based Open Mind Initiative (OMI) [1], which intends to develop intelligent software by collecting information from Internet users. The OMI aims to create software that is capable of comprehending general knowledge and concepts that underlie our general intelligence [6, 12]. To reach these goals the OMI relies on three types of people that contribute to the initiative as follows:

1. Field experts: Those who provide the problems to be solved and the technical information needed to solve them such as learning and fundamental algorithms.
2. Programmers: Those who provide software to retrieve and manipulate data from Internet users.

3. Internet users: Those who provide the data.

Figure 1: GWAP can help teach computers how we perceive the world around us

In an OMI-based software system, a user participates by answering yes or no questions, voting on the logical correctness of sentences, or playing 20 Questions-based games. However, to constantly provide data that regards our common knowledge or perform menial tasks in a straightforward fashion can be quite boring to us human beings. This can be very taxing, especially when programmers do not provide an attractive and entertaining interface to perform these tasks. Figure 2 shows us MIT's Open Mind Common Sense interface, which is an example of OMI-style software.

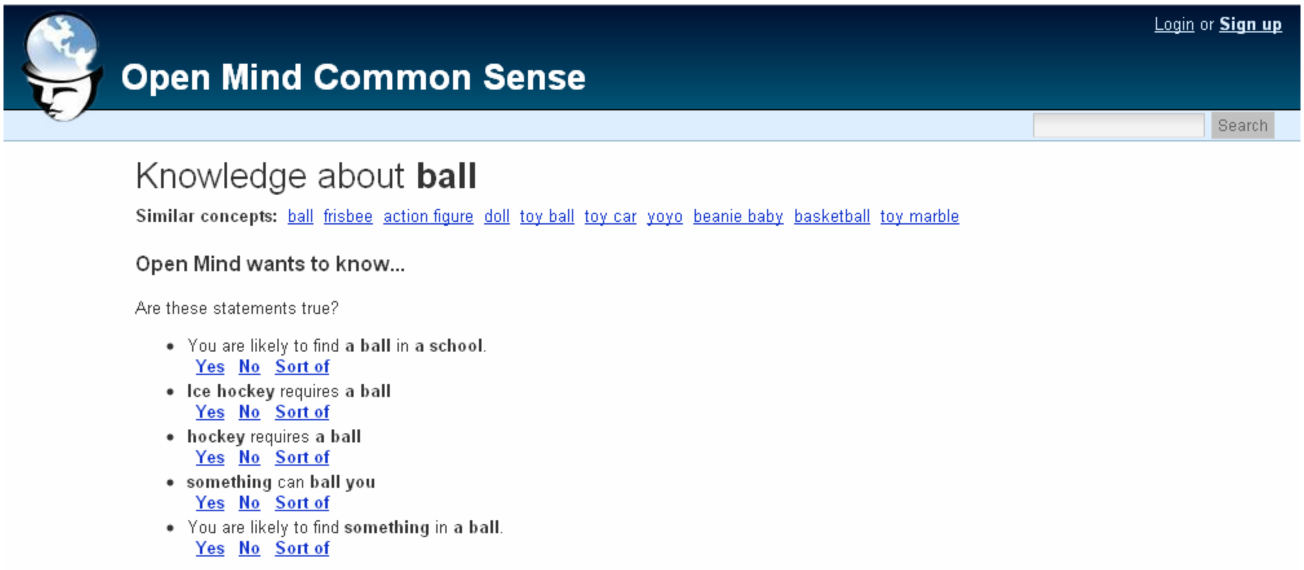

Figure 2: Users agreeing on a common-sense fact using the OMI approach.

This is where GWAP approach differs from the OMI approach. The OMI approach concentrates solely on retrieving data, while the GWAP approach utilizes games to attract and keep users around as much as possible. The GWAP approach concentrates on making the games easy to learn and play, fun, online, easily accessible, and, more importantly, able to produce data to solve a computational need. The idea here is that we use games so that more users can play and enjoy our games so that they will return often. Thus, with greater a flow of players in our games, we can gather great amounts of data and aim to solve bigger problems that involve today's computers systems and software.

2.2 Current Applications

As mentioned before, the first person to coin the term “human computation” and do research with GWAP was Carnegie Mellon University professor Luis Von Ahn. The purpose of his initial research was to use games to achieve the ambitious goal of precisely describing the majority of the images on the Internet. According to Ahn, the description of images, i.e., labels, on the Internet does not contain a method of standardized description, which can lead to poor textual depiction of a picture or none at all, [1].

This deeply affects applications that depend on such image labels such as image search engines or a Web browser for the visually impaired. Plus, manually labeling images is an arduous and boring task. Looking for a manner of making such tasks less laborious and much more entertaining, Prof. Von Ahn decided to use games as interfaces that would help to label images and let users enjoy the experience. With this game, Prof. Von Ahn would let users perform image-labeling without letting them know that they were working to solve a problem.

The ESP game which pairs two Internet users and asks them to describe images with one word is the result of his research. These users are randomly paired and unable to communicate with each other so that cheating can be avoided. When the game starts, players can see only the image that they have to describe and the input for the image description. The goal of the game is to have each player guess and type the word that the other player types. As soon as the game finds a word that both partners have typed, i.e., that the players have agreed on the image, the word is considered to be a possible label for the image, and the game moves on to the next image to be described. Figure 3 shows a snapshot in the game action of an ESP match.

To avoid the stagnation of image descriptions, words that have been agreed upon and accepted as a label for a certain image are considered taboo words. This means that the next time an image is used in an ESP, match players will not be able to use the taboo words or variations such as plurals, to describe an image. The idea behind using taboo words is that first labels have the tendency to be general. Thus, taboo words would enforce a greater amount and variety of labels, making an image description more specific.

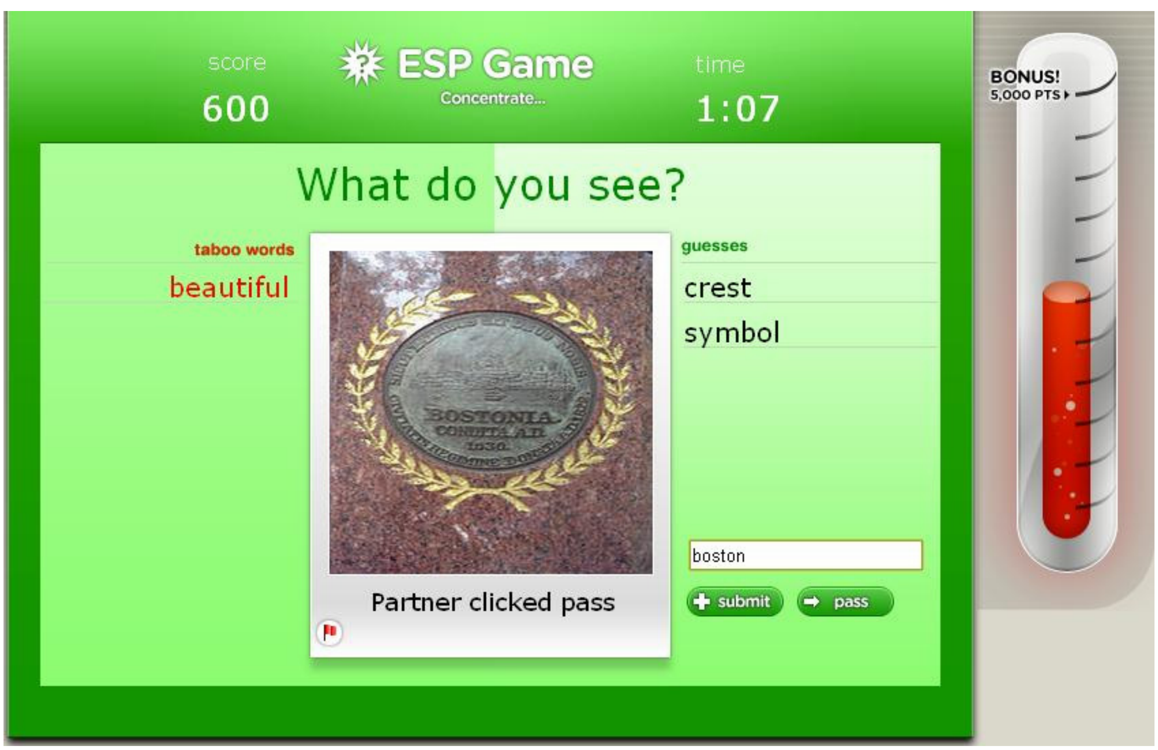

Figure 3: GWAPs encourage users to play by providing a colorful UI and competition.

Once an image acquires an extensive list of taboo words, it is considered to be fully labeled and is removed from the game. A list of fully labeled images can then be used by several applications that may improve Internet accessibility or be used as data for machine-learning algorithms in the computer vision field or even another GWAP. One important thing to note is that fully labeled images are re-inserted into the ESP game after some time, since images' descriptions might change over time because cultural values or the English language and the usage of its words may change over time. For example, clothes that were considered "fashionable" or "cool" may not be considered as such nowadays, but in the future they might be considered "cool" again.

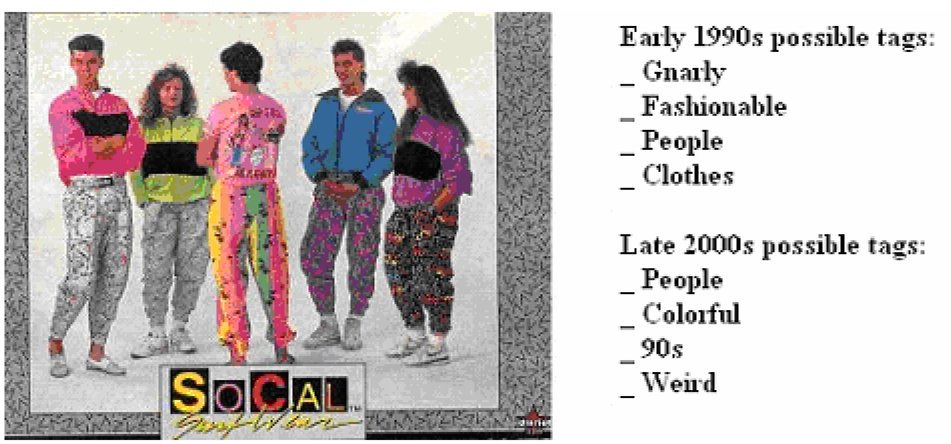

Figure 4: Possible changes in descriptions (through a decade).

Other games developed by Prof. Von Ahn follow the same idea of the ESP game, i.e., they let users describe things that computers are not quite capable of describing. The Tag a Tune and PopVideo games, like the ESP game, let users label song excerpts, and videos, respectively, but they differ in their execution. It is not possible to describe how these games work internally, since papers related to them have not been published.

However, when playing the games, you can have an idea of how they work. The difference between PopVideo and the ESP game is that instead of having only two players, a group of randomly chosen users labels a video by describing anything that they see and hear in the video. If at least two users agree on a label, they gain points and the label may become a part of the video description.

In the Tag a Tune game, again two users are randomly paired and unable to communicate, but instead of having the same song excerpts to describe, the users may be asked to describe different excerpts. As players enter possible labels for a song, they can see what their opponent is entering as a possible label. According to the list inputs created by both players, they can decide whether the song they are listening to is the same or not. Since players can get points only if they join forces to determine if their song is the same or not, the labels typed by both users tend to be very accurate and good enough to label a song.

Squigl is another game that uses visual cognition, but instead of asking users to describe images, it asks users to recognize objects in a picture. It is a game that aims to capture how we humans perceive objects in an environment, and its results can be used to improve algorithms that deal with image segmentation. Two randomly paired users are asked to trace an object or feature, determined by a label and displayed together with a related picture. The closer the outlines of both users are, the more points they gain, which encourages their traced outlines to be similar to each other. It is important to mention that the labels displayed in the game are the labels retrieved from the ESP game.

Figure 5: Agreeing on common-sense facts using the GWAP approach.

Not all games developed by Prof. Von Ahn strictly follow the same formula as the three presented above, and some even stray away from the problems that visual cognition can solve. Verbosity, for example, attempts to collect common-sense reasoning related to a word. Here, the game pairs two random players assuming the roles of a narrator and a guesser with the goal of collaborating with each other to gain as many points as possible. The role of a narrator is to describe a secret word to the guesser while using sentence templates with blanks that can be filled with a hint word related to the secret word, as shown in Figure 5. The role of the guesser is to guess the word according the received hints, i.e., received common-sense facts, as shown in Figure 6. Note how the GWAP approach greatly differs from the OMI approach.

Each round of the game has a secret word that must be discovered by the guesser or agreed to be skipped between by both the Narrator and Guesser. Throughout a round, the narrator can help the guesser by indicating whether the guesses made are close ("hot"), or far away ("cold") from the secret word. Once a round is finished, either with a correct guess or skipped word, the players switch roles. Skipping words does not provide points, which encourages players to enter hints that are related to a secret word, which, in turn, can be used as good, common-sense facts about a word.

Figure 6: The guesser reads common sense facts and guesses the related secret word.

Even though GWAP is a very new research area, there are already projects outside of Prof. Von Ahn's realm that draw from this idea. Foldit is a GWAP about protein folding developed at the University of Washington. It aims to improve current folding algorithms by using the human capacity for 3D environment perception and intuition to solve 3D puzzles [15]. Another GWAP outside of Prof. Von Ahn's realm is Phrase Detectives, developed at Essex University. The game intends to use the human capacity for understanding and distinguishing anaphora links in phrases to train anaphora resolution systems (ARS) to improve Web surfing by improving text summarization and search engine indexing methods [14]. An anaphora is, generally speaking, an expression that refers to another text. For example, the phrase "Jerry looked at the ball and kicked it" is anaphoric, since "it", an anaphoric link, refers to the word "ball".

2.3 Our Contributions

As mentioned before, research in the area of GWAP is rather new and, thus, full of opportunities that can help the development of new and better computers. More importantly, using a GWAP provides us with an exciting and powerful new way to tackle unsolved problems that affect several fields in computer science. Artificial intelligence, Internet searching and accessibility are just few of the domains that are benefiting from the use of GWAPs. However, due to their versatility, other domains in computer science could use a GWAP to address difficult problems. For example, security and computer vision are fields that can benefit from using or drawing ideas from the GWAP approach.

ReCAPTCHA is a CAPTCHA system that is an example of a solution to a security problem that uses base concepts of GWAP. As with any CAPTCHA service, ReCAPTCHA uses visual tests to determine whether access of to a Web page is done by a human being or not. The difference between ReCAPTCHA and any other CAPTCHA system is that it uses human-based computation to help in the digitization of books, newspapers, and old radio shows [16].

Employing a GWAP to solve a problem depends on the creativity of a person to come up with games that fit a certain need within a domain. Using my academic experience with computer vision and computer perception, plus my research, I can ascertain that computers have difficulty in emulating human senses. Computer vision, in particular, seems to have the greatest promise to fully copy our ability to see, but it currently suffers from several problems that hinder the realization of its full potential. One area that fits this scenario is facial recognition and facial recognition systems that work on human facial recognition

We look to help improve current facial recognition methods and FRSs by using a GWAP that deals with weaknesses that affect them. Although there have been advances in this area, facial recognition is still not precise. Also, facial recognition struggles to perform under normal conditions, e.g., to perform in a reasonable manner; some FRSs need neutral facial expressions or lights fixed at a certain spot. All of these restricting factors are crucial for the effectiveness of facial recognition; they need to be reinforced, or a FRS might fail. However, pictures used in facial recognition probably will never meet these restrictions posed by FRSs, especially for systems that attempt to perform matches on the fly, i.e., with any facial representation available, be it old photographs or video from a camera on the street.

In this thesis, we first look into the usage of FRSs, and more deeply into the weaknesses that affect them. We briefly introduce some of the techniques used by FRSs, and then we describe the problems that plague these systems. Also, we explain the scenarios in which these systems are generally used and how these scenarios help deter the functionality of FRSs. In this section, we want to illustrate how these problems affecting FRSs are hard for a machine to decipher but, at the same time, are easily solved by a human eye. We look into how a GWAP would fit the role of a FRS and how we can use a GWAP to explore the superiority of our visual capabilities.

Once we discuss the possibilities of using a GWAP as a FRS, we move on to the design of our GWAP. We show how our GWAP allows users to capture and filter data present in facial pictures that can cause confusion for FRSs. We discuss the logic of our game and the services it provides, as well as how it uses our vision and our capabilities for describing, comprehending, and recognizing visual information in order to help us match facial pictures. We also talk about the game flow and how to extract our facial data.

Finally, we look closely at the architecture of the game itself and how the background of our system works. We discuss techniques and algorithms used in the development of our GWAP. Also, we discuss the evaluation period of our system and the data we have collected from its use.

3 Facial Recognition

3.1 Facial Recognition Systems

Facial recognition systems (FRSs) are computer applications built to automatically identify or verify a person through a picture of his or her face. FRSs are non-intrusive biometric systems that use features of the human face to analyze and compare them against the facial features of other faces found in a database of face images. These systems apply several algorithms in order to identify faces in an image (face detection), retrieve biometric features, and compare them against others. Even though they may utilize different types of algorithms, they can be generally divided into two types: image-based, and feature-based.

Figure 7: Sample probe images and its resulting image space, i.e., eigenfaces.

An example of an image-based algorithm is the Principal Components Analysis (PCA). In a PCA system, several normalized images are gathered in order to train the system and calculate eigenfaces that are compared and matched against a normalized probe image. PCA treats each probe image as a vector in a very high-dimensional space, in which every pixel corresponds to a coordinate. This high-dimensional image space is actually represented by the eigenfaces, which are nothing more than a set of eigenvectors created from the images that trained the system. A PCA-based FRS compares a probe image against the image space created by the eigenfaces and finds a match by retrieving its nearest neighbor in the image space, [37, 42].

An example of a feature based algorithm is the Elastic Bunch Graph Matching (EBGM) a biology-based method that tries to simulate the processes present in the visual cortex of higher mammals [50]. To perform comparisons, the system contains models of faces that have a set number of reference points, such as the nose tip, or pupils. Each reference point in a model is a vertex belonging to a graph representing a standard face. Due to the variation that is found in faces throughout the globe, it is impossible to have models representing each variation.

To cover this wide range of variations, EBGM uses a finite set of graphs and bunches them together, according to a reference point, in order to create a greater range of face models. After entering some manually created face graphs representing different poses, the system generates graphs for each face that is entered in its database. Thus, facial recognition can be executed with little computational effort by simply entering a probe image, generating its face graph, and comparing the reference points in the probe's face graph with the existent ones in the database, [38, 42].

3.2 Uses for Facial Recognition

Due to its passive nature, FRSs, in conjunction with cameras, can be used for general surveillance and security. Primarily, FRSs have been used by law enforcement agencies to search for criminals in a crowd. There has been documented use of FRSs for such ends during the early part of the 2000s in American airports, football stadiums, and cities located in the United States and the United Kingdom.

An example of such system is the United States Visitor and Immigrant Status Indicator Technology (US-VISIT). When foreign travelers pass through a port of entry to the U.S.A. or apply for a visa to enter the country, pictures and fingerprints are taken by officials. The retrieved biometric data is checked with databases in order to determine whether a traveler may receive a visa or enter the country, [48, 49].

However, utilization of FRSs does not need to be restricted to these types of situations. Other scenarios that may profit from facial recognition are identity confirmation, searches for missing people, and prevention of identity theft. For example, in banks a quick face scan could be used to verify a costumer’s identity instead of using a bank card and a PIN number. Recently, Toshiba has released computers that give the option of using facial recognition to let users log into their computers.

Even though FRSs have been constantly improving and employed throughout the world, facial recognition is far from perfect. FRSs are still susceptible to external factors such as illumination, shadows, busy subject surroundings, and facial expressions. More importantly, while some FRSs work with the permission of their users, others, e.g. surveillance FRSs, can be used to invade the people's privacy by performing facial recognition without their knowledge.

3.3 Weaknesses

It is clear that the potential of facial recognition is immense, but there are weaknesses that affect its fulfillment. Weaknesses in facial recognition are nothing more than the challenges that a FRS encounters when it is in use. These challenges come from different sources, but can mostly be grouped into logical and legal issues. Logical issues are those challenges that affect an FRS during its execution. Legal challenges, on the other hand, are outside the realm of computer programs and deal with problems that arise from the use of FRSs.

3.3.1 Logical Issues

The toughest aspect of creating a FRS is that these systems have to be invariant to environment influences and the position and pose of a person in a picture. The source of images and the circumstances in which FRSs are used can also show how much these weaknesses affect facial recognition. To better discuss these difficulties, we use scenarios that better exemplify possible FRS utilization.

The first scenario presents the weaknesses that affect a FRS when facial images used by the system, probe and database images, are taken from a controlled environment. Images from controlled environments typically can be retrieved from situations similar to mug shots or shots for government-issued identification. Figure 8 shows us examples of such pictures. The FRSs that work in these scenarios are usually used for identity verification. They can be consistently accurate when high-resolution images, correct illumination, and proper face angles are used and maintained throughout the various pictures.

However, these FRSs are not free of execution challenges. Image-based FRSs here have trouble with matching pictures affected by aging, gender, ethnicity, facial expressions, and partial occlusion of the face with, for example, makeup, facial hair, or accessories. The background of pictures can also cause mismatches in image-based FRSs, since details in it will undoubtedly vary from picture to picture. Figure 9 shows examples of pictures that cause problems for image-based FRSs. A problem that feature-based FRSs encounter here is that locating reference points in the probe image usually requires manual checks. This is necessary because some facial features, e.g., thick eyebrows, can still confound systems and cause mismatches.

Figure 8: Sample pictures taken from a controlled environment.

The second scenario introduces weaknesses that affect a FRS, when the probe image is taken from a controlled environment and the database images are taken from uncontrolled sources, like video cameras and personal pictures, and vice versa. The major weakness that affects FRSs, in this scenario, deal with the fact that pictures retrieved from uncontrolled sources are usually of poor quality. Everything that is used to maintain picture quality in our first scenario such as facial expression and camera distance, are eliminated here.

Every detail from an uncontrolled environment, as in Figure 10, can affect FRS matching. A closed-circuit camera in a busy street is the prime example of an image that presents a problem for a FRS. Clothes, the position of the subject's head, subject pose, cluttered backgrounds, and foreground objects hiding a subject are examples of factors that hamper the performance of a FRS. Even systems that create a profile of a person containing pictures taken from different angles suffer here, since images retrieved from cameras tend to be aligned differently from that of the profile.

Figure 9: Variations of Figure 8 that cause matching problems in today's FRSs. The contrast, shadow, and background detail can all fool a facial recognition algorithm, but not a human being.

Finally, the last scenario deals with the case of images that are always retrieved from an uncontrolled source. In this scenario, besides suffering from the same problems as in the previous one, FRSs have to rely on different image sources. This means that images could come from several pictures and videos that depict a person's face from several angles and under different conditions. This could provide a high amount of facial data that could prove too complex for today's FRS to handle. Also, it is important to note that cautious people can evade or elude FRSs by carefully hiding from the cameras using masks, hats, facial hair, or even plastic surgery, [41, 42, 43, 47, 50, 51].

Figure 10: Facial images can be retrieved from closed circuit cameras, but they are hard to utilize in a FRS. Note how the visible faces are at different angles and partially occluded by other people, shadow, or objects such as a cap.

3.3.2 Legal Issues

The other difficulties that FRSs face are legal issues related to an individual's privacy and freedom. As the technology improves and its use becomes more widespread, there is a chance that people surveillance may become more aggressive and invasive. People could be subject to constant scrutiny from the moment they leave the privacy of their homes until they come back. In addition, there are no guarantees that using FRSs will help improve security, since there are still a lot of ways to evade these systems. Also note that the majority of the people evading the law will not willingly collaborate with law enforcement and have their information entered into these systems, [47, 50].

Due to its passive nature, there could be deviations from intended purpose, or abuse of its power. Law enforcement could attempt to reach even further than its natural reach, or even discriminate against individuals according to the personal biases of the people using the system. As the technology becomes cheaper and more accessible, there is a possibility that private companies or people could use facial recognition for spying on, stalking, or even blackmailing a person. These problems can be dealt with as long as laws that protect civilian rights and regulate the use of FRSs develop together and at the same pace at which FRSs improve, [51].

3.4 Using a GWAP for Facial Recognition

New methods for solving FRSs problems are being developed and researched throughout the world. Private companies and universities tend to look for algorithms that make FRSs more precise and, thus, more reliable. For example, continuous facial recognition research arrived at 3D image and skin texture recognition. These types of FRSs cover problems related to camera angles and illumination, and they allow additional features to be taken into account when looking for matches. However, these techniques also bring problems such as in searching for matches with a 3D image in a 2D-based database consisting of uncontrolled environment pictures, or in retrieving skin information from pictures that do not contain such data.

We try a different approach and look to take a different road to improve FRSs. Instead of trying to force our way the through the well-known path of image-based and feature-based systems, we try a different method. We decided to look at how a GWAP, used as a FRS, would fill the holes that are existent in today's FRSs. Since GWAPs use human brain power, we want to model our FRS according to human vision.

The main reason, to approach facial recognition with a GWAP, is that facial recognition is a type of problem that fits the GWAP principle that humans can deal with it trivially, while machines cannot. As long as we have clues, visual or conceptual, of someone's face, we can recognize this person. This is how police sketches of criminals and fugitives have been used for several years. For example, when somebody is robbed he or she gives a police officer a description of the suspect's appearance. The police then search for people who may match the description, and then the robbed person determines who the robber is.

This is clearly a method of facial recognition performed by human beings on a daily basis. Thus, we would like to adapt our capacity for giving conceptual clues of somebody's face and matching it to real people to our method of dealing with facial recognition. Also, we know that our vision is capable of dealing with illumination, camera angles, and busy backgrounds, among other details in a picture. Therefore, due to these advantages over FRSs, using human vision for facial recognition seems like a good choice.

The costs of using a GWAP for facial recognition are minimal if compared to an FRS, since GWAPs, by nature, use human-based computation to solve problems. We only need to create a clever game that allows us to retrieve the information we need. The benefits of using a GWAP are enormous; for example, the data used by a GWAP can be used to teach machines how we see the world and recognize objects. The only limitation to this method is the limitation of the human mind and eyes, which currently is better at recognizing faces than any current FRS.

An online GWAP offers the possibility of creating a community that is dedicated to its use and improvement. More importantly, we want users to willingly provide us with data. We want to create an interactive system that allows users to gather, analyze, and share knowledge used in describing and comparing faces; i.e., we want to create networks of knowledge based on human participation. We want to use the superiority of our vision to solve the many difficulties that FRSs have trouble solving, even though such an approach inherits the faults that are innate to human vision. Thus, using a GWAP can help to eliminate problems from FRSs that are simple to us, which could help to enhance these FRSs and allow them to tackle problems with which humans have difficulties.

4 The GuessMe GWAP

4.1 Our Idea

The idea of using a GWAP to deal with these facial recognition problems came from the fact that machines fail at some things that we can easily understand. This is the very definition of GWAP; i.e., we use human-based computation to solve problems that computers cannot. Being convinced that a GWAP can improve facial recognition methods compared to today's approaches, we decided to create a game that can aid in achieving this goal.

To address the weakness present in today's facial recognition, we need to develop a game that allows users to describe not only a person's facial features but also more subjective features, such as a smile, a frown, or tied hair. Also, this game would have to allow any facial picture to be used, be it from a controlled or uncontrolled source. We relax these source constraints since we want to describe objects such as sunglasses, braces, and various expressions and camera angles that might be present in an image.

There is already a board game that allows us to perform facial recognition and to capture such subjective information. Created in 1979 by Ora and Theo Coster, "Guess Who?" is a two-player game in which players ask each other yes-or-no questions in order to guess his or her adversary's mystery person. The mystery person is nothing more than a person selected from a pile of cards, each containing a name and a face picture in it. Each player has a board of face pictures used to eliminate people that fit the profile created from the yes-or-no questions they ask. The goal of the game is to correctly guess the mystery person of the other player first, [21, 22].

Figure 11: The “Guess Who?” board game.

Clearly, “Guess Who?” fits perfectly the role of our GWAP; i.e., it can use humans to perform reliable facial recognition computations while processing this data in any way that fits our needs. In this manner, our GWAP version of the “Guess Who?” game can create a deep database that represents the human capacity for understanding facial features and differentiating one person from another, helping us to determine how we make such decisions. Although using a GWAP to deal with facial recognition might not be the fastest method to deal with such problems, it is still a good approach to estimate our human facial recognition capacities and to use them to improve FRSs.

4.2 Game Features

As we said in section 3.4, a clever GWAP has to be created in order to retrieve information that we need. In our case, we are looking for descriptions of facial features that can be used to perform facial recognition among pictures of several people from different sources. Also, we want to retrieve these descriptions from people playing while they play our GWAP, i.e., perform human-based computation for us. In this thesis, we create a GWAP system named GuessMe, based on “Guess Who?”.

Aside from the GuessMe game itself, the GWAP system contains other features allowing users to add content to the game. The game itself contains several differences from “Guess Who?”. The game developed for this thesis is not the completely envisioned game; it is a functional version of it, which serves as a proof of concept. Other parts of the game’s design are later discussed in the section of this paper regarding future work.

To make GuessMe reach its goal, we need to deal with the limitations of the “Guess Who?” design. The most important detail is the fact that “Guess Who?” has a very limited library of face pictures; it only contains only twenty-four pictures distributed among forty-eight cards. Since the goal of GuessMe is to perform human-based facial recognition, we must avoid limiting the number of available facial pictures. Also at the same time, we want to deal with the weaknesses of today’s methods, mentioned in Section 3, and, thus, we want to include a great variety of face angles, expressions, and objects in the pictures.

Because we cannot provide a great number of pictures with this variety of details, we opted for the simplest solution: to let the users add the pictures used by our game. Users can add pictures of any person they want, as long they provide a name for this person. This is necessary because we need to have a common link between these pictures in order to select them for a match, and there is no way of reliably performing facial recognition as soon as the pictures are added.

These pictures, once added by any player, are stored in our database but are not automatically used in a GuessMe match. We want to screen these pictures first because we want to avoid those that are not appropriate for facial recognition. For example, we want to avoid pictures that may have such a low resolution that a face is unrecognizable or that contain objects, such as masks, that might render a face undistinguishable. To avoid such pictures, we introduce a voting system that let users compare existing pictures of a person with a new picture and decide whether a picture is acceptable or not. For now, once five voters decide that a picture is acceptable, it becomes usable by GuessMe.

Users who have signed up with the GWAP system can access, through their profile, the pictures they have added and pictures of them added and tagged by friends. Here, a user has the power to propose that a picture be vetoed from the game if it is tagged with the user's name. We use this feature as an extra tool that a user may use to screen bad pictures belonging to himself or herself. Every time a user vetoes a picture, it is put again to a vote so that users can decide to put it back in the game or not.

4.3 Game Flow

When the GuessMe GWAP is started, users have five options: play the game, add pictures, check their profile, vote on pictures, and logout. As previously explained, users can check their profile so they can see pictures added by themselves and pictures of themselves added by other users, and veto they can any picture. The vote option allows users to decide which pictures can be inserted in the game. As of now, there is not a limit on the number of votes a user can submit. We would like to change this in the future so that we can avoid cheating or the influence of users who might disturb voting sessions.

Adding pictures into the system differs from the norm of loading a picture and adding it immediately to the system. We take a different approach to adding pictures because we want to avoid the creation of multiple profiles that point to the same person. When users want to add a new picture into our system, he or she must first search and determine whether a profile exists for of the person who appears in the picture. Search results, as shown Figure 12, display pictures belonging to people's profiles for easy recognition of a search target. If users cannot find a suitable profile to add pictures for, they can create a new profile that will not be assigned to any user.

We want to provide this option since users can tag pictures of people who are not members, like tagging pictures in any social networking site. In our GWAP, we have an extra reason to allow the creation of such profiles since we want to keep these pictures within a unique profile. Avoiding multiple profiles of pictures of the same person allows for the centralization of content and better facial descriptions by preventing valuable descriptions from being distributed in several places. This method of adding pictures serves as a pre-GWAP facial recognition step in which users link different pictures of the same person for us.

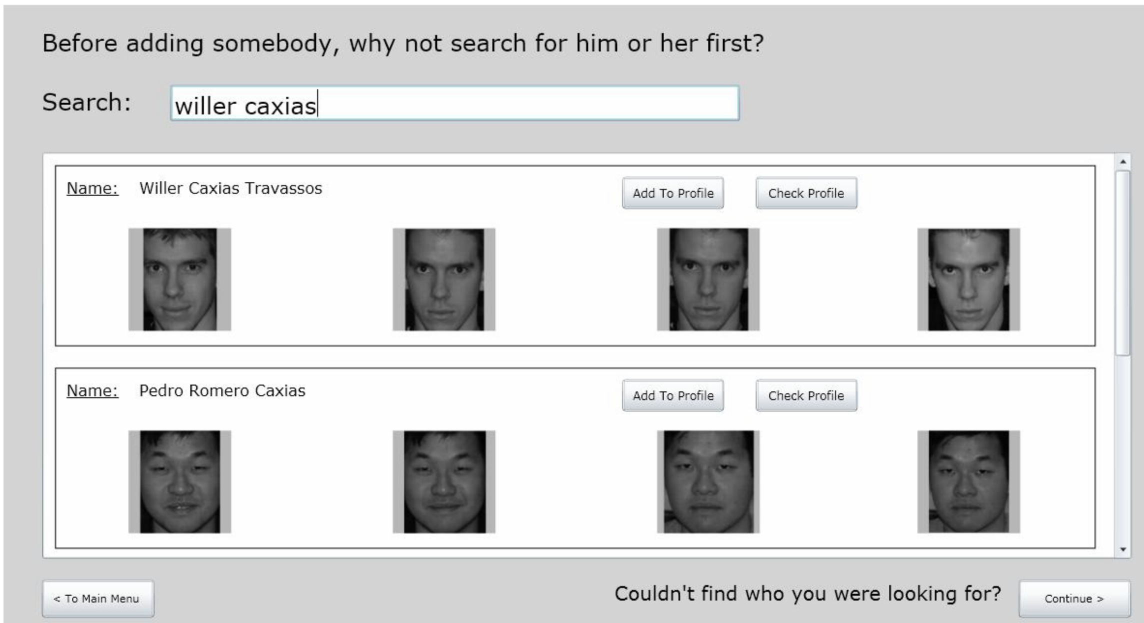

Figure 12: A search result and possible profiles to which a user can add a picture.

The game, if a user decides to play it, is where we can gather descriptions of face pictures. It is through the yes-or-no questions asked by the users playing a match that we generate descriptive tags that refer to the mystery person's facial features. The game play is simple and easy to learn. Once a match starts, the system randomly decides who will start the game. After a decision is made, the system looks for twenty cards containing facial pictures for player one. One of these cards is then chosen as the mystery person for player one. Following this step, the system chooses twenty cards containing different facial pictures of the same people chosen on step for player one and are given to player two. Finally, the system selects the mystery person for player two. Note that profiles containing only one picture are never selected to be used in a GuessMe match.

Once the pictures are loaded, the match is ready to start. Each player has a five by four grid that containing the system-selected pictures as the possible opponent's mystery person, plus a separate picture of the mystery person on the screen. The grid of pictures serves as a helping device for narrowing down the identity of the opponent's mystery person. A player turns down a face picture according to the answers that are received throughout the match. Players can turn their pictures as many times as they want, but it is advantageous to turn down pictures only when they do not fit the answers received from their opponent. Each card has a menu of actions, which players can use to turn a card, guess the mystery person, and see the name of the person in the picture.

Player one asks the first question of the match. These questions are used to help determine player two's mystery person. As in "Guess Who?" these questions can only be answered with a yes or no. For example, they can be formulated as "Does your mystery person have a mustache?". After player two answers the first question of the game, player one can turn down any card from the grid of pictures that does not fit the answer. For example, player one may turn down all pictures of people who do not have a mustache. After answering player one's question, player two can immediately ask his or her own question and perform all of the same actions that player one did. The game follows this trading of questions until the players have eliminated enough people to guess the adversary's mystery person. For each turn, a player can guess the mystery person or ask a question, but not both. Once any of the players make a correct guess, the match ends.

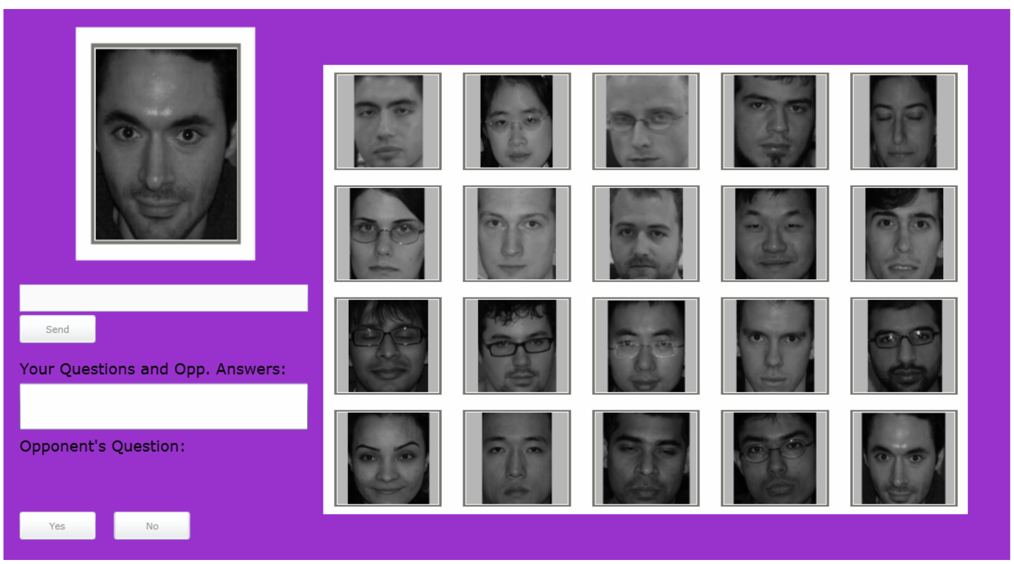

Figure 13 and Figure 14: The grid of pictures and a mystery person as seen by two players in a match. Notice that the pictures on both grids are different, but the people are the same.

One interesting thing to note regarding the game approach is that the cards each player has in his or her grid of pictures are different pictures that represent the same people. The card of a person on player one's grid, for example on index (1, 1), contains a different picture from that of player two's (1, 1) card, even though they are of the same person. These pictures are different in that they might have been taken from different angles, or the person in it might have had a different haircut. Figures 13 and 14 give us a better idea how we implement this idea. This greatly differs from how "Guess Who?" works since what we propose with this thesis is not only to individually tag face pictures of a person but also to give a better description of a face picture.

We intended to use pictures this way since we want to attack the problem that FRSs have in dealing with the same situation. We want the users to play our GWAP and link all of the various pictures of one person by describing the common traits that exist between all of the pictures tagged to them. We want to capture how humans can identify human faces through a diversity of conditions, i.e., the conditions that make FRSs fail. More importantly, we would like to capture the capacity and subjectivity of the human mind for recognizing faces. In this way, we can understand our logic for recognizing people, and we can use our data to improve facial recognition.

This may lead to some doubts as to whether this method can actually work. It is understandable to ask ourselves this question, since we are trying to have a player ask questions about a picture of one person while the other player has a different picture of the same person on which to base his or her answers. We run the risk of having pictures in the system that do not have anything in common with others in the same profile. For example, one picture may have visible facial features and the other might be a picture of the back of the same person‘s head. This is where the voting system steps in and acts as the screener for bad pictures.

When users vote, they determine which pictures are acceptable to be used in the game. As mentioned before, users use voting as a check for picture quality and usefulness. However, what was not mentioned is that they can also compare the new picture to the ones that are in the profile of a person. Whenever a user decides to add a new picture into a profile, the new proposed picture will always be presented with two other pictures that pertain to a profile, or the user can check this person’s profile and see every picture in it. If a user is creating a new picture profile, the newly added picture will not be put to a vote until a second picture is added. Thus, the user who is voting can always compare pictures and make sure they belong together in the same profile.

One detail to note is that not every user can vote. To avoid cheaters or people who may disrupt the goal of the game, we use a player reputation rating system. Every player has a reputation score, and only players with good reputations may vote or add pictures in the system. The reputation of each player is determined at the end of each game, when we encourage players to rate each other by asking about the game experience provided by each player.

Players can rate each other positively or negatively, and the reputation of a player is calculated as a ratio of positive votes and negative votes. Players with a ratio of one or above can vote and add pictures, while those with ratios below one have permission only to play the game. By being allowed to play only matches, users with low reputations can redeem themselves by playing the game honestly and increasing their ratios. While not currently implemented, there are still plans to adding rating mechanisms for players that add pictures. In this fashion, we can also identify players who might be trying to harm the system.

5 System Details

5.1 The GWAP Platform

The game created to support this thesis used a Web platform framework called Microsoft Silverlight. It is a cross-browser, cross-platform Web plug-in that allows the creation of Web applications for Windows and Mac systems, currently in its third version. Also, there is a version being developed in a collaboration between Novell and Microsoft for Linux systems based on the second and first versions of Silverlight, called Moonlight, [23].

Silverlight applications can be written using any .Net programming language, Visual Studio 2008 and its Web Platform Express variant, or an environment plug-in for Eclipse. One thing to note is that even though Silverlight utilizes .Net languages, it does use the Common Language Runtime (CLR) that is shared by all of them. Instead Silverlight uses a modified CLR optimized to host Web applications.

The Silverlight framework is a competitor to Adobe's Flash, even though the technology employed in Silverlight is different from that of Flash. In reality, Silverlight contains a subset of the Windows Presentation Foundation (WPF), which separates graphical interface from the logic used in coding. WPF is built using DirectX and provides hardware acceleration and a variety of user interface features such as transparency, 3D drawing, and transforms.

Being a subset of WPF, the main difference between them is that Silverlight presents its applications using a language called Extensible Application Markup Language (XAML), a markup language based on XML, instead of the compiled applications used with Flash. The use of XAML allows developers to write less code aimed at making applications more readable to search engines, since they can read XAML files. Thus, developers can make their applications yield better indexing and, thus, search results by using Silverlight.

Another important fact of using Silverlight is that even though Silverlight uses a modified version of the .Net CLR, its applications still have the ability to communicate with any application created with the .Net framework. More specifically, any application that has access to the Windows Communication Foundation (WCF) can send and receive messages from and to a Silverlight application as long the information sent is supported by the Silverlight CLR.

5.2 GWAP Architecture

The GWAP developed for this thesis is based on client-server architecture using Visual Studio 2008. It uses ASP .Net to deal with the server-side functions of the application, such as handling user information and creating new users. Silverlight is used to code the game's user interface and the client side of the application. C# is the chosen language to handle the "code behind" files of every ASP and XAML pages, as well as the socket and policy servers. WCF is used to maintain communications among the client side, the game, and the server side.

As for the Web server and database, the XAMPP package was used, consisting of the Apache HTTP Server, MySQL, and interpreters for the PHP and Perl scripting languages. The machine used for testing the whole solution was a 2.0 GHz Pentium 4 computer with 2 gigabytes of RAM and 50 gigabytes of memory. A little over 4300 lines of code were written to create this GWAP system.

5.2.1 Server Side

The server side of the GWAP is responsible for the socket servers and providing services that are not related to the game itself. These services are those related to user profile manipulation, user registration, login, password recovery, and mailing systems. The creation and manipulation of user information is done using the Membership and Roles classes which are provided by the ASP platform. It is important to note that important information pertaining to a user, such as first and last names, are stored in a separate database table, which the Membership and Role services cannot access.

The Membership service provides a series of convenient functions allowing user information to be easily created, deleted, changed, retrieved, and used for user authentication. The Roles service provides a means to categorize users into groups that specify the restrictions on users within a series of Web pages. In the case of this GWAP, restrictions are applied not only to Web pages but also to the game. Here, roles are used to distinguish users who can or cannot contribute to the game, e.g., adding new face pictures or playing the game against other users.

Roles are also used to determine registered users from guests. Even though a person has to register with the GWAP system to access all of the services it provides, there is still an option for people to login as guests and play the GWAP. Guests in the system are treated as restricted users such that their contribution to the game is diminished since there is not a consistent manner in which to hold a guest player accountable for his or her in-game behavior. Guest login might be removed in the future, or have the roles changed if the potential contributions of such players are great enough, thus avoiding waste of facial recognition information.

The server side includes more than the Membership and Role services. More importantly, it contains the servers that are responsible for maintaining the data flow between the client and server sides of the GWAP system, namely the policy socket, and the socket servers. The policy socket server is responsible for dealing with the incoming connections of a remote client. It allows the server side to restrict what such a client may access, and it can avoid security threats such as denial of service, or reverse tunnel attacks, [26].

The Silverlight runtime allows the client side to connect to the GWAP server over TCP on port 943. It first downloads an XML security policy file that indicates to the client whether a connection is possible or not. If the downloaded policy file grants permission, then the client finally sets a connection on the GWAP server. Also, this security policy file specifies which ports through which a Silverlight can receive or send data, namely ports 4502 to 4534, [25, 26].

Once a connection is set between the client and the server applications any data sent or received by them is handled by the Socket server. The duty of this server is to handle data that pertains to the chat room, game lobby, and the game contained in the system. The chat room duties consist of simply relaying messages written by a logged user to all other logged users through the system's chat room. As of now, there is only one universal chat room in the system, but there are plans to create private chat rooms.

The game lobby duties of the server are a bit more complex than the chat room responsibilities. Here, the server behaves according to the input provided by the user, which can be a user asking the server to create a game room or to join a game room so that a game can be played. Also, once two users are in a room the server is responsible for relaying data between both users indicating whether both users are ready to start a match.

As soon as the Socket server receives confirmation from both users, it informs the client side to start a GWAP match. Once the match is started, the Socket server relays information from player to player that can contain information such as questions describing the mystery person, answers to these questions, and guesses about whom the mystery person is. When the match ends, depending on the actions of a player on the client side, the Socket server then receives information about the behavior of the players in the game.

5.2.2 Client Side

The client side of the GWAP system is the interface that provides the information managed by the server side that is needed to achieve the goal of the system. Currently, it provides the user with views for adding pictures, chatting, and playing the GWAP. The architecture used here is based on game states; i.e., each screen seen by the user acts as if it is a separate program, and all of the game states combine to make a complete game.

A game state is responsible for drawing its own animation, handling events associated with it, and starting and stopping its execution by releasing resources used throughout the game. The flow of execution of each game state is overseen by the game state manager (GSM), which ensures that states start and stop properly. Every screen displayed using Silverlight is a game state itself, with the exception of the chat room, [34].

The GWAP client starts as soon as a user logs into the system. Once the user is granted access to the game, the system initializes the GWAP client and sends user information from the GWAP server to the client through WCF. Unfortunately, the initialization and execution of pages with Silverlight is sequential, i.e., not multi-threaded. This causes a problem with the GWAP system because the retrieval of data from the server side using WCF is always executed after the initialization of a Silverlight page. Remember that user information must be known before the initialization of the game so that user scores and page customizations may correctly load.

This why the chat room is different; it forces the client to wait for the WCF data arrival before initializing the game. Once the chat room receives the WCF data it signals the GWAP client to retrieve the user information and to start the game. Unfortunately, this is the only way to retrieve WCF initialization, as Silverlight execution is not multi-threaded, and it is not possible to change the initialization sequence of a Silverlight application and its pages.

Once the game is started, the GSM initializes the welcome game state (WGS), which the user can skip or wait for its execution. As the WGS ends, the GSM switch states to the Menu Game State (MGS), as shown in Figure 15. Here, the user has two options: to enter the game lobby, i.e., the Game Lobby Game State (GLGS), or to enter the "Add a Picture" page, i.e., the Add Picture Game State (APGS).

Figure 15: From the main menu, users play, chat, or add pictures to the game.

In the GLGS state, as shown in Figure 16, the user can either choose to create a game room, as shown in Figure 17, and wait for an opponent, or to join a room and challenge another user. Regardless of the path taken, once a user enters a room, he or she needs to confirm that he or she is ready to start a match. Once two users occupy a room and both confirm that they are ready, the match starts. Before starting a match, users in a room can choose to abandon it.

If the user who created the room leaves it, he or she promptly returns to the lobby. If there is a challenger present, the moment the creator decides to leave the room, the client side asks the server side to notify the challenger that the room will be closed since its creator left. Then, both users are returned to the lobby. Otherwise, if only the challenger leaves a room, then a vacant spot opens up in this room.

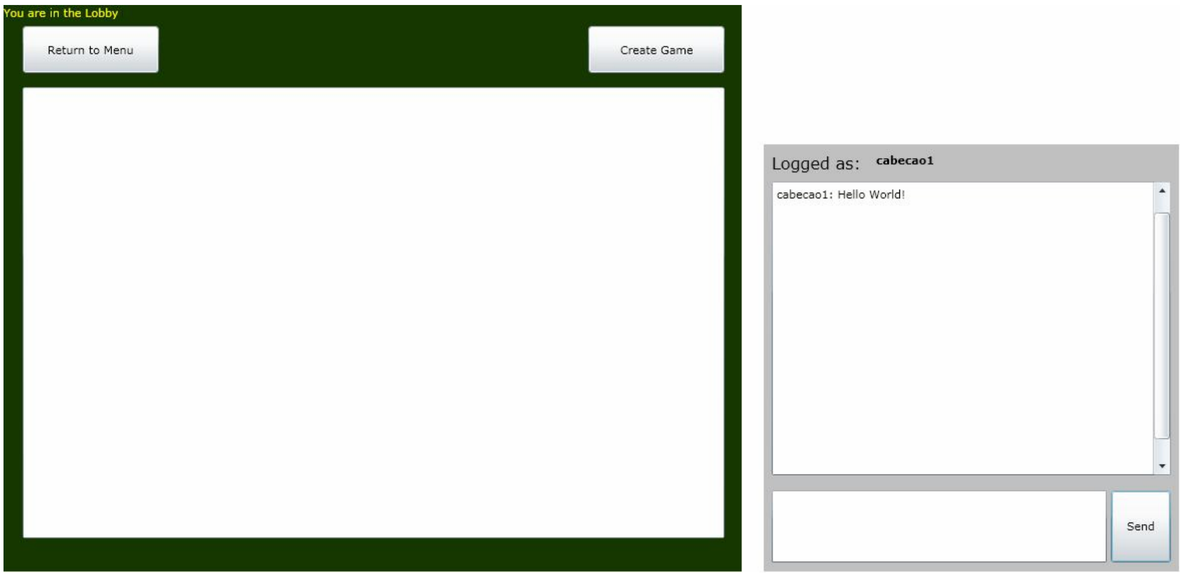

Figure 16: In the game lobby, a user can chat or create a room to play GuessMe.

When the confirmations of readiness to start a match are received from both users, the game starts automatically. First, a user is chosen to go first; as this player is chosen, the GWAP client asks the server to retrieve a set of face pictures to be used in the game. When the server retrieves these pictures, it searches for similar pictures in the database and then sends them to the other user playing the game.

After the facial pictures and its variations are loaded for both users, questions can be sent back and forth between players. Figures 13 and 14 demonstrate two players engaging in a match. When the first correct guess of the mystery person is made, the game ends, and a popup window asks the user to rate his or her opponent. Once an answer is given, both users return to the game lobby. One important thing to note is that, as of now, if a user quits the game early, the challenger returns to the game lobby automatically, the game is not saved, and the users are not able to rate their opponents.

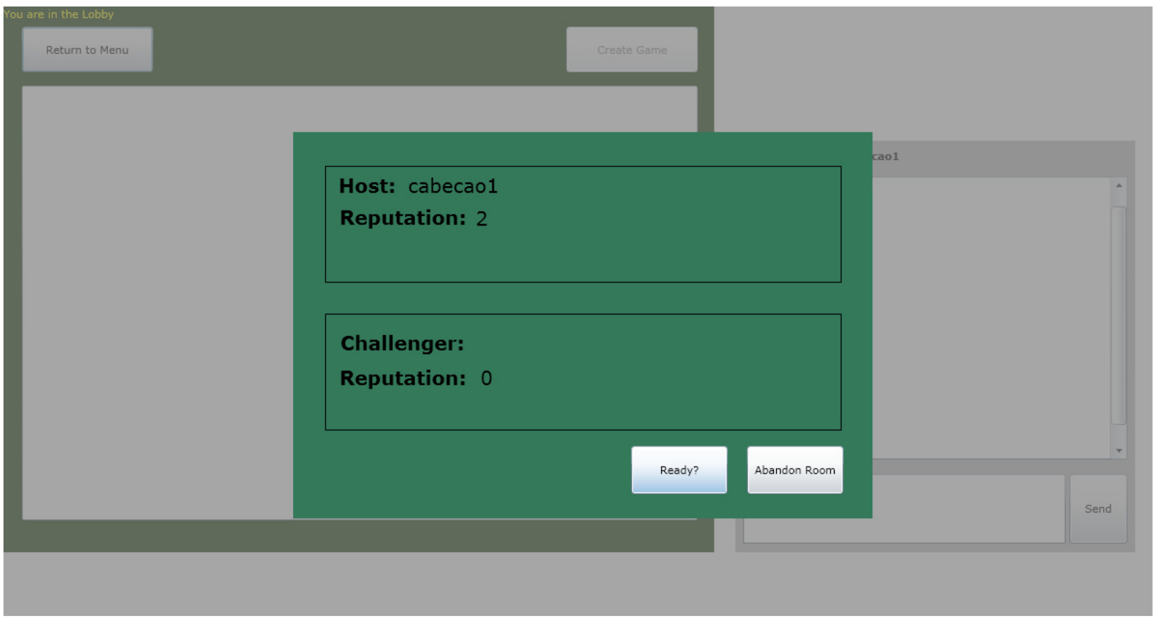

Figure 17: A user is ready to play a GuessMe match and creates a game room.

If the user intends to add a picture to the database of facial pictures, the Silverlight application enters the APGS state. Before adding a picture, the user must first search for the person contained it. Since every picture is named after somebody, the user performs searches by typing the name of the person they are looking for. Depending on the name, searches may produce a result or not. Depending on the results of a search the user can add a picture to an existing person's image profile or add a new profile to the GWAP database. If a new profile is added to the GWAP database, the user loads the image file and types the name of this person before clicking Save. Before this new person is added to the database, the client side runs the Double Metaphone algorithm that calculates a key value representing the name of a person that is used for future search matches. If a user finds the person he or she is seeking, then the user merely loads an image to an existent profile.

5.2.2.1 Double Metaphone Algorithm

The Double Metaphone algorithm (DMA), developed by Lawrence Philips, is a phonetic encoding algorithm, which generates encoded key values of four or three characters representing the sound of a word according to its English pronunciation. The DMA also provides alternate keys that represent the alternative pronunciation of non-homophonic words and names such as gigabyte, and foreign names of Spanish, Italian, French, Germanic, and Slavic origins, [17, 18].

The DMA used in the GWAP, as implemented by Adam Nelson, produces phonetic keys for any first, middle, and last names that a person might have, [17]. Before, being saved in the database, the produced keys and their alternatives are combined to create a tag key that is used by our picture search engine. This tag key preserves the position of the names entered and always adds the primary phonetic key before its alternative. For example, John Smith would bear a tag key of the format, first name key, key divider, first name alternative key, position divider, last name key, key divider, and last name alternative key.

One of the main reasons for using DMA is that DMA allows names that have spelling variations such as Stephen and Steven, to produce the same phonetic key. This is very important because people may misspell names when searching for people's names, which can produce less-than-desirable results. Thus, using DMA allows result sets to be much more meaningful since comparisons during a search can produce a bigger range of matches.

5.2.2.2 Levenshtein Distance Algorithm

The Levenshtein Distance algorithm (LDA) was invented by Vladimir Levenshtein, and it measures the distance between two different strings. It measures this distance by counting the number of edit operations, i.e., the number of character insertions, deletions, and substitutions that are necessary to transform one string into another. For example, if we calculate the distance between "aunt" and "ant" using LDA, we have an edit distance of 1 since it takes one deletion to turn "aunt" into "ant", [19, 20].

The LDA, as implemented by Sam Allen, is used to calculate the ranking of search results. More specifically, LDA is used to calculate the edit distance between the tag keys produced by the DMA. First, the tag key is divided into the keys pertaining to each word in a name. Once that is done, the LDA calculates the distance between the phonetic keys of a search result to the phonetic keys produced from the search query word.

The distance value between keys is added to the total distance between the search result and the search. This process repeats for every word in a name, and all search result words are compared to all query words. This is done because people will sometimes omit middle names or extra last names of a person during a search. The rankings are based on the total distance of a search result name compared to the queried name. The search results with the smallest distance are then placed at the top of the result table, while names with bigger distances are at the bottom.

6 Evaluation

6.1 Evaluation Preparation

For the GuessMe evaluation, we created a prototype that could show that the GWAP approach to facial recognition can be both efficient, and accurate. The prototype was developed as described in section 4, the only difference being our decision to exclude the voting system. We made this decision because we wanted to concentrate on the GuessMe's capabilities for amassing facial descriptions.

During the evaluation period, we gathered fifteen volunteers to play GuessMe. We explained to our volunteers the rules of the game and our intended goal of gathering facial descriptions so that we can use them in facial recognition. However, we did not tell them how to provide these descriptions; as stated in the rules of the "Guess Who?", we only asked our volunteers to ask only yes-or-no questions that would help them win a match.

Matches generally took five to seven minutes to complete, and players were not allowed to communicate with each other but could inquiry us in order to clear doubts related to the game play. Volunteers were asked to play at least two matches so that one match could help them get acquainted with the game's mechanism. The other matches were then played by the volunteers without assistance.

Volunteers usually could guess their mystery person after four questions and could make a correct guess on their seventh question. We used eighty black and white face pictures that were taken for a facial recognition assignment in a graduate course as the cards in our GWAP. These eighty pictures are split among twenty people, each having four face pictures taken at different angles. We had plans to take pictures of our volunteers and add them to the GWAP, but due to time constraints we decided to forego the plan.

6.2 Evaluation Results

At the end of every match, we asked our volunteers for ideas on how the game can be improved and for details that we might have left out of our implementation. After the evaluation was over, we assessed our findings during this period. We obtained encouraging and exciting results showing that the GWAP approach could be beneficial for facial recognition, as well as how we can improve GuessMe to better gather information.

The descriptions gathered during game play helped us to create a text-based silhouette of the faces of the people pictured in our database. These silhouettes are nothing more than grouped face descriptions and picture details. Even though we had limited time to test our prototype, these grouped descriptions can give us enough data to decide to which person they belong. This is possible because each person has unique traits that are easy to spot, allowing us to deduce easily whether they are distinguishable from each other or not.

Users tended to look at the grid of pictures to compare our face pictures and detect such traits that would help in their guessing. Due to the nature of our game, users tended to concentrate on these discernible traits as soon as possible since they wanted to win their matches faster than their opponents. For example, when pictures of people with moles, glasses, or broken noses were present in the grid of pictures, our volunteers were inclined to ask about them right away.

When these types of traits are not detected right away, or when our volunteers exhausted these options, an interesting phenomenon occurred. Unaware or unable to ask about easily discernible traits, users started concentrating their descriptions on the more general features of a person's face. Lip size, nose width and length, forehead length, and eyebrow size are some examples of the types of features that our volunteers used in their questions during the game.

All of the retrieved information from played matches allowed us to perform searches or compare people against one another. Our testing time provided us with enough data to differentiate between certain people and identify them. Due to the tendencies of our volunteers, those who had pronounced unique traits were the easiest to identify. For others, we would have to add more variables to our searches in order to pinpoint them.

Unfortunately, there were pictures of some people that were not used or were not used very often. Due to the limited time of our evaluation, we were not able to retrieve enough data to differentiate, i.e., to create detailed text-based silhouettes for every person contained in our database of face pictures. As mentioned in [1], if we were to make our GWAP public and test it for a longer period of time, the rate of facial descriptions would be much higher. Our pictures would gather much more detailed facial descriptions and thus, make our searches and comparisons much more accurate.

An important factor is the need for taboo descriptions, like those of the ESP game [1]. Our prototype did not contain such a mechanism and, as we expected, our volunteers tended to ask the same types of questions to discard possible non-matches for their opponent's mystery person. For example, when people with glasses were present in the grid of pictures of our volunteers would, most of the time, use their first question to ask whether a mystery person had glasses or not. The use of taboo descriptions would be essential to ensure variety and greater precision in facial descriptions if our GWAP were made publicly available.

As we mentioned during our evaluation period, we found that not every picture got a chance to be used. This raised an important question regarding users adding their own pictures. If users were allowed to add as many pictures as they wanted, these pictures could run the risk of not being used in our game. We would, thus, lose important information that could be added to the description profile of persons in the database.

To avoid such losses, we plan to limit the number of pictures that a user can have in his or her image profile. As in [1], in order to allow for a greater variety of pictures, which will lead to richer descriptions, we could stop using pictures after they reach a certain number of descriptions. In this way, we could ensure that different pictures and pictures with different angles, environmental conditions, and so on, are used in our GWAP, thus enabling users give us diverse facial descriptions of a person.

7 Conclusion

7.1 Conclusion

In this thesis, we approach the current problems faced by facial recognition systems with a “game with a purpose”. We show how we can utilize the superiority of our vision to better match facial pictures from controlled and uncontrolled sources. This approach allows us to deal with weaknesses that affect today’s FRSs. Using a method simpler than that of today’s FRSs, we facilitate the retrieval of descriptive information regarding the features contained in a facial picture.

We base our GWAP on the “Guess Who?” board game in order to channel the human capacity for creating and comprehending verbal cues related to a facial picture. These cues are used by the players of our game to describe and recognize faces displayed by the game. We demonstrate how our approach to facial recognition can provide accurate descriptions of a person and how the data retrieved from our GWAP can be used to create better face recognition methods.

7.2 Future Work Discussion

As it was mentioned in section 4.2, the GWAP system used to support this thesis is a prototype of its envisioned final version. There are still some features to be added that would make gathering facial descriptions and new pictures more effective. These changes would affect not only the features that are currently provided but also the way in which the game is played. The most important improvement of our system is related to the GuessMe game play and how users enter facial descriptions.

As of now, users enter facial descriptions during a match by typing the whole question into a text box. After receiving an answer, our system saves the whole question into our descriptions database table. Saving a description in this manner introduces several nonessential words such as verbs and pronouns. During the development stage of the GWAP, we thought about filtering these words using text parsers. Unfortunately, this adds unnecessary overhead in our system and unreliability that may cause our system to filter words that are important to a description. To solve this problem, we decided to take a cue from the Verbosity GWAP and use sentence templates that would allow users to type only data that is pertinent to facial features.

Even though we settled early on this method of inputting data, we decided to leave it out of from our current system implementation. We decided to remove sentence templates for now because while they allow us to easily retrieve descriptions, they also introduce difficulties for users typing questions. For example, very specific templates may not allow users to ask question the types of questions they want. In the future, we would like to explore this matter more deeply, since it is very important that we provide sentence templates that do not undermine a person's descriptive power. One solution that we would like to research is the use of partial sentence templates.

Another game play aspect that we would like to add to our game is the saving of descriptions for the pictures in the grid of pictures. As mentioned before, users play a GuessMe match by asking several yes-or-no questions to each other. According to the answer of the opponent, a user would then turn down the cards that do not fit it. As of now, we do not link these descriptions with the grid pictures, but the action of turning down or leaving a card up tells us whether a description fits a picture in the grid or not. Saving these descriptions for the grid pictures may result in the collection of valuable facial descriptions.

To accommodate this idea in our GWAP, some changes would have to be made to its game play. As of now, users playing a match may turn cards face down or up at any time they please. This can be troublesome for us since we would not know when a description truly fits a picture in the grid. The first action that we plan to take is to allow users to turn a card down only when they receive answers to their questions and before asking a new question. Once they ask a new question, the GWAP system would save these descriptions for the grid picture. There are weaknesses and strengths to this approach, and we intend to investigate it in the future.

Another feature that we would like to add to our implementation is the option of editing pictures before adding them. Here we took a cue from networking Web sites such as Facebook and Orkut, which let users enter pictures and tag the people present in it. When a user tags a person, he or she has the option of specifying where this person is located in the picture. In our implementation, just as these sites do, we would ask our users to tag the people present in the picture and locate their faces. Similarly to the current method of adding pictures, our system would then perform searches on the names of the tagged people and retrieve the profiles probably related to tagged people. The user would decide how to add the pictures in the same way as described in sections 4.2 and 5.2.2.

Other features that we would like to add to our GWAP are a merit per vote system, a score system, and support for other languages in our GWAP. The merit per vote system would take care of the unlimited number of times a user can vote. This system would let user vote a limited number of times, by giving a certain number of votes per month according to their reputation and current score. Lastly, we would like to make this game reachable to people of all languages so that we can retrieve more data on how different people from different locations and cultures describe a face.

References

[1] Luis von Ahn, and Laura Dabbish, “Labeling Images with a Computer Game”, ACM Conference on Human Factors in Computing Systems (CHI 2004), April 2004.

[2] Luis von Ahn, Ruoran Liu and Manuel Blum, “Peekaboom: A Game for Locating Objects in Images”, ACM Conference on Human Factors in Computing Systems (CHI 2006), April 2006.

[3] Luis von Ahn, Mihir Kedia and Manuel Blum, “Verbosity: A Game for Collecting Common-Sense Facts”, ACM Conference on Human Factors in Computing Systems (CHI 2006), April 2006.

[4] Luis von Ahn, Shiry Ginosar, Mihir Kedia, Ruoran Liu and Manuel Blum, CHI 2006, “Improving Accessibility of the Web with a Computer Game”, April 2006.

[5] George W. Furnas, Kevin Fox, Caterina Fake, Marc Davis, Luis von Ahn, Cameron Marlow, Joshua Schachter, Mor Naaman, Scott Golder, “Why Do Tagging Systems Work?”, CHI 2006 (Panel), April 2006.

[6]David G. Stork and Chuck P. Lam, Ricoh Silicon Valley, "Open Mind Animals: Insuring the quality of data openly contributed over the World Wide Web", AAAI Technical Report WS-00-05 (www.aaai.org), 2000.

[7] Chad Carson and Virginia E. Ogle, "Storage and Retrieval of Feature Data for a Very Large Online Image Collection", Computer Science Division, University of California at Berkeley, pages 21-28.

[8] Kobus Barnard and David Forsyth, "Learning the Semantics of Words and Pictures", Computer Division, University of California, Berkeley.

[9] Ingmar Weber, Stephen Robertson, Milan Vojnovic, "Rethinking the ESP Game", Technical Report MSR-TR-2008-132, September 2008.

[10] Ling-Jyh Chen, Bo-Chun Wang, Kuan-Ta Chen, "Playing GWAP With Strategies", Institute of information Science, Academia Sinica.

[11] Luis von Ahn and Laura Dabbish, “Designing Games With a Purpose”, Communications of the ACM, August 2008, vol. 51, no. 8, pages 58-67.

[12] Open Mind Initiative, “Open Mind Initiative”. Internet: http://www.openmind.org/, January 1, 2001, [February 2009]

[13] Luis von Ahn, “GWAP – Play games that change the Web”. Internet: http://www.gwap.com/gwap/, [February 2009]

[14] Jon Chamberlain, Udo Kruschwitz, Massimo Poesio, “Phrase Detectives”. Internet: http://anawiki.essex.ac.uk/phrasedetectives/, [February 2009]

[15] “FoldIt – Solve Puzzles for Science”. Internet: http://fold.it/portal/, [February 2009]

[16] Luis von Ahn, Ben Maurer, Colin McMillen, Mike Crawford, David Abraham, Edison Tan, “ReCAPTCHA”. Internet: http://recaptcha.net/, [February 2009]

[17] Lawrence Phillips, “The Double Metaphone Search Algorithm”. Internet: http://www.ddj.com/cpp/184401251, June 2000, [June 2009]

[18] Adam Nelson, “Double Metaphone implementation by Adam Nelson”. Internet: www.codeproject.com/KB/string/dmetaphone1.aspx, (last updated) March 19, 2007, [June 2009]

[19] Wladimir Levenshtein , “The Levenshtein Algorithm”. Internet: http://www.levenshtein.net/, [June 2009]

[20] Sam Allen, “Levenshtein algorithm Implementation”. Internet: http://dotnetperls.com/levenshtein, (last updated) January 8, 2010, [June 2009]

[21] Wikipedia, “Guess Who ?”. Internet: http://en.wikipedia.org/wiki/Guess_Who%3F, [February 2009]

[22] eHow Contributing Writer, “How to play Guess Who”. Internet: http://www.ehow.com/how_2054106_play-guess-who.html, [February 2009]

[23] “Get Started the Official Microsoft Silverlight Site”. Internet: www.silverlight.net/getstarted/, [February, 2009]

[24] MSDN Silverlight Developer Center, “Silverlight”. Internet: http://msdn.microsoft.com/en-us/library/cc838158(VS.95).aspx, [March 2009]

[25] Microsoft Developer Network Library, “Silverlight: Working with Sockets”. Internet: http://msdn.microsoft.com/en-us/library/cc296248(VS.95).aspx, [April 2009]

[26] Microsoft Developer Network Library, “Network Security Access Restrictions in Silverlight”. Internet: http://msdn.microsoft.com/en-us/library/cc645032(VS.95).aspx, [April 2009]

[27] Microsoft Developer Network Library, “Accessing Web Services in Silverlight”. Internet: http://msdn.microsoft.com/en-us/library/cc296254(VS.95).aspx, [April 2009]

[28] Dan Wahlin, “Creating a Silverlight 2 Client Access Policy Socket Server”. Internet: http://weblogs.asp.net/dwahlin/archive/2008/06/08/creating-a-silverlight-2-client-access-policy-socket-server.aspx, June 8, 2008, [April 2009]

[29] Dan Wahlin, “Pushing Data to a Silverlight Client with Sockets”. Internet: http://weblogs.asp.net/dwahlin/archive/2008/04/10/pushing-data-to-a-silverlight-client-with-sockets-part-i.aspx, April 10, 2008, [April 2009]

[30] Jayan Nair, “Asynchronous Socket Programming in C#”. Internet: http://www.codeguru.com/csharp/csharp/cs_network/sockets/article.php/c8781, October 6, 2005, [June 2009]

[31] Matthew Casperson, “Silverlight Game Programming Tutorial - State Management”. Internet: http://www.brighthub.com/internet/web-development/articles/15029.aspx, (last updated) Dec 15, 2009 [June 2009]

[32] Micorsoft Corporation, “The Official Microsoft Silverlight Forums”. Internet: “http://forums.silverlight.net/forums/”, [May 2009]

[33] Micorsoft Corporation, “ASP .NET Forums”. Internet: http://forums.asp.net/, [May 2009]

[34] Daniel Schuller, “C# Game States”. Internet: http://einfall.blogspot.com/2005/02/game-states.html, February 2005, [May 2009]

[35] Smartyp, “Matchingo - A Silverlight Memory Game”, Internet: http://matchingo.codeplex.com/, (last updated) Sep 29, 2009, [May 2009]

[36] Brad Merril, “C# Regular Expressions”. Internet: http://oreilly.com/windows/archive/csharp-regular-expressions.html, January 18, 2001, [June 2009]

[37] M.A. Turk and A.P. Pentland, “Face Recognition Using Eigenfaces”, IEEE Conf. on Computer Vision and Pattern Recognition, pages 586-591, 1991

[38] Laurenz Wiskott, Jean-Marc Fellous, Norbert Kruger, and Christoph von der Malsburg, “Face Recognition by Elastic Bunch Graph Matching”, Intelligent Biometric Techniques in Fingerprint and Face Recognition, eds. L.C. Jain et al., publ. CRC Press, ISBN 0-8493-2055-0, Chapter 11, pages 355-396, 1999.

[39] Jun Zhang, Yong Yan, and Martin Lades, “Face Recognition: Eigenface, Elastic Matching, and Neural Nets”, PROCEEDINGS OF THE IEEE, Vol.85, No. 9, pages 1423-1435, September 1997

[40] FedericoMatta, Jean-LucDugelay, “Person recognition using facial video information: A state of the art”, Journal of Visual Languages and Computing 20, pages 180–187, 2009

[41] Ambika Suman, “Automated Face Recognition: Applications with Law Enforcement. Market and Technology Review”, British National Policing Improvement Agency (NPIA), June 2006

[42] Kelly Smith, Arun Ross, and Dirk Colbry, “Face Recognition”, National Science and Technology Council (NTSC), Committee on Technology, Committee on Homeland and National Security, Subcommittee on Biometrics, August 2006

[43] Wikipedia, “Face Recognition System”, Internet: http://en.wikipedia.org/wiki/Face_recognition, [June 2009]

[44] Kevin Bonsor and Ryan Johnson, “How Facial Recognition Systems Work”. Internet: http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/facial-recognition5.htm, [June 2009]

[45] Ron Kimmel, and Guillermo Sapiro, “The Mathematics of Face Recognition”, Society for Industrial and Applied Mathematics (SIAM). Internet: http://www.siam.org/news/news.php?id=309, April 30, 2003, [June 2009]

[46] Mark Williams, “Better Face-Recognition Software”, Technology Review published by MIT. Internet: http://www.technologyreview.com/computing/18796/page1/, May 30, 2007, [June 2009]

[47] American Civil Liberties Union (ACLU), “Q&A on Face-Recognition”. Internet: http://www.aclu.org/technology-and-liberty/qa-face-recognition, September 2, 2003 [June 2009]

[48] Wikipedia, “United States Visitor and Immigrant Status Indicator Technology”. Internet: http://en.wikipedia.org/wiki/United_States_Visitor_and_Immigrant_Status_Indicator_Technology, [August 2009]

[49] US Homeland Security, “US VISIT”. Internet: http://www.dhs.gov/files/programs/usv.shtm, [August 2009]

[50] John D. Woodward, Jr., Christopher Horn, Julius Gatune, and Aryn Thomas, “A Look at Facial Recognition”, RAND Public Safety and Justice for the Virginia State Crime Commission, ISBN: 0-8330-3302-6, 2003

[51] Amy Zalman, “Profiling Passengers Pros & Cons”. Internet: http://terrorism.about.com/od/issuestrends/i/Profiling.htm, [June 2009]