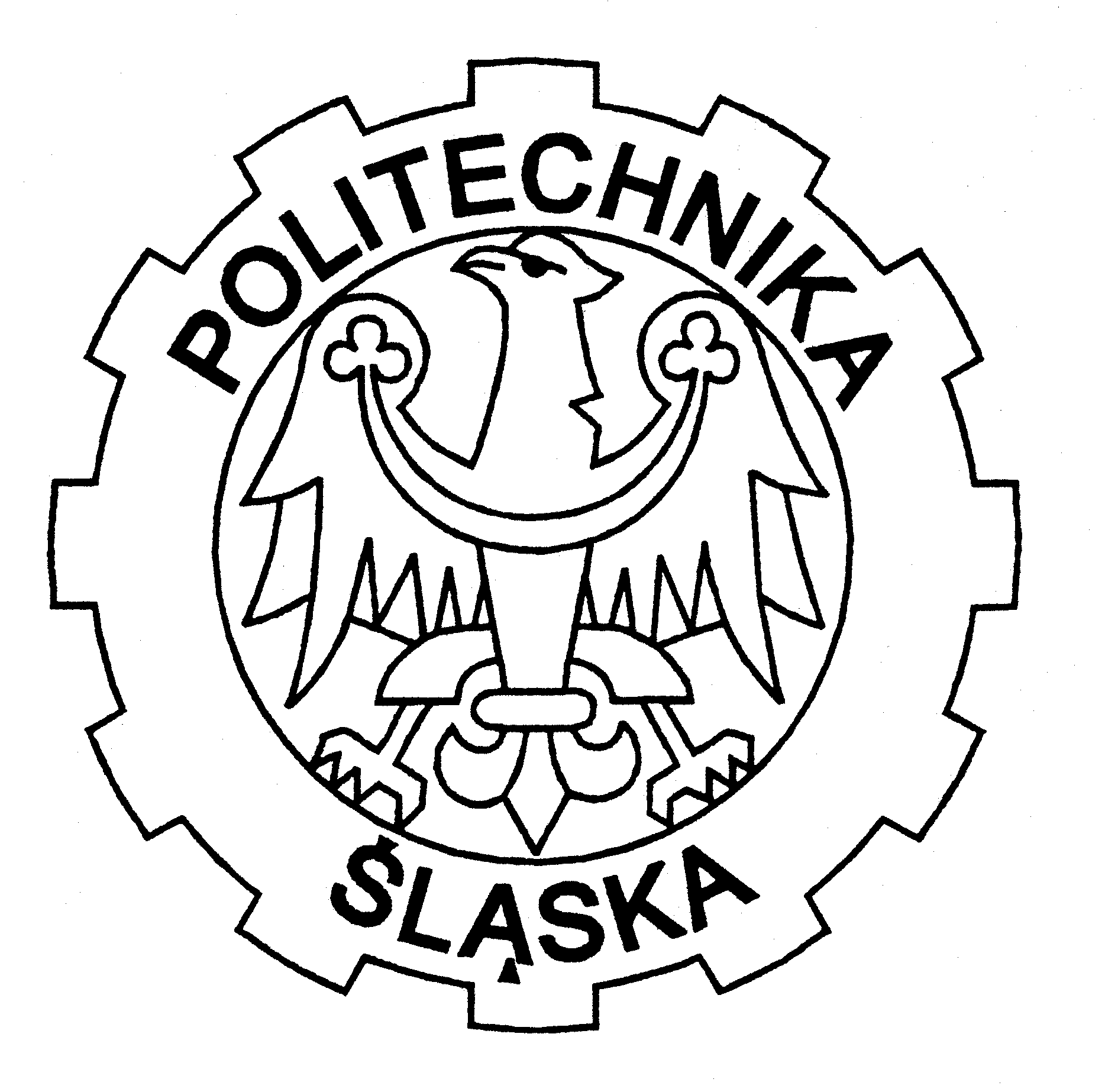
The Silesian University of Technology

– Faculty of Automatic Control, Electronics and Computer Science –



**Master Thesis**

Interactive business cards based  
 on Augmented Reality idea

**Author: Filip Sitko**

**Supervisor: dr inż. Michał Kawulok**

**Gliwice, September 2011**

Contents

1. [Introduction 3](#_Toc298453783)

[1.1. Problem Definition 3](#_Toc298453784)

[1.2. History 6](#_Toc298453785)

[1.3. Applications 9](#_Toc298453786)

[1.4. Requirements 12](#_Toc298453787)

1. [Augmented Reality problem analysis 13](#_Toc298453788)

[2.1. Display 13](#_Toc298453789)

[2.2. Video capture 21](#_Toc298453790)

[2.3. Marker detection 24](#_Toc298453791)

[2.4. Square detection **Error! Bookmark not defined.**](#_Toc298453792)

1. [Project Design 30](#_Toc298453793)

[3.1. Library choice 30](#_Toc298453794)

[3.2. Development Process 32](#_Toc298453795)

[3.3. Project Algorithm 33](#_Toc298453796)

[3.4. GUI 34](#_Toc298453797)

1. [Internal Specification 34](#_Toc298453798)

[4.1. Main program functions 34](#_Toc298453799)

[4.2. Graphical User Interface 34](#_Toc298453800)

1. [External Specification 35](#_Toc298453801)

[5.1. ‘How to’ instruction 35](#_Toc298453802)

[5.2. Errors handling 35](#_Toc298453803)

1. [Testing and results analysis 36](#_Toc298453804)

[6.1. Marker choice analysis 36](#_Toc298453805)

[6.2. Environment dependencies 36](#_Toc298453806)

[6.3. Threshold methods 36](#_Toc298453807)

[6.4. Displaying static image and video 36](#_Toc298453808)

[6.5. Camera parameters 36](#_Toc298453809)

1. [Summary 37](#_Toc298453810)
2. [Bibliography 38](#_Toc298453811)
3. [Contents of the CD 40](#_Toc298453812)

**Chapter 1**

# Introduction

Perception can be certainly considered as an essential factor of human life. Every information about the environment we are living in is received by our senses. Despite of human body imperfections people always tried to improve their perception skills and their inventions helped to find the new ways to explore and understand the surrounding world. Augmented Reality idea introduces a new dimension of perception and opens vast new possibilities that will aid nearly every area of human life.

## Problem Definition

In fundamental terms, the Augmented Reality, often abbreviated to **AR** is an area of Mixed Reality that refers to the real-time view of a physical world which is augmented by elements generated or triggered by a computer input and can be considered as the connection between the real world and the virtual one. Given real subject image captured by a camera is processed and combined with virtual layers (such as graphics, sounds, data and even smells which are triggered by computer input).

Most common definition was created by Ronald Azuma who described it as follows: “*Augmented Reality* (AR) is a variation of *Virtual Environments* (VE), or Virtual Reality as it is more commonly called.

VE technologies completely immerse a user inside a synthetic environment. While immersed, the user cannot see the real world around him. In contrast, AR allows the user to see the real world, with virtual objects superimposed upon or composited with the real world. Therefore, AR supplements reality, rather than completely replacing it. (…) this survey defines AR as systems that have the following three characteristics:

* Combines real and virtual
* Interactive in real time
* Registered in 3-D” **[1]**



Figure 1.1.1. Real desk with virtual lamp and two virtual chairs. (Courtesy ECRC)

Augmented Reality is commonly mistaken with Virtual Reality, hence to provide better understanding of Augmented Reality the Paul Milgram’s Virtuality Coninuum**[2]** graph (Figure 1.1.2.) should be introduced to show the general classification of Mixed Reality areas and their unique features.

**MIXED REALITY(MR)**

**VIRTUAL  
ENVIRONMENT**

**AUGMENTED  
VIRTUALITY (VR)**

**AUGMENTED  
REALITY (AR)**

**REAL  
ENVIRONMENT**

Figure 1.1.2. Paul Milgram’s Virtuality Continuum graph.

Nowadays technology based on Mixed Reality is rapidly developed and distinct boundaries of each area are impossible to define.**[3]** However to remark the main differences between them each one can be described by a short definition and unique features:

* **Real environment:**

View of the real, physical world as it can be perceived directly.

* **Augmented Reality(AR):**

Real world view augmented by a computer- generated inputs which create a possibility of interaction.

* **Augmented Virtuality(AV):**

Virtual space view augmented by a real world inputs most commonly used for Human-Computer Interaction(HCI).

* **Virtual Reality(VR):**

Fully simulated world view which provides environment elements controlled by a real world input.

Reproduction Fidelity (FR) of the virtual image should be proportional to the quality of captured image of the real world.  
Basing on Reproduction Fidelity graph (Figure 1.1.3.) it can be noticed that to obtain the most realistic views combination the computer-generated models details should be real world image fidelity dependent (e.g. High fidelity 3D model with shadings and textures would look unnatural if projected on low-resolution monoscopic video).

**STEREOSCOPIC VIDEO**

**HIGH DEFINITION VIDEO**

**3D HDTV**

**COLOUR VIDEO**

**CONVENTIONAL  
(MONOSCOPIC)  
VIDEO**

**SHADING, TEXTURE, TRANSPARENCY**

**REAL-TIME  
 HIGH FIDELITY 3D ANIMATION**

**VISIBLE SURFACE IMAGING**

**RAY TRACING, RADIOSITY**

**SIMPLE WIREFRAMES**

Figure 1.1.3. Paul Milgram’s Reproduction Fidelity(FR) graph based on Naimark’s Taxonomy.

## History

Augmented Reality in form which is known nowadays was imagined as a technology of the future since the first computer was designed. People could observe multiple applications of AR in science-fiction movies but did not know that this concept was already researched. Rapid development of AR can be noticed within last 10 years and is commonly considered to be the one of the inventions of XXI century.

This belief is incorrect as the beginning of AR is dated for 1962 as Morton Heilig created a bicycle simulator called Sensorama (Figure 1.2.1.) based on multimodal (multi-sense) technology. The machine could provide stereoscopic 3D vision in wide-angle view, body tilt, stereo sound and even wind tracks and smells triggered as the film was displayed. As nearly all senses were involved during the simulation Sensorama gave the general idea of Augmented Reality which was developed further using the computer.

Ivan’s Sutherland invention of the first head-mounted display named Sword of Damocles (Figure 1.2.2.) in 1968 was the next m ile step in AR history. Allowing to see computer-generated wireframe rooms according to user head position it gave the background for user interaction with virtual world.

In 1975 Myron Krueger established artificial reality laboratory called Videoplace. Based on cameras, projectors and computer hardware it created an interactive artificial environment for the first time.

These inventions aroused interest of Augmented Reality and from this point it became a popular subject of studies and computer science research. In 1989 Jaron Lanier coined the ‘Virtual Reality’ phrase by leading the company that sold VR goggles and gloves. In 1990 Tom Caudell an aircraft manufacturer popularized ‘Augmented Reality’ phrase.

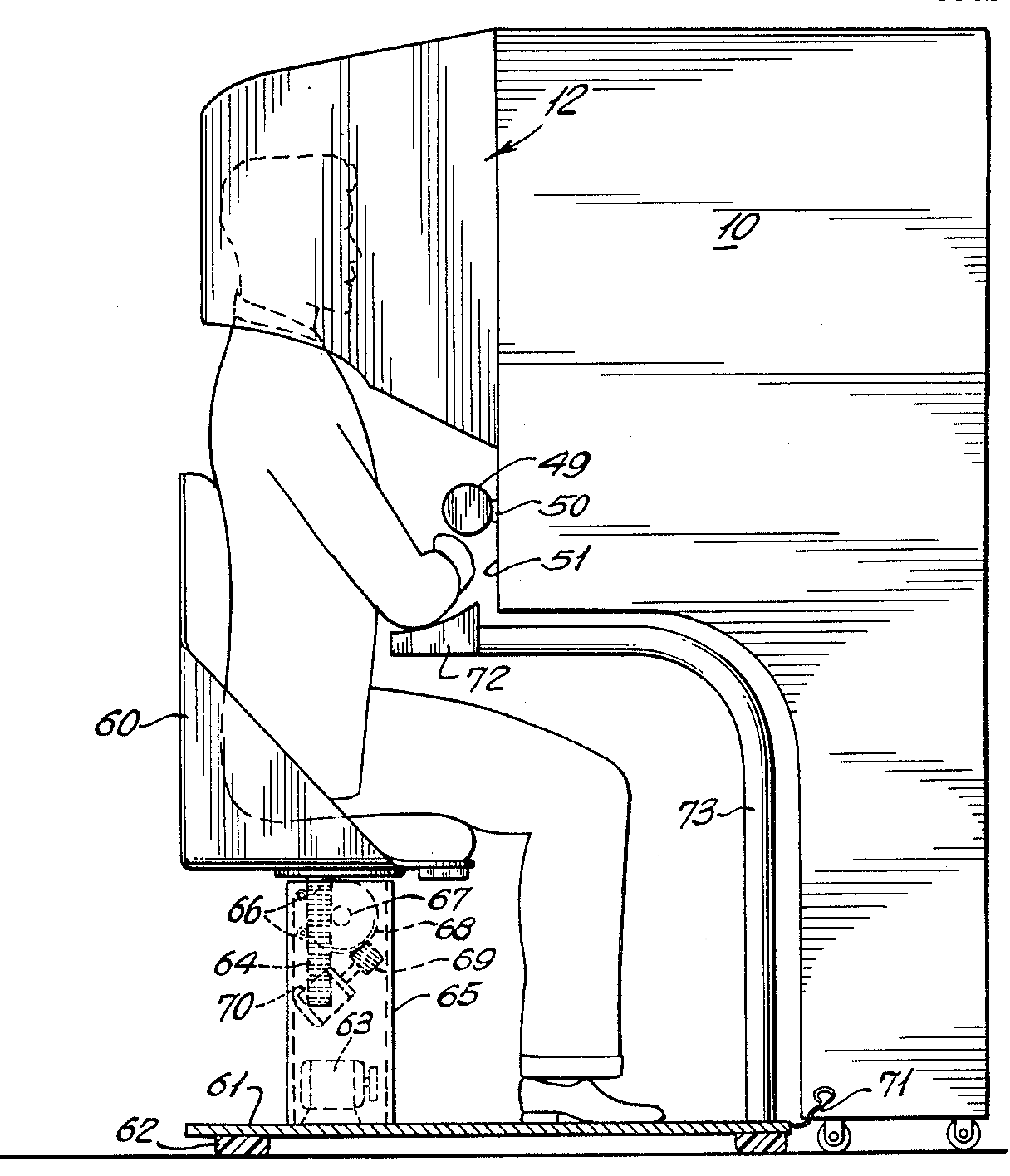


Figure 1.2.1. Morton Heilig's Sensorama. [Figure 5 of U.S. Patent #3050870]

Figure 1.2.2. Ivan’s Sutherland Sword of Damocles**[4]**

1994 introduced Paul Milgrim’s Vrtuality Continuum concept (Figure 1.1.1.) and classified Augmented Reality as an area of Mixed Reality specifying it’s boundaries and unique features.

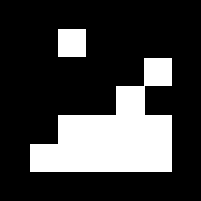
In 1996 Jun Rekimoto presents 2D matrix markers (Figure 1.2.3) as a square-shaped barcodes, one of the first marker systems to allow simultaneously identify real world objects and estimate their coordinate systems.

Figure 1.2.3. Jun Rekimoto’s exemplary 2D matrix marker**[4]**

Commonly known definition of the Augmented Reality term and it’s field was defined by Ronald Azuma in “A Survey of Augmented Reality” in 1997. His paper survey described first AR systems from basis and it is popular material even for today’s research purposes.

The release of the ARToolkit- open source computer vision tracking library developed by Hirokazu Kato at the HITlab in 1999 was the real milestone for the AR research. It began new wave of interest among developers and opened a new possibilities of AR programming.

The first game based on Augmented Reality concept “ARQuake” was developed in 2002 and started by it’s inventor Bruce H. Thomas. It provided outdoor first-person shooter based on virtual environment generated upon real world captured images.

ARToolkit was redesigned and ported to Adobe Flash (FLARToolkit) by Tomohiko Koyama (Saqoosha) in 2009 bringing Augmented Reality to web browsers and starting a new trend wave of web- based AR applications.



Figure 1.2.4. ARToolkit based 3D model projection **[4]**

Figure 1.2.5. ARQuake outdoor combined views image **[4]**

Augmented Reality technology is rapidly developed nowadays for variety of platforms starting with personal computers through mobile devices and ending with touch screens and technical aparatures.According to Rick’s Oller: “The future of augmented reality holds tremendous promise. With display technology getting better, smaller, lighter and requiring less power every year, it is only a matter of time before augmented reality displays can be fitted to an ordinary pair of glasses, or even contact lenses.” **[5]**

## Applications

Augmented Reality idea provides variety of new perception and interaction possibilities which are used in almost every field of human life. The applications of AR are practically limited only by imagination and technical development process. Although AR is rather new technology the rapid thrive in development brought it to daily life often staying unnoticed. Focusing only on common applications Augmented Reality can be used for:

* **Navigation:**AR can be used as a navigation tool which helps to define object positions in challenging environmental conditions. Combined with GPS functionality outdoor objects recognition and virtual routes can provide all the necessary data to enhance navigation process. Digital maps and object/human position indication aid the exploring and tracking tasks.
* **Sightseeing:**

Stored data with history and object descriptions, interactive animations and audio guides and even visual routes to places of interest are really attractive additions for sightseeing tours. AR technology can provide it almost everywhere, so usually head- mounted or mobile displays are designed for this purpose as their size and mobility are essential.

* **Military:**Detailed mission briefings, GPS-based navigation routes and location data, digital maps, enemy locations or even enhanced firing systems are introduced to the military with AR technology usage. Head- mounted display and goggles with addition of other devices (GPS receivers, orientation trackers, computers and handheld control devices) are the most suitable solutions for this propose as it requires to be mobile since it is used by troops.
* **Medicine:**

Medical students use the technology to practice surgery in a controlled environment. Visualizations aid in explaining complex medical conditions to patients. Surgery risk ratio is significantly reduced as surgeons get improved sensory perception during complicated operations. AR combined with MRI or X-ray systems can be an invaluable tool bringing all of them into one view.

* **Maintenance and task support:**

Using head worndisplay a mechanic can be aided by additional data, instructions and label for specified object’s parts. Repairing procedures can be provided in adverse environmental conditions and specialist training expenses are reduced by using simulations.

* **Advertising and promotion:**

Promotion through interactive animations, 3d models and games is becoming popular. New web services, company products and even movies are advertised with simple AR gadgets enticing to play and interact.

* **Entertainment:**AR technology influenced the gaming market starting new type of interactive entertainment applications and games to emerge. Bringing 3D virtual world into reality created fertile ground for developing new mobile and outdoor games and social such as sporting events and concerts.
* **Education:**Since beginnings AR technology stayed in association with educational institutions. Many AR research breakthroughs have been accomplished by college and universities teams, as prototypes and still developed devices were available as multipurpose educational tools. Providing possibility of real-time processing it can be used for presentations, training simulations and development research.
* **Industry:**Augmented reality can provide hands-free visual overlays of dynamic manufacturing information targeted to specific, highly controllable automated and semi-automated assembly environments. Computer generated virtual project prototypes can replace the real ones reducing the final product expenses.
* **Architecture:**Virtual models mockups and simulations could be projected on one platform aiding the design and planning process. As it provides possibility of collaboration on shared models it can be used as a powerful tool improving planning and communication process.
* **Translation:**Real-time dynamic subtitles display and text translation can really enhance communication process. Font and text recognition and even simple mathematical problems solution can be achieved using this unique feature.

These are only the main applications of Augmented Reality narrowed to the ones we are using nowadays.**[5][6]** Taking into account technical development progress we can expect more of them in near future as every unique area of life can be simplified, each object can be augmented to be more usable and each action can give extraordinary experience which cannot be obtained in real life.

Figure 1.3.1. Virtual fetus inside womb of pregnant patient. (Courtesy UNC Chapel Hill Dept. of Computer Science.)**[1]**

Figure 1.3.2. Head-up display for a fighter plane**[3]**

## Requirements

Final project’s design should provide simple AR elements to the final business cards labeled by 2D black and white markers. Created application should meet the following requirements:

* **Main detection algorithm written in C++:**

C++ usage should enhance the algorithm speed and influence on overall performance level.

* **High accuracy marker recognition:**

Business card markers must be detected and matched with selected template to display a virtual graphical element.  
High spectrum of marker angle and environment lightness level acceptance would improve the interaction process.

* **Real- time image processing (min 12 FPS):**

To provide high quality interaction acceptable Frames Per Second ratio has to be achieved. Lower FPS processing would create an illusion of delay.

* **Image and video display:**

Simple computer-generated graphical inputs should be combined with real, physical world image capture. Usage of AR technology in this case will project photos and videos that will aid the personal identification of the business card holder.

* **Graphical User Interface:**

Simple and intuitive GUI that would make the whole application more user- friendly and allow to use every program feature in a convenient way.

**Chapter 2**

# Augmented Reality problem analysis

There are several approaches to obtain the desired effect for Augmented Reality implementation. However each of them has the unique features designed for specific type of devices and environments in which they are used in. Some of them should not be used for this project purposes, hence strengths- weaknesses analysis of each approach property will reveal most suitable solution to achieve the project goals.

## Display

Augmented Reality based technology is classified in regard of displays used for combined computer- generated input and real physical world captured images visualization. According to Oliver Bimber and Ramesh Rascal:

“Augmented reality displays are image-forming systems that use a set of optical, electronic, and mechanical components to generate images somewhere on the optical pathin between the observer’s eyes and the physical object to be augmented” **[8]**.

**SPATIAL**

**HEAD-ATTACHED**

**HANDHELD**

**REAL OBJECT**

**PROJECTOR**

**RETINAL DISPLAY**

**HEAD- MOUNTED DISPLAY**

**HAND- HELD DISPLAY**

**SPATIAL OPTICAL  
SEE-THROUGH DISPLAY**

**PROJECTOR**

**PROJECTOR**

Figure 2.1.1. Image generation graph for augmented reality displays inspired by Oliver Bimber and Ramesh Raskar **[8]**

There are 3 major display techniques used for design of Augmented Reality based devices and applications:

* **Head-Attached displays:**

Head attached displays require the AR display system worn on user’s head. Depending on the image generation technology there are 3 main types distinguished:

**Retinal Displays:**

Instead of providing screens in front of the eyes, retinal displays utilize low-power semiconductors lasers or special light emitting diodes to scan modulated light directly onto retina of the human eye. The advantage of using this type of display is a higher resolution and a potentially wider field of view in comparison of screen-based displays. Vivid contrast, brightness and low-power consumption make it perfect for mobile outdoor applications (nowadays retinal displays can be fitted into casual glasses and goggles). However there are some disadvantages which made it not so popular. Due to low-power consumption and cheap components mainly monochrome (red) lasers utilized. Moreover due to the complete bypass of the ocular motor system by scanning directly onto the retina, the sense of ocular accommodation is not supported hence the focal length is fixed. Stereoscopic versions of retinal displays do not exist yet so the application possibilities are very limited.

**LASER(S), OPTICS AND SCANNING UNIT**

**BEAM SPLITTER**

**PROJECTED IMAGE ON RETINA**

Figure 2.1.2. Simplified diagram of a retinal display inspired by Oliver Bimber and Ramesh Raskar **[8]**

**Head Mounted Displays:**

Two different head-mounted display technologies exist to superimpose graphics onto the user’s view of the real world:

Video see-through: makes use of video-mixing and display the merged images within a closed-view head-mounted display.

Optical see-through head-mounted displays that make use of optical combiners (essentially half-silvered mirrors or transparent LCD displays). Main advantage of this technology are it’s mobility and possibility of full color spectrum and fidelity level image generation. Several disadvantages which are inherited from general limitation of head-attached display technology should be also noticed. Lack in resolution of generated image( limitations of attached miniature displays), limited field of view (due to limitations of applied optics), visual perception issues( Fixed focal length problem occurs as the eyes are constantly are forced to either continuously shift focus between the different depth levels, or perceive one depth level as unsharp- mostly for see-through displays), Increased incidence of discomfort due to simulator sickness because of head-attached image plane (especially during fast head movements). Optical see-through devices require difficult (user- and sessiondependent) calibration and precise head tracking to ensure a correct graphical overlay. Nevertheless head-mounted displays were the dominant display technology within the AR research field. They support mobile applications and multi-user applications if a large number of users need to be supported.

**MINIATURE DISPLAY**

**IMAGE**

**BEAM SPLITTER**

**COMPUTER**

**REFLECTED IMAGE**

Figure 2.1.3. Simplified diagram of an optical see through head-mounted display inspired by Oliver Bimber and Ramesh Raskar **[8]**

**CAMERA**

**COMPUTER**

**IMAGE**

**MINIATURE DISPLAY**

Figure 2.1.4. Simplified diagram of a video see through head-mounted display inspired by Oliver Bimber and Ramesh Raskar **[8]**

**Head- Mounted Projectors:**

Head-mounted projective displays redirect the frustum of miniature projectors with mirror beam combiners so that the images are beamed onto specified surfaces (usually retro-reflective surfaces because of their light-reflective visual properties) that are located in front of the viewer. As whole technology is projector-based it provides wider field of view as he effect of inconsistency and convergence is decreased in comparison with other HMD types, however there are also some shortcomings like limited projector generated image resolution and brightness, necessity of special display surfaces, strong light conditions dependencies and limitations to closed space environment (as the ceiling is necessary for system installation). These factors made head-mounted projector more suitable for Virtual Reality applications.

**PROJECTOR**

**COMPUTER**

**BEAM SPLITTER**

Figure 2.1.5. Simplified diagram of projector head-mounted

display inspired by Oliver Bimber and Ramesh Raskar **[8]**

* **Handheld displays:**

Hand-held display concept provides generating images within arm’s reach within single device which combines processor, memory, display, and interaction technology. Whole idea aims at supporting a wireless and unconstrained mobile handling, hence this technology can be utilized using a Tablet PCs, PDAs, personal digital assistants and mobile phones which brought the AR technology to mass market (nowadays AR based entertainment applications are very popular among smartphone users). Mobile optical see-through and hand-held mirror beam combiners and other mobile technology combinations also exist, however classic concept is the most common one due to it’s advantages like mobility, simplicity and cheap, available device components. There are also some disadvantages that should be introduced. The image analysis and rendering units are processor and memory intensive and require floating point units, what is critical for low-end devices (mobile phones, PDAs) and causes slow frame rates and decrease the precision of generated image. Mobile device screen size restricts the covered field of view, however since it can be moved and navigate through an information space which is definitely larger than a screen size this technology supports a visual perception phenomenon called Parks effect. Thus far mobile devices cameras were not adapted for AR tasks due to resolution or focus planes limitation, however recent rapid development of mobile devices technologies made them perfect for simple AR tasks. Unfortunately in comparison with other types of displays hand-held devices do not allow to complete any tasks in completely hand-free working environment.



Figure 2.1.6. a first prototype on a conventional

consumer cell phone. **[8]**

* **Spatial displays (SAR):**

Spatial displays refer to those display techniques that detach most of the technology from the user and integrate it into the environment. There are 3 unique approaches which differ in environment augmentation process:

**Screen-Based Video See-Through Displays:**

These systems make use of video mixing and display merged images directly onto monitor’s screen. As the output signal video can be provided for regular monitors or projectors the main advantage of this method is the hand-free usage convenience and wide range of observers for a single AR application instance. As to adapt this technique only off-the-shelf hardware components and standard PC equipment is required- this method is considered to be the most cost-efficient. Naturally it also has disadvantages connected with general modern screens issues. Limited screen resolution and size decrease the image quality and user’s field of view, however modern development in displays and projectors area makes these factors negligible. The most disadvantage is a remote view rather than a see-through concept which decrease the level of view interactivity.

**Spatial Optical See-Through Displays:**

The main idea of this technique is to generate images that are aligned within the physical environment using spatial optical combiners, such as planar or curved mirror beam combiners, transparent screens, or optical holograms. This method provides more realistic image marked out with easy eye accommodation property, high and scalable resolution, wider field of view and more controllable environment( by tracking, illumination level etc.).Spatially aligned optics generate some shortcoming as lack of support for mobile devices, observers and environment interaction limitations. Due to a screen size limitation the virtual objects outside the display area are unnaturally cropped which is called a window violation effect. Since the generated image is realistic and superimposed to natural environment this technique is commonly used to create holograms and optical illusions.

****

Figure 2.1.7. Transparent projection screen. **[8]**

**Projection-Based Spatial Displays:**

Projector-based spatial displays use front-projection to seamlessly project images directly on a physical objects surfaces instead of displaying them on an image plane (or surface) somewhere within the viewer’s visual field. To enhance the image quality and interaction level the whole system is a combination of single or multiple static or steerable projectors. Main disadvantage of this technique is shadow- casting of physical objects and interacting users due to front-projection. Display area constrained to the size and shape of the physical object and increased complexity of consistent geometric alignment and color calibration as the number of applied projectors increases makes the whole system hard to calibrate especially for 3D objects. This technique is commonly used for presentations and outdoor structures visualizations.

As the purpose of this project is based on the business card image augmentation it is aimed to a bulk consumer. Taking it into consideration the display technique should mark out with availability and mobility to provide fast and simple augmentation method. These requirements can be satisfied by hand-held displays and spatial screen-based video displays. Hence the PC application can be applied for wider spectrum of consumers and cost-efficient approach needed the screen-based video display technique was chosen for this project purposes. This solution applied to notebooks combines advantages of both techniques bringing all the disadvantages to the minimum.

## Video capture

Choice of Augmented Reality display technique depends on the video capture method used for designed application. As AR environment should present the area as 3D plane there are two main methods that provide this desired effect. To introduce them both it is essential to understand the main properties of human perception. Interesting definition and main principles of 3d viewing was formulated by Hong Hua:

“Human eyes rely on many visual cues to perceive and interpret depth in the real world. Such depth cues can be monocular or binocular. Monocular depth cues are observed only with one eye and common examples include perspective, occlusion, texture gradients, distribution of light and shadows, and motion parallax. Binocular depth perception is based on displacements (i.e.binocular disparity) between the projections of a scene object onto the left and right retina due to eye separation. The binocular disparity is processed by the brain, giving the impression of relief in an effect known as stereopsis. Stereoscopic displays enable depth sensation by exploring the binocular disparity.”**[9]**

**Single camera video capture:**

This image capture method is commonly used in a web-based AR applications focused on the bulk consumer due to it’s simplicity. As only single camera is required the webcams already installed in mobile devices or cheap standalone cameras are sufficient. The main idea of AR environment creation then is to generate a proper algorithm to evaluate the 3D space basing on the objects position and perspective on the captured image. Then the virtual objects are superimposed onto the image according to evaluated object location properties. This method however generates 2D images only with an illusion of 3 dimensional space. The modern development of image processing introduced the stereoscopic 3D image generation out of 2D one, however it requires application of complex algorithms which would decrease the FPS ratio and certainly would not be as accurate as stereoscopic video capture (but is perfect for already captured video files which do not need to be processed in real-time).

**Stereoscopic video capture:**

Stereoscopic video capture method is more complex and requires the usage of two cameras, however it provides more detailed, virtual 3D environment. As two cameras capture the image simultaneously the main issue is the proper synchronization of both videos. Having this problem solved it is possible to specify each object’s position in 3D space and generate more detailed virtual environment. It provides a possibility to superimpose 3D virtual model onto the obtained image very accurately in two different perspectives. In conclusion with the usage of simple 3D glasses the viewer gets very realistic and precise image which can be generated in real-time.

**Camera calibration:**

Both approaches are affected by shortcoming of the camera usage. In theory there is a possibility to define a perfect lens that introduce no distortions. However every lens used in camera generates two types of distortion. The first one is the radial distortion caused by the shape of lens. It occurs as an image bulging phenomenon known as a barrel effect-rays farther from the center of the lens are bent more than those closer in. For radial distortions, the distortion is equal to 0 at the (optical) center of the imager and increases each step toward the periphery.

**LENS**

**SQUARE OBJECT**

**IMAGE PLANE**

Figure 2.2.1. Radial distortion: rays farther from the center of a simple lens are bent too much compared to rays that pass closer to the center; thus, the sides of a square appear to bow out on the image

plane (this is also known as barrel distortion) **[7]**

Tangential distortion results when the lens is not fully parallel to the image plane. This effect can be usually observed in cheap cameras in which the imager is glued to the back of the camera. It occurs as an tangential image perspective deformation. There exist other kinds of lens distortion but typically they have lesser effect on obtained image. To obtain the corrected image the camera calibration should be applied. In practice, this distortion is small and can be characterized by the first few terms of a Taylor series expansion around *r* = 0. For cheap web cameras, we generally use the first two such terms: the first of which is conventionally called *k*1 and the second *k*2. For highly distorted cameras such as fish-eye lenses we can use a third radial distortion term *k*3. In general, the radial location of a point on the imager will be rescaled according to the following equations:



Figure 2.2.2. Camera image before undistortion (left ) and after undistortion (right) **[7]**

For the purposes of this project video capturing mode selection is essential( as the see-through technique is unobtainable and not suitable for this project). Single camera video capture was chosen. This method is commonly used within PC entertainment applications as most of mobile computers are equipped with single webcam and a vast majority of consumers would use them to run the application. This solution also eliminates the issues connected with videos synchronization so it was found to be the most suitable choice.

## Marker detection

The marker detection process is the essential step in designing AR 2D marker based application. Accuracy of this algorithm influence deeply on further steps as if no marker is found the virtual models should not be superimposed onto the captured image. There are several detection methods which should be introduced to fully understand the detection problem and choose the most suitable solution. Only overall view on detection methods can be found here as the detailed algorithm process is fully described in the next Chapter (Chapter 3.3.). Whole marker detection analysis would be provided as the full most common algorithm steps with openCV library methods applied as this library was chosen for final project implementation (Chapter 3.1.):

* **Blob detection**

To obtain a maximum accuracy for finding markers the captured image should be processed to distinguish and labelize every object that can be the marker. There are two main approaches to evaluate it, however only the first steps differ and the rest of the algorithm stays unchanged( Figure 2.3.1):

The first one assumes color extraction process as the marker can be designed as a monochromatic (usually red, green or blue) shape or only it’s corner are colored with different colors to evaluate it’s rotation value. In this case captured RGB image should be splitted into 3 channels and each of them should be thresholded to get the final 3 binary images.

The second one is based on the black and white 2D markers marked out by the unique shape. As the colors are not relevant in this approach whole image should be converted into one channel grayscale and then the threshold filter should be applied to measure only the single pixels luminosity values. This solution is faster but would not work for color detection.

The rest of the algorithm is the same for both approaches.   
To enhance the detection process the binary image(s) should be sharpened by the morphology filters erode and dilate. They are usually applied to get rid of noises and image distortions as the single pixels surrounded by the opposite value one are absorbed to the surrounding regions in the resulting image. Note that this type of correction can slightly change the object’s contours so these filters should be used with relevant parameters( since the overall algorithm performance can be decreased there is no specific requirement to use them if the benefits are too small). Having the luminosity value of each pixel the blob detection should be performed to find all pixel groups with the similar values that creates convex shapes. The most commonly used type of blob detection in OpenCV is the Laplace operator method. Every time a new pixel is assigned to a blob, the values for the bounding box, area and centroid are updated.

**CONVERT TO GRAYSCALE**

**EXTRACT RGB CHANNELS**

**APPLY THRESHOLD FILTER**

**APPLY THRESHOLD FOR SELECTED CHANNEL**

**(OPTIONALLY)  
APPLY MORPHOLOGY FILTERS**

**DETECT BLOBS**

**LABELIZE OBJECTS**

**COLORED MARKER**

**SHAPED MARKER**

Figure 2.3.1. Block diagram of full blob detection procedure.

* **Corner finding**

Having the objects labelized it is essential to evaluate their position and size parameters which would be used for perspective estimation. In fact depending on the chosen algorithm the blob detection step can be completely omitted and corner/edge finding can be applied to raw binary image.   
In order to obtain the corner values various algorithm could be utilized, however only most common, OpenCV supported ones would be described:

First of all the processed image should be converted to a binary image to provide better accuracy for corner detection. (there is no need to binarize the image if the blob detection was performed since the result is already binarized). It can be obtained by applying one of the threshold filters. The classic simple threshold or filters which can provide variable threshold level: adaptive threshold (Gaussian/mean value based) or histogram-based hysteresis threshold should be chosen depending on environmental conditions in which the application would be used. Instead of simple thresholding the Canny edge detection combing threshold and contours finding main features can be used. It is based on computation of the first derivatives in *x* and *y* and then combined into four directional derivatives. The points where these directionalderivatives are local maxima are then candidates for assembling into edges. The most significant new dimension advantage is assembling the individual edge candidate pixels into contours*.* They are formed by applying a hysteresis threshold to the pixels. This means that there are two thresholds, an upper and a lower. If a pixel has a gradient larger than the upper threshold, then it is accepted as an edge pixel- if a pixel is below the lower threshold, it is rejected. If the pixel’s gradient is between the thresholds, then it will be accepted only if it is connected to a pixel that is above the high threshold.

Depending on the shape of the marker to be found there are several choices to be made in order to extract shape features:   
If the desired shape is a polygon the best choice is to use the subpixel corner finding algorithm. The actual computation of the subpixel location uses a system of dot-product expressions that all equal 0, where each equation arises from considering a single pixel in the region around the iterated pixel.

OpenCV library provides also the features finding algorithms as the Shi and Tomasi corner finding implementation (based on the second derivatives evaluation (using the Sobel operators) that are needed to compute eigenvalues). This method however is not so accurate for extracting geometrical measurements so it is not popular for marker detection procedures. Another supported method is the Hough transformation algorithm. Even though for polygon-shaped markers Hough line transform is not the best choice it is very good solution to use Hough circular transform to extract circular –shaped markers. However the most common 2D markers are square- shaped so this is not popular method of marker extraction.

Although algorithms like the Canny edge detector can be used to find the edge pixels that separate different segments in an image, they do not tell you anything about those edges as entities in themselves. Having contours evaluated the polygon can be approximated by matching the found contour lines and basing on their intersection points obtain the corners coordinates and create polygon sides. Additionaly convexity check should be performed and if the result is positive the full marker shape is ready for extraction using the corner position coordinates. Both side count and convexity check are supported by OpenCV function cvApproxPoly and was used in the final algorithm.

**BINARIZE  
(SIMPLE THRESHOLD/CANNY)**

**FIND CORNERS  
(SUBPIXEL/FEATURES)**

**FIND CONTOURS**

**APPROXIMATE POLYGON  
(INTERSECTION MATCH +   
CONVEXITY CHECK)**

Figure 2.3.2. Block diagram of full corner finding procedure.

* **Marker tracking**

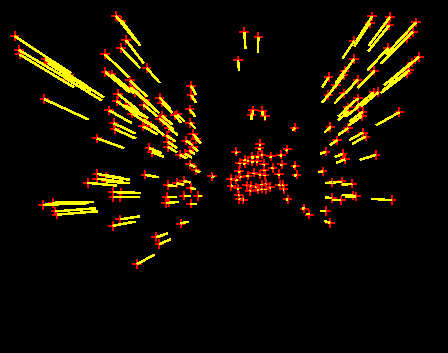
To obtain a better detection accuracy the marker tracking algorithm can be applied. Since the general idea assumes that if marker is found once there is no need to look for it on entire image and only neighbourhood pixel regions are scanned for the desired pattern. This way the application performance can be greatly increased as for high resolution images (HD, Full HD and higher) only fixed region would be set as Region of Interest (ROI) in each cycle. Taking into consideration the optical flow of processed image there are two kinds of marker tracking techniques can be distinguished:

**Sparse optical flow** based technique implemented in Lucas- Kanade tracking algorithm. It require previously specified template which should be tracked with unique features like corners, edges or patterns so the marker detection is needed to be performed at least once for the first captured frame. Assuming that the camera view can be changed during the tracking process the marker detection have to be performed each time the specified object is out of reach as the lighting conditions may vary for each view. Sparse optical flow based algorithms are marked out by 3 assumptions:   
1. Brightness constancy:A pixel from the image of an object in the scene does not change in appearance as it (possibly) moves from frame to frame. For grayscale images the brightness of a pixel does not change as it is tracked from frame to frame.

2. Temporal persistence or “small movements”: The image motion of a surface patch changes slowly in time. In practice, this means the temporal increments are fast enough relative to the scale of motion in the image that the object does not move much from frame to frame.

3. Spatial coherence: Neighboring points in a scene belong to the same surface, have similar motion, and project to nearby points on the image plane. **[7]**

**Dense optical flow** based technique represented by a Horn- Schunck tracking and block matching algorithms. This method is based on pixel displacement measurements so there is possibility to track specified patterns even without previous marker detection. Some kind of velocity can be associated with each pixel in the frame or, equivalently, some displacement that represents the distance a pixel has moved between the previous frame and the current frame. Although this method is more accurate than the previous one it is also more complex what manifest very clearly in high computional cost of algorithm. Some additional disadvantages can occur as this technique is related mainly to the motion capture process. If the object we want to track is homogenous (e.g. white sheet of paper) it’s motion would be very hard to track as most of the white pixels from the previous frame will simply remain white in the next one. Only the edges coordinates may change, and even then only those perpendicular to the direction of motion can be accurately tracked. **[7]**



**INITIAL SEQUENCE**

**FEATURES TRACKING**

**TRACKED FEATURES**

Figure 2.3.3. Example of features tracking algorithm based on Jean-Yves Bouguet and Pietro Perona Visual Navigation project**[11]**

As both of these techniques introduce additional computation cost which would negatively influence on overall project performance none of them was used in the final algorithm. Moreover sparse optical based one requires brightness constancy which cannot be satisfied as the environmental lighting conditions can vary. In fact tracking process is very useful for multiple or unknown object tracking, but for a single fully specified marker there is no need to use additional detection techniques.

**Chapter 3**

# Project Design

## Library choice

During last 10 years the Augmented Reality became a popular object of interest, hence multiple open source libraries were developed and aimed to mass consumer. Nowadays practically everyone can create simple AR effects using ready-to-use applications scattered throughout the internet. However this work is focused on designing the AR algorithm from the beginning so the existing libraries should be only used for analysis and requirements statement. The majority of available open source libraries (GNU General Public Licence) were developed basing on ARToolKit created in 1999 by Hirokazu Kato at the HITlab. Next versions were firstly created to support different programming languages giving the opportunity to add new features and improve the existing algorithms. As Tomohiko Koyama introduced FLARToolKit supporting ActionScript3 the AR based applications started to emerge throughout the internet and could be run online on websites as simple .swf files. Since these two libraries are known to be revolutionary and each of them introduce new features- both of them were analysed to create an unique algorithm for this work. Many other libraries were created basing on ARToolKit e.g. NyARToolKit( Java, C++, C#, Android), FLARToolKit Alchemy(improved performance), ARToolKit Professional(extended features)- each of them supported other open source libraries for rendering functionality (Papervision3D,OpenSceneGraph,OpenGL) as basicly they were created as tracking libraries and nodekits. The key libraries development process can be presented as an evolution graph introduced by Tomohiko Koyama:

**ARToolKit (C)  
Prof. Hirokazu Kato**

**NyARToolKit (Java)  
Ryo Iizuka (aka Nyatla)**

**FLARToolKit (AS3)  
Tomohiko Koyama (aka Saqoosha)**

**NyARToolKit (C++)  
Ryo Iizuka (aka Nyatla)**

**FLARToolKit Alchemy  
Ryo Iizuka (aka Nyatla)**

Figure 3.1.1. ARToolKit library evolution graph**[12]**

Despite of variety of Augmented Reality technology based libraries the Open Source Computer Vision (OpenCV) library was chosen for this project purposes. Since it supports single Computer Vision based actions like image capture, image processing, transformations and calibrations the whole project algorithm could be created from scratch. This choice created an opportunity to develop basic AR algorithm based on 2D markers and gave the possibility to improve it and extend it with new features in future. Since the whole project was created step by step the simple debug mode could be created to monitor all application variables needed for performance analysis (full performance and results analysis and other libraries features comparison can be found in Chapter 6). OpenCV version 2.1. was chosen due to the version stability as later versions were reported to be unstable and need to be fixed.

## Development Process

Tools used in development process have the direct impact on code quality, overall software delivery time and file archivisation. The choice of the proper engineering model was aimed to fast delivery time and secure and reliable code and documentation archives. All these requirements are satisfied by the Continuous Integration software engineering by means of using the Apache Subversion (SVN) software versioning. It gave the opportunity to use simple revision control system hence the project could be developed in small parts giving the access to whole development process and documents creation history.

As this work is basing on Open Source libraries and tools the chosen SVN system was supported by the google code project hosting as it turned out to be a solution for simple repository creation, control and gave the possibility to share this project under Open Source License.

The main project’s page which would be extended by external documentation and ‘how to’ instructions can be found at:

<http://code.google.com/p/augmented-reality-cards/>

Whole project’s repository which includes code, documentation and images, video, templates files can accessed by:

<https://reaveth@augmented-reality-cards.googlecode.com/svn/>

Simplified repository structure graph looks as follows:

**BRANCHES (IN-DEVELOPMENT PROJECT VERSIONS)**

**DOCS (DOCUMENTATION AND INSTRUCTIONS)**

**TRUNK (ACTUAL WORKING PROJECT)**

**TEMPLATES (TEMPLATES, IMAGE AND VIDEO FILES)**

**WIKI (FUTURE WIKI RESOURCES)**

## Project Algorithm

The main project algorithm is based on standard 2D Augmented Reality implementation procedures used in the most common Toolkits, however there are many small changes which have a major influence on project performance. These algorithm alterations were focused on the marker detection accuracy improvement mainly, remembering that the frames per second ratio should meet the real-time performance requirement. Each step was designed to create unique marker based AR technique implementation supported by OpenCV library solutions. Simplified algorithm block diagram for single frame is presented below (Figure 3.3.1.):

**Capture camera frame**

**Binarize the input image**

**Find contours and detect squares**

**Match detected squares with marker template**

**Render virtual object**

**Extract squares and apply perspective transformation**

Figure 3.3.1. Simplified Augmented Reality algorithm block diagram

In order to fully understand this unique and improved 2D augmented reality implementation each step should be described in detail and enhancements should be introduced.

1. **Capture the camera frame.**

Any type of static image can be used as an input for introduced algorithm. However to satisfy the real-time processing requirements the analyzed image should be streamed by the camera device. In order to get the input for each cycle the camera has to provide single frames processed and augmented by the virtual element ( Naturally the video file previously saved on disk can be also used- it has to be split into single frames and each of them is then read as input for each cycle. Such method was used for rendered video file frames acquisition in rendering step). This task was done by means of using the cvCreateCameraCapture() which handles the camera image streaming. If the time of getting and displaying each frame is short enough the generated image sequence is perceived as a video.

Figure 3.3.2. Original image captured from the camera device

**Enhancements:**

As the camera capture issue is rather simple and handled by used functions the Frames per Second ratio counter was implemented to provide the necessary information about the application performance (mostly used in debug mode). To achieve satisfactory FPS accuracy the timer has to be initialized when the streaming starts. The number of cycles (processed frames) are counted and divided by the exact time difference (time of the frame acquisition – time of the timer initialization). This way the overall performance can be monitored. It was essential to implement this improvement to check if the real-time processing requirement was satisfied.

1. **Binarize the input image.**

Method selected for this work applies to RGB images only as this is the most common and default color model used to represent and display images in electronic systems.

The input frame cannot be binarized directly from the color image so it should be converted to one-channel grayscale model. This can be obtained by splitting the original image into 3 single 8-bit color channels (Red, Green, Blue). Having these evaluated the grayscale 1-channel image can be computed using the perceptually weighted formula for each pixel:

As a result 1-channel 8-bit image is obtained (each pixel represents the luminosity value in range of 0-255). This method is supported by a cvCvtColor() function which converts input image’s color space into another. Alternatively there is possibility to use each 8-bit channel to generate 3 binarized images which should be processed later. This method however is used for color detection. For the simple contour detection only one image is necessary and simultaneous 3-chanell processing would significantly decrease the overall project performance. As it is essential to compute the final output image in real-time the first method was selected.

Figure 3.3.3. Grayscale model image acquisition diagram

The obtained image should be binarized then. The main idea of this technique assumes that each processed pixel of the input image should be set to max or min value (0 and 255 have been chosen for this project to provide the best contrast) depending if it’s luminosity level is above or below the specified threshold value. Simple threshold algorithm was not sufficient for this project purposes as the environment light sources can change decreasing the marker contour visibility as well as the emerging reflections can vastly decrease the detection accuracy. To deal with this problem the adaptive threshold (dynamic threshold) method has been introduced as it is commonly used to eliminate the noises and reflections. The main difference between traditional threshold and adaptive threshold method is that the latter one computes the new pixel value basing not only the single pixel but also the neighbourhood pixels inside the specified region (with computed pixel coordinates as a center) so the threshold value is different for different regions in the image. The Gaussian method have been chosen as it gave the most accurate and clear results. (For threshold methods results comparison see the chapter 6.2). It implements the idea that the pixels in the region around (*x, y*) are weighted according to a Gaussian function of their distance from that center point. This specific binarization process is supported by cvAdaptiveThreshold() function which was used in this project.

**Enhancements:** As the simple adaptive threshold method occurred to be not sufficient and the binarized image could not be used as a final source for the later computations the major changes had to be applied to the algorithm. First of all the binarization step have proven to be the most essential part for marker contours detection so the improvements had to be focused on the square detection rather than the whole marker extraction.   
 The image binarization process has been divided into two phases. First one corresponds to accurate contours exposition and is applied in this step. The second one focus on the extracted and cropped square binarization to provide the most accurate results in the matching step (See 5. of this algorithm analysis).  
 To improve the implemented adaptive thresholding method the marker tracking elements have been introduced. The main idea of this improvement assumes that the processed image dynamic threshold block size parameter fully depends on the previously computed marker bounding box size. This way if the previously found marker is bigger (closer to the camera device) the block size proportionally increases- the contours are thicker and easier to find. It is similar to the marker tracking methods which set the ROI area near the previously found marker to increase the algorithm performance.

Figure 3.3.4. Binarized image example

1. **Find contours and detect squares.**

Selected square detection method is based on contours finding. To extract the contours sequences from the input binary image the cvFindCountours() function is used. It’s main advantage is that it compresses horizontal, vertical, and diagonal contours segments, leaving only their ending points. Having these coordinates the simple shapes can be computed by means of polygon approximation algorithms. Based on simple Freeman chain approximation method (Figure 3.3.5) the shape is represented as a sequence of steps in one of eight directions designated by an integer from 0 to 7. To convert the Freeman chains to polygonal representation the cvApproxPoly() is utilized as the next step. Having the contours ending points evaluated earlier it returns the sequence of straight contours which create an approximated polygon. As the desired result is the quadrangle the additional requirements have to be satisfied:

- sides (contours) number equal to 4

- convex polygon found

- polygon area bigger than specified value ( as the marker size

should be limited to a minimal value)

Figure 3.3.5. Obtained contours visualization based on the binary image

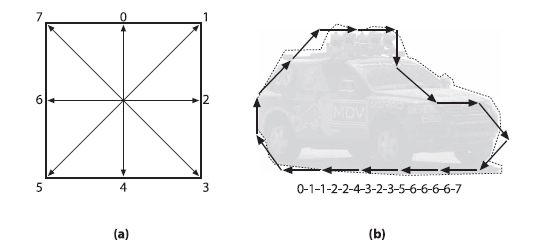


Figure 3.3.6. Panel a, Freeman chain moves are numbered 0–7; panel b, contour converted to a Freeman chain-code representation starting from the back bumper **[7]**

Having the result quadrangle contours sequences evaluated the corners coordinates can be evaluated by means of using OpenCV sequence reader. As this step computes every quadrangle coordinates and do not affect the performance visibly- there is no need to enhance this step implementation.

1. **Extract squares and apply perspective transformation.**

Sdfsdf

1. **Match detected squares with marker template.**

Sdfsdf

1. **Render virtual object.**

Sdfsdf

## GUI

Opis wyboru srodowiska do tworzenia GUI I krotki opis jego powstania

**Chapter 4**

# Internal Specification

## Main program functions

Dokladny opis funkcjio uzytych w programie oraz rozwieniecie teaoretyczne + wzory przy kluczowych funkcjach (thresholding, homograficzna transforacja, podkladanie obrazu itp)

## Graphical User Interface

Dokladniejszy opis GUI I podzial na poszczegolne elementy.

**Chapter 5**

# External Specification

## ‘How to’ instruction

Instrukcja w jaki sposob uzywac aplikacji

## Errors handling

Opis najczestrzych errorow (np brak zaladowanego markera lub obrazka czy wideo)

**Chapter 6**

# Testing and results analysis

## Marker choice analysis

Porownanie roznych markerow I dokladnosci ich wykrywania

## Environment dependencies

Porowanie wynikow wykrycia markera dla roznych srodowisk (jasno, normalnie, ciemno)

## Threshold methods

Rozne metody thresholdu

## Displaying static image and video

Porownanie wyswietlania obrazkow I wideo

## Camera parameters

Wyniki dla roznych kamer (w laptopie, statyczna Logitech, creative social hd z autofocusem)

**Chapter 7**

# Summary

Krotkie podsumowanie calosci

# Bibliography

[1] Ronald Azuma. “A Survey of Augmented Reality”, Presence: Teleoperators and Virtual Environments vol. 6, no. 4, pp. 355-385, August 1997.

[2] P. Milgram and A. F. Kishino, “Taxonomy of Mixed Reality Visual Displays” IEICE Transactions on Information and Systems, vol. E77-D no. 12, pp. 1321-1329, 1994.

[3] Joe Lamantia. “Inside Out: Interaction Design for Augmented Reality” UXmatters, August 17, 2009.

[4] Christian Doppler. “History of Mobile AR”, Laboratory for Handheld Augmented Reality,   
(accessed 12/07/2011)

[5] Rick Oller. “Augmented Reality”, Scribd.com, April 2010.  
http://www.scribd.com/doc/44664310/Augmented-Reality  
(accessed 14/07/2011)

[6] Tim Perdue. “Applications of Augmented Reality”, About.com Guide,  
http://newtech.about.com/od/softwaredevelopment/a/Applications-Of-Augmented-Reality.htm  
(accessed 14/07/2011)

[7] Gary Bradski, Adrian Kaehler. “Learning OpenCV Computer Vision with the OpenCV Library”, O’Reilly Media, 2008.

[8] Oliver Bimber, Ramesh Raskar. „Spatial Augmented Reality

Merging Real and Virtual Worlds”, A K Peters, Ltd., pp. 71-90, 2005.

[9] Hong Hua. “Stereoscopic Displays” Optical Sciences Center, University of Arizona, Tucson, AZ 85721, publications materials, June 2004.

[10] Gianpierre Villagomez. “Augmented Reality” University of Kansas, lectures materials, vol.3, December 2010.

[11]Jean-Yves Bouguet, Pietro Perona.”Visual Navigation” California Institute of Technology, November 1998.

[12] Tomohiko Koyama.”Introduction to FLARToolKit”.  
http://saqoo.sh/a/labs/FLARToolKit/Introduction-to-FLARToolKit.pdf  
(accessed 24/07/2011)

# Contents of the CD

The content of attached CD directory structure is described by the following diagram:

**SRC**…………………………………………………………..………….**SOURCE CODE**

**EXC**………………………………………………………………...…….**EXECUTABLES**

**LIB**……………………………………………………..………………**USED LIBRARIES**

**OPECV**

**RSC**…………………………………………..……………………………..**RESOURCES**

**TEMPLATE.JPG**

**TEMPLATE.PSD**

**IMAGE.JPG**

**VIDEO.AVI**

**DOC**………………………………………………………………**DOCUMENTATION**